<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Principles of Endocrinology</td>
</tr>
<tr>
<td>2</td>
<td>The Endocrine Patient</td>
</tr>
<tr>
<td>3</td>
<td>Genetic Control of Peptide Hormone Formation</td>
</tr>
<tr>
<td>4</td>
<td>Mechanism of Action of Hormones That Act on Nuclear Receptors</td>
</tr>
<tr>
<td>5</td>
<td>Mechanism of Action of Hormones That Act at the Cell Surface</td>
</tr>
<tr>
<td>6</td>
<td>Laboratory Techniques for Recognition of Endocrine Disorders</td>
</tr>
<tr>
<td>7</td>
<td>Neuroendocrinology</td>
</tr>
<tr>
<td>8</td>
<td>Anterior Pituitary</td>
</tr>
<tr>
<td>9</td>
<td>Posterior Pituitary Gland</td>
</tr>
<tr>
<td>10</td>
<td>Thyroid Physiology and Diagnostic Evaluation of Patients with Thyroid Disorders</td>
</tr>
<tr>
<td>11</td>
<td>Thyrotoxicosis</td>
</tr>
<tr>
<td>12</td>
<td>Hypothyroidism and Thyroiditis</td>
</tr>
<tr>
<td>13</td>
<td>Nontoxic Goiter and Thyroid Neoplasia</td>
</tr>
<tr>
<td>14</td>
<td>The Adrenal Cortex</td>
</tr>
<tr>
<td>15</td>
<td>Endocrine Hypertension</td>
</tr>
<tr>
<td>16</td>
<td>The Physiology and Pathology of the Female Reproductive Axis</td>
</tr>
<tr>
<td>17</td>
<td>Fertility Control: Current Approaches and Global Aspects</td>
</tr>
<tr>
<td>18</td>
<td>Disorders of the Testes and the Male Reproductive Tract</td>
</tr>
<tr>
<td>19</td>
<td>Sexual Dysfunction in Men and Women</td>
</tr>
<tr>
<td>20</td>
<td>Endocrine Changes in Pregnancy</td>
</tr>
<tr>
<td>21</td>
<td>Endocrinology of Fetal Development</td>
</tr>
<tr>
<td>22</td>
<td>Disorders of Sex Differentiation</td>
</tr>
<tr>
<td>23</td>
<td>Normal and Aberrant Growth</td>
</tr>
<tr>
<td>24</td>
<td>Puberty, Ontogeny, Neuroendocrinology, Physiology, and Disorders</td>
</tr>
<tr>
<td>25</td>
<td>Endocrinology and Aging</td>
</tr>
<tr>
<td>26</td>
<td>Hormones and Disorders of Mineral Metabolism</td>
</tr>
<tr>
<td>27</td>
<td>Metabolic Bone Disease</td>
</tr>
<tr>
<td>28</td>
<td>Kidney Stones</td>
</tr>
<tr>
<td>29</td>
<td>Type 2 Diabetes Mellitus</td>
</tr>
<tr>
<td>30</td>
<td>Type 1 Diabetes Mellitus</td>
</tr>
<tr>
<td>31</td>
<td>Complications of Diabetes Mellitus</td>
</tr>
<tr>
<td>32</td>
<td>Glucose Homeostasis and Hypoglycemia</td>
</tr>
<tr>
<td>33</td>
<td>Obesity</td>
</tr>
<tr>
<td>34</td>
<td>Disorders of Lipid Metabolism</td>
</tr>
<tr>
<td>35</td>
<td>Pathogenesis of Endocrine Tumors</td>
</tr>
<tr>
<td>36</td>
<td>Multiple Endocrine Neoplasia</td>
</tr>
<tr>
<td>37</td>
<td>The Immunoendocrinopathy Syndromes</td>
</tr>
<tr>
<td>38</td>
<td>Gastrointestinal Hormones and Gut Endocrine Tumors</td>
</tr>
<tr>
<td>39</td>
<td>Endocrine-Responsive Cancer</td>
</tr>
<tr>
<td>40</td>
<td>Humoral Manifestations of Malignancy</td>
</tr>
<tr>
<td>41</td>
<td>Carcinoid Tumors, Carcinoid Syndrome, and Related Disorders</td>
</tr>
</tbody>
</table>

**Appendix: Reference Values**
Contributors

Eli Y. Adashi

Presidential Professor of Obstetrics and Gynecology
and John A. Dixon Professor and Chair,
Department of Obstetrics and Gynecology,
University of Utah Health Sciences Center,
Salt Lake City, Utah
The Physiology and Pathology of the Female Reproductive Axis;
Fertility Control: Current Approaches and Global Aspects

Lloyd P. Aiello

Associate Professor of Ophthalmology,
Harvard Medical School;
Assistant Director,
Beetham Eye Institute;
Investigator and Head,
Section of Eye Research,
Joslin Diabetes Center,
Boston, Massachusetts
Complications of Diabetes Mellitus

Andrew Arnold

Murray-Heilig Chair in Molecular Medicine
and Professor of Medicine and Genetics,
University of Connecticut School of Medicine;
Director, Center for Molecular Medicine,
and Chief, Division of Endocrinology and Metabolism,
University of Connecticut,
Farmington, Connecticut
Pathogenesis of Endocrine Tumors

Jennifer Berman
Laura Berman

Assistant Professor,
Department of Urology,
David Geffen School of Medicine at University of California, Los Angeles;
Attending Surgeon,
UCLA Medical Center;
Co-Director, Female Sexual Medicine Center at UCLA,
Los Angeles, California

Sexual Dysfunction in Men and Women

Shalender Bhasin

Professor of Medicine,
University of California, Los Angeles,
School of Medicine;
Chief, Division of Endocrinology, Metabolism, and Molecular Medicine,
Charles Drew University and King-Drew Medical Center,
Los Angeles, California

Sexual Dysfunction in Men and Women

Andrew J. M. Boulton

Professor of Medicine,
University of Manchester;
Consultant Physician,
Manchester Royal Infirmary,
Manchester, United Kingdom

Complications of Diabetes Mellitus

Robin P. Boushey
General Surgery Resident,
Department of Surgery,
University of Toronto,
Toronto, Ontario, Canada
Gastrointestinal Hormones and Gut Endocrine Tumors

Glenn D. Braunstein

Professor of Medicine,
University of California, Los Angeles,
School of Medicine;
Chairman, Department of Medicine,
The James R. Klinenberg, MD, Chair in Medicine,
Cedars-Sinai Medical Center,
Los Angeles, California
Endocrine Changes in Pregnancy

F. Richard Bringhurst

Associate Professor of Medicine,
Harvard Medical School;
Physician,
Massachusetts General Hospital,
Boston, Massachusetts
Hormones and Disorders of Mineral Metabolism

Michael Brownlee

Anita and Jack Saltz Professor of Diabetes Research
and Professor of Medicine and Pathology,
Albert Einstein College of Medicine,
Bronx, New York
Complications of Diabetes Mellitus

Serdar E. Bulun
Associate Professor of Obstetrics-Gynecology and Molecular Genetics;
Director, Division of Reproductive Endocrinology and Infertility,
University of Illinois,
Chicago, Illinois
*The Physiology and Pathology of the Female Reproductive Axis*

Charles F. Burant

Associate Professor,
Department of Internal Medicine,
University of Michigan;
Attending Physician,
University of Michigan Hospitals,
Ann Arbor, Michigan
*Type 2 Diabetes Mellitus*

John B. Buse

Associate Professor of Medicine;
Chief, Division of General Medicine and Clinical Epidemiology;
Director, Diabetes Care Center,
University of North Carolina School of Medicine,
Chapel Hill, North Carolina
*Type 2 Diabetes Mellitus, Type 1 Diabetes Mellitus*

David A. Bushinsky

Professor of Medicine and of Pharmacology and Physiology,
University of Rochester School of Medicine and Dentistry;
Chief, Nephrology Unit,
Strong Memorial Hospital,
Rochester, New York
*Kidney Stones*

Judy L. Cameron
Irene E. and Michael M. Karl Professor of Endocrinology and Metabolism in Medicine, Washington University School of Medicine; Chief of Endocrinology, Diabetes, and Metabolism, Barnes-Jewish Hospital, St. Louis, Missouri

**Glucose Homeostasis and Hypoglycemia**

Terry F. Davies

Professor of Medicine and Director, Division of Endocrinology, Diabetes and Bone Diseases, Mount Sinai School of Medicine; Attending Physician, Mount Sinai Hospital, New York, New York

**Thyroid Physiology and Diagnostic Evaluation of Patients with Thyroid Disorders; Thyrotoxicosis; Hypothyroidism and Thyroiditis**

Marie B. Demay

Associate Professor of Medicine, Harvard Medical School; Associate Physician, Massachusetts General Hospital, Boston, Massachusetts

**Hormones and Disorders of Mineral Metabolism**

Robert G. Dluhy

Professor of Medicine, Harvard Medical School; Clinical Chief, Division of Endocrinology, Diabetes and Hypertension, Brigham and Women's Hospital, Boston, Massachusetts

**Endocrine Hypertension**

Daniel J. Drucker
Professor of Medicine, University of Toronto Faculty of Medicine; Director, Banting and Best Diabetes Centre, Toronto General Hospital, Toronto, Ontario, Canada
Gastrointestinal Hormones and Gut Endocrine Tumors
George S. Eisenbarth

Professor of Pediatrics, Medicine, and Immunology, University of Colorado School of Medicine; Executive Director, Barbara Davis Center for Childhood Diabetes, Denver, Colorado
Type 1 Diabetes Mellitus; The Immunoendocrinopathy Syndromes
Joel K. Elmquist

Associate Professor of Medicine and Neurology, Harvard Medical School, Division of Endocrinology, Beth Israel Deaconess Medical Center, Boston, Massachusetts
Neuroendocrinology
Robert V. Farese Jr.

Professor of Medicine, University of California, San Francisco; Senior Investigator, Gladstone Institute of Cardiovascular Disease, San Francisco, California
Disorders of Lipid Metabolism
Daniel D. Federman
Professor of Medicine,
Division of Internal Medicine,
University of Texas/M.D. Anderson Cancer Center;
Adjunct Professor of Medicine, Internal Medicine, and Cell Biology,
Baylor College of Medicine,
Houston, Texas
Multiple Endocrine Neoplasia

Peter A. Gottlieb

Assistant Professor of Pediatrics and Medicine,
Barbara Davis Center for Childhood Diabetes,
University of Colorado Health Sciences Center,
Denver, Colorado
The Immunoendocrinopathy Syndromes

James E. Griffin

Professor of Internal Medicine
and Diana and Richard C. Strauss Professor in Biomedical Research,
The University of Texas Southwestern Medical Center,
Dallas, Texas
Disorders of the Testes and Male Reproductive Tract

Melvin M. Grumbach

Edward B. Shaw Emeritus Professor of Pediatrics
and Emeritus Chairman, Department of Pediatrics,
University of California, San Francisco,
San Francisco, California
Disorders of Sex Differentiation:
Puberty: Ontogeny, Neuroendocrinology, Physiology, and Disorders

Joel F. Habener
Professor of Medicine,
Harvard Medical School;
Associate Physician and Chief,
Laboratory of Molecular Endocrinology,
Massachusetts General Hospital;
Investigator,
Howard Hughes Medical Institute,
Boston, Massachusetts
Genetic Control of Peptide Hormone Formation

Ian D. Hay

Professor of Medicine,
Mayo Medical School;
Consultant in Endocrinology and Internal Medicine,
Mayo Clinic,
Rochester, Minnesota
Thyroid Physiology and Diagnostic Evaluation of Patients with Thyroid Disorders;
Nontoxic Goiter and Thyroid Neoplasia

Wayne J. G. Hellstrom

Professor of Urology
and Chief, Section of Andrology and Male Dysfunction,
Tulane University Health Sciences Center,
New Orleans, Louisiana
Sexual Dysfunction in Men and Women

Ieuan A. Hughes

Professor of Paediatrics
and Head, Department of Paediatrics,
University of Cambridge;
Honorary Consultant,
Paediatric Endocrinology,
Addenbrooke Hospital,
Cambridge, United Kingdom
Disorders of Sex Differentiation

Michael Kafrissen
Adjunct Professor,
Maternal and Child Health,
University of North Carolina School of Medicine, Chapel Hill, North Carolina;
Vice President,
Ortho-McNeil Pharmaceutical Co.,
Raritan, New Jersey
Fertility Control: Current Approaches and Global Aspects

George G. Klee

Professor of Laboratory Medicine,
Department of Laboratory Medicine,
Mayo Medical School;
Consultant in Clinical Pathology,
Department of Laboratory Medicine and Pathology,
Mayo Clinical Hospitals,
Rochester, Minnesota
Laboratory Techniques for Recognition of Endocrine Disorders

Samuel Klein

Danforth Professor of Medicine,
Department of Medicine,
Washington University School of Medicine,
St. Louis, Missouri
Obesity

David L. Kleinberg

Professor of Medicine and Director,
Neuroendocrine Research Medicine,
New York University School of Medicine;
Attending, Medicine, New York University Medical Center;
Consultant in Medicine, New York Harbor Healthcare Veterans Administration Hospital,
New York, New York
Anterior Pituitary

Barbara E. Kream
Henry M. Kronenberg

Steven W. J. Lamberts

P. Reed Larsen

Jennifer E. Lawrence
Director,
South Georgia Medical Center Diabetes Center;
Chief,
Division of Endocrinology,
South Georgia Medical Center,
Valdosta, Georgia
Endocrine Hypertension

Mitchell A. Lazar

Sylvan H. Eisman Professor of Medicine and Genetics and Director,
Penn Diabetes Center,
University of Pennsylvania School of Medicine;
Chief, Division of Endocrinology, Diabetes, and Metabolism,
Hospital of The University of Pennsylvania,
Philadelphia, Pennsylvania
Mechanism of Action of Hormones That Act on Nuclear Receptors

Joseph A. Lorenzo

Professor of Medicine,
University of Connecticut School of Medicine;
Attending Physician,
John Dempsey Hospital/University of Connecticut Health Center,
Farmington, Connecticut
Metabolic Bone Disease

Malcolm J. Low

Professor,
Department of Behavioral Neuroscience,
Oregon Health and Science University;
Scientist, Vollum Institute,
Oregon Health and Science University,
Portland, Oregon
Neuroendocrinology

Robert W. Mahley
Professor of Pathology and Medicine,  
University of California, San Francisco;  
Director, Gladstone Institute of Cardiovascular Disease,  
San Francisco, California  
*Disorders of Lipid Metabolism*

Stephen J. Marx

Chief, Metabolic Diseases Branch,  
and Chief, Genetics and Endocrinology Section,  
National Institute of Diabetes and Digestive and Kidney Diseases,  
National Institutes of Health,  
Bethesda, Maryland  
*Multiple Endocrine Neoplasia*

Shlomo Melmed

Senior Vice President,  
Academic Affairs,  
Cedars Sinai Medical Center;  
Professor of Medicine and Associate Dean,  
University of California, Los Angeles,  
School of Medicine,  
Los Angeles, California  
*Principles of Endocrinology;  
Anterior Pituitary*

Rebeca D. Monk

Assistant Professor of Medicine,  
University of Rochester School of Medicine and Dentistry;  
Head, Urolithiasis Clinic,  
Strong Memorial Hospital,  
Rochester, New York  
*Kidney Stones*

Richard W. Nesto
Associate Professor of Medicine,
Harvard Medical School, Boston;
Chairman, Department of Cardiovascular Medicine,
Lahey Clinic Medical Center,
Burlington, Massachusetts
Complications of Diabetes Mellitus

Kjell E. Öberg

Chairman, Department of Medical Sciences,
Medical Faculty,
Uppsala University;
Professor of Endocrine Oncology,
University Hospital,
Uppsala, Sweden
Carcinoid Tumors, Carcinoid Syndrome, and Related Disorders

Kenneth S. Polonsky

Adolphus Busch Professor and Chairman,
Department of Medicine,
and Professor, Department of Cell Biology and Physiology,
Washington University School of Medicine,
St. Louis, Missouri
Principles of Endocrinology;
Type 2 Diabetes Mellitus;
Type 1 Diabetes Mellitus

Lawrence G. Raisz

Professor of Medicine,
University of Connecticut School of Medicine and Health Center, Farmington;
Physician, Hartford Hospital and St. Francis Medical Center,
Hartford, Connecticut
Metabolic Bone Disease

Edward O. Reiter
Professor of Pediatrics,
Tufts University School of Medicine, Boston;
Chairman, Department of Pediatrics,
Baystate Medical Center,
Children's Hospital,
Springfield, Massachusetts
Normal and Aberrant Growth

Alan G. Robinson

Vice Provost,
Medical Sciences,
and Executive Associate Dean and Professor of Medicine,
University of California, Los Angeles,
School of Medicine,
Los Angeles, California
Posterior Pituitary Gland

Johannes A. Romijn

Professor of Medicine and Endocrinology,
Department of Endocrinology,
Leiden University School of Medicine and Medical Center,
Leiden, The Netherlands
Obesity

Ron G. Rosenfeld

Credit Union Endowment Professor and Chairman,
Department of Pediatrics, and
Professor, Cell and Developmental Biology,
Oregon Health and Science University;
Physician-in-Chief,
Doernbecher Children's Hospital,
Portland, Oregon
Normal and Aberrant Growth

Richard Santen
Martin-Jean Schlumberger

Professor, University of Paris, Sud;
Chief, Department of Nuclear Medicine and Endocrine Tumors,
Institute Gustave Roussy,
Villejuif, France
Thyroid Physiology and Diagnostic Evaluation of Patients with Thyroid Disorders;
Nontoxic Goiter and Thyroid Neoplasia

Allen Spiegel

Director, National Institute of Diabetes and Digestive and Kidney Diseases,
National Institutes of Health,
Bethesda, Maryland
Mechanism of Action of Hormones That Act at the Cell Surface

Paul M. Stewart

Professor of Medicine, School of Medicine,
University of Birmingham;
Consultant Physician, Queen Elizabeth Hospital,
Birmingham, United Kingdom
The Adrenal Cortex

Gordon J. Strewler
Dennis M. Styne

Simeon I. Taylor

Joseph G. Verbalis

Aaron I. Vinik
Karl H. Weisgraber

Gordon H. Williams

Jean D. Wilson
Foreword

Robert H. Williams

The publication of the tenth edition comes 52 years after the first edition of Williams' Textbook of Endocrinology. There had been other large textbooks of endocrinology, such as Biedl's *Innere Sekretion* in 1916 and Rolleston's *The Endocrine Organs in Health and Disease* in 1936, but only a handful of physicians could be identified as endocrinologists by the middle of the twentieth century. Consequently, Robert H. Williams exercised "powerful persuasion" to overcome the reluctance of the sales staff of the W. B. Saunders Company to publish a book that had no visible audience. In fact, however, Williams was correct in predicting a large readership, because its publication coincided with an explosive increase in basic endocrine science and in the application of this basic information to patients, and with the evolution of endocrinology into a recognized subspecialty of several branches of medicine. Indeed, the book has had a profound impact on endocrine science and on the development of the clinical discipline, and it is appropriate at this time to remember Robert Williams and his contributions to the field and to Textbook of Endocrinology.

Williams described in a memoir the training and the background that led to his development of the textbook. After his graduation from Washington and Lee University, he obtained the M.D. degree at Johns Hopkins University Medical School in 1934. His house staff training was spread between the Mallory Institute of Pathology at the Boston City Hospital, the Department of Medicine at Vanderbilt (where he did research with Tinsley R. Harrison), and the Department of Medicine at Johns Hopkins (where he worked with Warren Longcope). He finished his training at the Massachusetts General Hospital as an endocrine fellow at a time when there were "many quacks in this area throughout the world." He and his mentor, Fuller Albright, became good friends and maintained close contact over the years.

In 1940, Williams was appointed to the staff of the Endocrine Unit of the Thorndike Laboratory in the Harvard Medical Unit at the Boston City Hospital. In 1942, he became head of the Unit, where his research focused principally on the biochemistry and physiology of thyroid disease, including pioneering work on the treatment of thyrotoxicosis with thionamide drugs and with radioactive iodine. In addition, he described the syndrome of biotin deficiency and published papers on adrenal physiology, obesity, and nephrogenic diabetes insipidus.

To attract students and fellows into the field, he developed the concept that "endocrinology is the backbone of metabolism and metabolism is the interstitium of medicine." His students and trainees included at least three future contributors to his textbook, Sidney Ingbar, Peter Forsham, and William H. Daughaday. Daughaday describes Williams as a man of extraordinary exuberance and enthusiasm who took great pleasure in lecturing and in bedside teaching and whose motto was "B(bright) and E(early) and on the B(basis)."

Williams considered himself first and foremost an educator, and in 1948 he moved to the University of Washington as Chairman of the Department of Medicine, where his extraverted and outgoing personality made him a superb teacher, recruiter, administrator, and institution builder. The Endocrine Division in Seattle was very broad and encompassed diabetes mellitus, clinical nutrition, and metabolism, as well as endocrinology. Williams served as President of the Endocrine Society, the American Society for Clinical Investigation, and the Association of American Physicians, and he was the founder of the Association of the Professors of Medicine. In brief, he was an academic giant of twentieth century medicine.

The founding of the Textbook of Endocrinology evolved from his interest in education: "In view of the rapid progress in endocrinology and metabolism, and the fact that our unit at Boston City Hospital had registered very high in undergraduate, graduate, and postgraduate teaching, I decided that there was a great need for a new textbook in endocrinology." The arrangements for the textbook were completed before Williams left Boston, and the aims were clearly described in the preface to the first edition:

The rapidity and extent of advances in endocrinology have made it increasingly difficult for the student and physician to take full advantage of information available for understanding, diagnosis and treatment of clinical disorders. It is the realization of these difficulties that prompted the writing of this book. The main

objective is to provide a condensed and authoritative discussion of the management of clinical endocrinopathies, based upon the application of fundamental information obtained from chemical and physiologic investigations.

The product was a book that over the years has served as an effective bridge between clinical medicine and the science of endocrinology. There may be no other arena of medicine in which the basic and clinical sciences are so tightly interwoven into one discipline. On the one hand, the clinical discipline profits immensely from scientific advances; on the other hand, clinical observations often raise important questions for investigation and on occasion provide answers that have an impact on basic science. By reflecting advances in both areas, the Williams Textbook was designed to convey the intellectual excitement of a rapidly changing scientific base and, at the same time, to promote the integration of a spectrum of disciplines ranging from molecular genetics to patient care into a unified discipline. The achievement of this aim was possible because, from the initial edition, Williams chose contributors who were at the forefront of the field, thereby ensuring the freshness of each edition.

Now, of course, there are several textbooks of endocrinology, but Williams' pioneering book continues to enjoy a growing readership of both the English and the foreign language editions. Williams edited the first five editions, and Edwin L. Bierman completed the editing of the sixth edition after Williams' death in 1979. Jean D. Wilson and Daniel W. Foster edited the seventh and eighth editions and were joined by Henry M. Kronenberg and P. Reed Larsen as editors for the ninth edition. For the tenth edition, Larsen and Kronenberg are joined by editors Kenneth S. Polonsky and Shlomo Melmed, and continuing the tradition set by Williams, the editors of the tenth edition have enlisted an outstanding group of new and former contributors. Saunders continues as publisher.
Endocrinology has changed in many ways during the past 50 years, and the editorial challenges likewise change with each edition. On one level, these challenges reflect scientific advances, such as the explosion of knowledge about hormone action, the development of new and improved diagnostic techniques and imaging modalities, and the application of molecular genetics to biology. On another level, the concept that the discipline of endocrinology was defined by the concept of humoral control mechanisms has become blurred by recognition that the endocrine, immune, and neurologic signaling systems constitute a single integrated system rather than separate control mechanisms. The most significant challenge now, however, is the same as that faced by all textbooks at a time when the volume of published information is rapidly increasing, namely the dilemma of how to take full advantage of developments in electronic publishing and the evolving revolution in information retrieval systems to devise effective learning systems for the near and remote futures. The fundamental educational issues are the same that Williams faced in 1948, namely the need to integrate rapidly evolving basic and clinical science in a cohesive format appropriate for undergraduate, graduate, and continuing medical education. Now, however, the prose text can be enhanced by multimedia additions, and updating can be done in a continuum. How these tools will be utilized to create new types of teaching materials by academicians faced with multiple demands on their time is not entirely clear, but the response to this challenge will determine whether Williams Textbook of Endocrinology will continue to have the same impact over the next 50 years.

Jean D. Wilson

References

5. Daughaday WH. Personal communication.
This tenth edition of Williams Textbook of Endocrinology is a milestone in many respects. A tenth edition per se is testimony to the enduring accomplishments of our predecessors, who have consistently created a product that remains the most popular textbook in this field. It is also now a half century since the publication of the first edition of Williams, a landmark suitably celebrated in a foreword by our coauthor and former editor, Dr. Jean Wilson. Two internationally renowned endocrinologists, Drs. Shlomo Melmed and Kenneth Polonsky, have joined Drs. Reed Larsen and Henry Kronenberg to formulate, co-edit, and assemble this volume. We will strive to meet the high standards maintained by Drs. Wilson and Daniel Foster during their editorial leadership.

Our goal for this first edition of the new millennium was to emulate the achievements of our predecessors by producing a definitive and fresh approach to the presentation of the essentials of clinical endocrinology. Accordingly, we invited a number of new authors, including several European colleagues, to prepare 23 of the 41 chapters. Our challenge to them and to those updating their material was to distill the burgeoning molecular and physiological knowledge into a complete, but relevant, scholarly presentation. Where appropriate, this would include a practical experienced guide as to how the author uses this information in the diagnosis and management of his or her own patients. Achieving such relevance, thoroughness, and practicality requires a unique combination of scientific knowledge and total clinical familiarity best encapsulated in the term “physician-scientist.” We believe that our physician-scientist authors have again met Robert Williams’ stipulation that this text should provide “a condensed and authoritative discussion of the management of clinical endocrinopathies based upon the application of fundamental information obtained from chemical and physiologic investigation.” We hope our readers will agree.

Both new and revised chapters are replete with tables and figures. Highlights of this edition include a new and expanded diabetes section, new chapters on many old and new topics including endocrinology and aging, female reproduction and fertility control, sexual function and dysfunction, kidney stones, the adrenal cortex, endocrine hypertension, endocrine-responsive tumors, and non-insulin-secreting tumors of the gastrointestinal system. A largely new, concise introductory section includes several new chapters discussing mechanisms of hormone action and the clinical approach to the endocrine patient, as well as a thorough guide to the intricacies of the rapidly changing laboratory techniques. The entire section containing chapters on the hypothalamus and both anterior and posterior pituitary disorders is original, and the thyroid section has been thoroughly revised and divided into expanded disorder-based presentations.

Stylistic innovations include page numbers in the introductory outlines for each chapter, which we hope will permit the reader ready access to specific topics. We have also introduced algorithms and clinical guidelines for diagnostic and treatment strategies to crystallize recommendations for each disease. Readers will also note that this edition is published only five years following its predecessor, reflecting the all-too-familiar rapidity with which new knowledge is accumulating in the biomedical disciplines. Its timely appearance despite so much new material reflects the diligent efforts of our editorial staff and especially the new authors.

We would like to express our deep gratitude to the coworkers in our offices, Anita Nichols, Debra Hession, Lynn Moulton, Grace Labrado, Linda Walker, Louise Ishibashi, and Sherri Turner, without whose dedication this project could not have been completed. We also wish to thank our colleagues at Elsevier, Richard Zorab and Cathy Carroll, and our tireless and effective developmental editors, Faith Voit and Joanne Husovsky. Their painstaking attention to every detail is a major contribution to this new edition.

P. Reed Larsen
Henry M. Kronenberg
Shlomo Melmed
Kenneth S. Polonsky
INTRODUCTION

Roughly a hundred years ago, Starling coined the term *hormone* to describe secretin, a substance secreted by the small intestine into the blood stream to stimulate pancreatic secretion. In his Croonian Lectures, Starling considered the endocrine and nervous systems as two distinct mechanisms for coordination and control of organ function. Thus, endocrinology found its first home in the discipline of mammalian physiology.

Work over the next several decades by biochemists, physiologists, and clinical investigators led to the characterization of many hormones secreted into the blood stream from discrete glands or other organs. These investigators showed for the first time that diseases such as hypothyroidism and diabetes could be treated successfully by replacing specific hormones. These initial triumphs formed the foundation of the clinical specialty of endocrinology.

Advances in cell biology, molecular biology, and genetics over the ensuing years began to help explain the mechanisms of endocrine diseases and of hormone secretion and action. Although these advances have embedded endocrinology into the framework of molecular cell biology, they have not changed the essential subject of endocrinology—the signaling that coordinates and controls the functions of multiple organs and processes. Here we would like to survey the general themes and principles that underpin the diverse approaches used by clinicians, physiologists, biochemists, cell biologists, and geneticists to understand the endocrine system.
Hormones can be defined as chemical signals secreted into the blood stream that act on distant tissues, usually in a regulatory fashion. Hormonal signaling represents a special case of the more general process of signaling between cells. Even unicellular organisms such as baker's yeast, Saccharomyces cerevisiae, secrete short peptide mating factors that act on receptors of other yeast cells to trigger mating between the two cells. These receptors resemble the ubiquitous family of mammalian 7-transmembrane spanning receptors that respond to ligands as diverse as photons and glycoprotein hormones. Because these yeast receptors trigger activation of heterotrimeric G proteins just as mammalian receptors do, this conserved signaling pathway must have been present in the common ancestor of yeast and humans.

Signals from one cell to adjacent cells, so-called paracrine signals, often trigger cellular responses that use the same molecular pathways used by hormonal signals. For example, the sevenless receptor controls the differentiation of retinal cells in the Drosophila eye by responding to a membrane-anchored signal from an adjacent cell. Sevenless is a membrane-spanning receptor with an intracellular tyrosine kinase domain that signals in a way that closely resembles the signaling by hormone receptors such as the insulin receptor tyrosine kinase. Since paracrine factors and hormones can share signaling mechanisms it is not surprising that hormones can, in some settings, act as paracrine factors. Testosterone, for example, is secreted into the blood stream but also acts locally in the testes to control spermatogenesis. Insulin-like growth factor I (IGF-I) is a hormone secreted into the blood stream from the liver and other tissues, but it is also a paracrine factor made locally in most tissues to control cell proliferation. Further, one receptor can mediate the actions of a hormone, such as parathyroid hormone, and of a paracrine factor, such as parathyroid hormone-related protein.

Target cells respond similarly to signals that reach them from the blood stream (hormones) or from the cell next door (paracrine factors); the cellular response machinery does not distinguish the sites of origin of signals. The shared final common pathways used by hormonal and paracrine signals should not, however, obscure important differences between hormonal and paracrine signaling system (Fig. 1-1). Paracrine signals do not travel very far; consequently, the specific site of origin of a paracrine factor determines where it will act and provides specificity to that action. When the paracrine factor BMP4 is secreted by cells in the developing kidney, it regulates the differentiation of renal cells; when BMP4 is secreted by cells in bone, it regulates bone formation. Thus, the site of origin of BMP4 determines its physiologic role. In contrast, since hormones are secreted into the blood stream, their sites of origin are often divorced from their functions. We know nothing about thyroid hormone function, for example, that requires that the thyroid gland be in the neck.

Because the specificity of action of paracrine factors is so dependent on their precise site of origin, elaborate mechanisms have evolved to regulate and constrain the diffusion of paracrine factors. Paracrine factors of the hedgehog family, for example, are covalently bound to cholesterol to constrain the diffusion of these molecules in the extracellular milieu. Most paracrine factors interact with binding proteins that block their action and control their diffusion. Chordin, noggin, and many other distinct proteins all bind to various members of the BMP family to regulate their action, for example. Proteases such as tolloid then destroy the binding proteins at specific sites to liberate BMPs so that the BMPs can act on appropriate target cells.

Hormones have rather different constraints. Because they diffuse throughout the body, they must be synthesized in enormous amounts relative to the amounts of paracrine factors needed at specific locations. This synthesis usually occurs in specialized cells designed for that specific purpose. Hormones must then be able to travel in the blood stream and diffuse in effective concentrations into tissues. Therefore, for example, lipophilic hormones bind to soluble proteins that allow them to travel in the aqueous media of blood at relatively high concentrations. The ability of hormones to diffuse through the extracellular space means that the local concentration of hormone at target sites will rapidly decrease when glandular secretion of the hormone stops. Because hormones diffuse throughout extracellular fluid quickly, hormonal metabolism can occur in specialized organs such as the liver and kidney in a way that determines the effective concentration of the hormones in other tissues.

Paracrine factors and hormones thus use several distinct strategies to control their biosynthesis, sites of action, transport, and metabolism. These differing strategies may partly explain why a hormone such as IGF-I, unlike its close relative insulin, has multiple binding proteins that control its action in tissues. As noted earlier, IGF-I has a double life as both a hormone and a paracrine factor. Presumably, the local actions of IGF-I mandate an elaborate binding protein apparatus.

All the major hormonal signaling programs are G protein-coupled receptors, tyrosine kinase receptors, serine/threonine kinase receptors, ion channels, cytokine receptors, nuclear receptors are also used by paracrine factors. In contrast, several paracrine signaling programs are used only by paracrine factors and are probably not used by hormones. For example, Notch receptors respond to membrane-based ligands to control cell fate, but no bloodborne ligands use Notch-type signaling (at least none is currently known). Perhaps the intracellular strategy used by Notch, which involves cleavage of the receptor and subsequent nuclear actions of the receptor's cytoplasmic portion, is too inflexible to serve the purposes of hormones.

The analyses of the complete genomes of multiple bacterial species, the yeast Saccharomyces cerevisiae, the fruit fly Drosophila melanogaster, the worm Caenorhabditis elegans, the plant Arabidopsis thaliana, and humans have allowed a comprehensive view of the signaling machinery used by various forms of life. As noted already, S. cerevisiae uses G protein-linked receptors; this organism, however, lacks tyrosine kinase receptors and nuclear receptors that resemble the estrogen/thyroid receptor family. In contrast, the worm and fly share with humans the use of each of these signaling pathways, although with substantial variation in number of genes committed to each pathway. For example, the Drosophila genome encodes 20 nuclear receptors, the C. elegans genome encodes 270, and the human genome encodes (tentatively) more than 50. These patterns suggest that ancient multicellular animals must have already established the signaling systems that are the foundation of the endocrine system as we know it in mammals.

Even before the sequencing of the human genome, sequence analyses had made clear that many receptor genes are found in mammalian genomes for which no clear ligand or function was known. The analyses of these "orphan" receptors has succeeded in broadening the current understanding of hormonal signaling. For example, the liver X receptor (LXR) was one such orphan receptor found when searching for unknown nuclear receptors. Subsequent experiments showed that oxygenated derivatives of cholesterol are the ligands for LXR, which regulates genes involved in cholesterol and fatty acid metabolism. The example of LXR and many others raise the question of what constitutes a hormone. The classical view of hormones is that they are synthesized in discrete glands and have no function other than activating receptors on cell membranes or in the nucleus. In contrast, cholesterol, which is converted in cells to oxygenated derivatives that activate the LXRs, uses a hormonal strategy to regulate its own metabolism. Other orphan nuclear receptors respond similarly to ligands, such as bile acids and fatty acids. These "hormones" have important metabolic roles quite separate from their signaling properties, although the hormone-like signaling serves to allow regulation of the
metabolic function. The calcium-sensing receptor is an example from the G protein-linked receptor family of receptors that responds to a nonclassical ligand, ionic calcium. Calcium is released into the blood stream from bone, kidney, and intestine and acts on the calcium-sensing receptor in parathyroid cells, renal tubular cells, and other cells to coordinate cellular responses to calcium. Thus, many important metabolic factors have taken on hormonal properties as part of a regulatory strategy.
ENDOCRINE GLANDS

Hormone formation may occur either in localized collections of specific cells, in the endocrine glands, or in cells that have additional roles. Many protein hormones, such as growth hormone, parathyroid hormone, prolactin, insulin, and glucagon, are produced in dedicated cells by standard protein synthetic mechanisms common to all cells. These secretory cells usually contain specialized secretory granules designed to store large amounts of hormone and to release the hormone in response to specific signals. Formation of small hormone molecules initiates with commonly found precursors, usually in specific glands such as the adrenals, gonads, or thyroid. In the case of the steroid hormones, the precursor is cholesterol, which is modified by various hydroxylations, methylations, and demethylations to form the glucocorticoids, androgens, and estrogens, and their biologically active derivatives. In contrast, the precursor of vitamin D, 7-dehydrocholesterol, is produced in skin keratinocytes, again from cholesterol, by a photochemical reaction. Leptin, which regulates appetite and energy expenditure, is formed in adipocytes, thus providing a specific signal reflecting the organism’s nutritional state to the central nervous system.

Thyroid hormone synthesis occurs via a unique pathway. The thyroid cell synthesizes a 660,000-kd homodimer, thyroglobulin, which is then iodinated at specific iodotyrosines. Certain of these “couple” to form the iodothyronine molecule within thyroglobulin, which is then stored in the lumen of the thyroid follicle. In order for this to occur, the thyroid cell must concentrate the trace quantities of iodide from the blood and oxidize it via a specific peroxidase. Release of thyroxine (T₄) from the thyroglobulin requires its phagoctyosis and cathepsin-catalyzed digestion by the same cells.

Hormones are synthesized in response to biochemical signals generated by various modulating systems. Many of these systems are specific to the effects of the hormone product; for example, parathyroid hormone synthesis is regulated by the concentration of ionized calcium, whereas gonadal, adrenal, and thyroid hormone synthesis is achieved by the hormonostatic function of the hypothalamic-pituitary axis. Cells in the hypothalamus and pituitary monitor the circulating hormone concentration and secrete trophic hormones that activate specific pathways for hormone synthesis and release. Typical examples are luteinizing (LH) follicle-stimulating (FSH), thyroid-stimulating (TSH), and adrenocorticotropic (ACTH) hormones.

These trophic hormones increase rates of hormone synthesis and secretion and also may induce target cell division, thus causing enlargement of the various target glands. For example, in hypothroid individuals living in iodine-deficient areas of the world, TSH secretion causes a marked hyperplasia of thyroid cells. In such regions, the thyroid gland may be 20- to 50-fold its normal size. Adrenal hyperplasia occurs in patients with genetic deficiencies in cortisol formation. Hypertrophy and hyperplasia of parathyroid cells, in this case initiated by an intrinsic response to the stress of hypocalcemia, occur in patients with renal insufficiency or calcium malabsorption.

Hormones may be fully active when released into the blood stream (e.g., growth hormone or insulin) or may require activation in specific cells to produce their biological effects. These activation steps are often highly regulated. For example, the T₄ released from the thyroid cell is a prohormone that must undergo a specific deiodination to form the active 3,5,3’ triiodothyronine (T₃). This deiodination reaction can occur in target tissues, such as in the central nervous system; in the thyrotrhophs, where T₄ provides feedback regulation of TSH production; or in hepatic and renal cells, from which T₃ is released into the circulation for uptake by all tissues. A similar post secretory activation step catalyzed by a 5-reductase causes tissue-specific activation of testosterone to dihydrotestosterone in target tissues, including the male urogenital tract and genital skin, as well as in the liver. Vitamin D undergoes hydroxylation at the 25 position in the liver and in the 1 position in the kidney. Both hydroxylations must occur to produce the active hormone 1,25(OH)₂ vitamin D. The activity of the 1-hydroxylase, but not that of the 25-hydroxylase, is stimulated by parathyroid hormone and reduced plasma phosphate but is inhibited by calcium and 1,25(OH)₂ vitamin D.

Hormones are synthesized as required on a daily, hourly, or minute-to-minute basis with minimal storage, but there are significant exceptions. One such exception is the thyroid gland, which contains enough stored hormone to last for about two months. This permits a constant supply of this hormone despite significant variations in the availability of iodine. If iodine deficiency is prolonged, however, the normal reservoirs of thyroxine can be depleted.

The various feedback signaling systems exemplified above provide the hormonal homeostasis characteristic of virtually all endocrine systems. Regulation may include the central nervous system or local signal recognition mechanisms in the glandular cells, such as the calcium-sensing receptor of the parathyroid cell. Superimposed, centrally programmed increases and decreases in hormone secretion or activation through neuroendocrine pathways also occur. Examples include the circadian variation in the secretion of ACTH directing the synthesis and release of cortisol. The monthly menstrual cycle exemplifies a system with much longer periodicity that requires a complex synergism between central and peripheral axes of the endocrine glands. Disruption of hormonal homeostasis due to glandular or central regulatory system dysfunction has both clinical and laboratory consequences. Recognition and correction of these are the essence of clinical endocrinology.
TRANSPORT OF HORMONES IN BLOOD

Protein hormones and some small molecules such as the catecholamines are water-soluble and readily transported by the circulatory system. Others are nearly insoluble in water (e.g., the steroid and thyroid hormones), and their distribution presents special problems. Such molecules are bound to 50 to 60-kd carrier plasma glycoproteins such as thyroxine-binding globulin (TBG), sex hormone-binding globulin (SHBG), and corticosteroid-binding globulin (CBG), as well as to albumin. These ligand-protein complexes serve as reservoirs of these hormones, ensure ubiquitous distribution of their water-insoluble ligands, and protect the small molecules from rapid inactivation or excretion in the urine or bile. Without these proteins, it is unlikely that hydrophobic molecules would be transported much beyond the veins draining the glands in which they are formed. The protein-bound hormones exist in rapid equilibrium with the often minute quantities of hormone in the aqueous plasma. It is this "free" fraction of the circulating hormone that is taken up by the cell. It has been shown, for example, that if tracer thyroid hormone is injected into the portal vein in a protein-free solution, it is bound to hepatocytes at the periphery of the hepatic sinusoid. When the same experiment is repeated with a protein-containing solution, there is a uniform distribution of the tracer hormone throughout the hepatic lobule. Despite the very high affinity of some of the binding proteins for their ligands, one specific protein may not be essential for hormone distribution. For example, in humans with a congenital deficiency of TBG, other proteins, transthyretin and albumin, subsume its role. Because the affinity of these secondary thyroid hormone transport proteins is several orders of magnitude lower than that of TBG, it is possible for the hypothalamic-pituitary feedback system to maintain free thyroid hormone in the normal range at a much lower total hormone concentration. The fact that the "free" hormone concentration is normal in subjects with TBG deficiency indicates that it is this free moiety that is defended by the hypothalamic-pituitary axis and is the active hormone.

The availability of gene-targeting techniques has allowed specific tests of the physiologic role of several hormone-binding proteins. For example, mice with targeted inactivation of the vitamin D-binding protein (DBP) have been generated. Although the absence of DBP markedly reduces the circulating concentration of vitamin D, the mice are otherwise normal. However, they show enhanced susceptibility to a vitamin D-deficient diet because of the reduced reservoir of this sterol. In addition, the absence of DBP markedly reduces the half-life of 25(OH)D$_2$ by accelerating its hepatic uptake, making the mice less susceptible to vitamin D intoxication.

In rodents, transthyretin (TTR) carries retinol-binding protein and is also the principal thyroid hormone-binding protein. This protein is synthesized in the liver and in the choroid plexus. It is the major thyroid hormone-binding protein in the cerebrospinal fluid of both rodents and humans and was thought to perhaps serve an important role in thyroid hormone transport into the central nervous system. This hypothesis has been disproven by the fact that mice without TTR have normal concentrations of T$_4$ in the brain as well as of free T$_4$ in the plasma. To be sure, the serum concentrations of vitamin A and total T$_4$ are decreased, but the knockout mice have no signs of vitamin A deficiency or hypothyroidism. Such studies suggest that these proteins primarily serve distributive/reservoir functions.

Protein hormones and some small ligands (e.g., catecholamines) produce their effects by interacting with cell surface receptors. Others, such as the steroid and thyroid hormones, must enter the cell to bind to cytosolic or nuclear receptors. In the past, it has been thought that much of the transmembrane transport of hormones was passive. Evidence is now in hand that there are specific organic anion transporters involved in cellular uptake of thyroid hormone (see reference ). This may be found to be the case for other small ligands as well, revealing yet another mechanism for ensuring the distribution of a hormone to its site of action.
TARGET CELLS AS ACTIVE PARTICIPANTS

Hormones determine cellular target actions by binding with high specificity to receptor proteins. Whether a peripheral cell is hormonally responsive depends to a large extent on the presence and function of specific and selective hormone receptors. Receptor expression thus determines which cells will respond, as well as the nature of the intracellular effector pathways activated by the hormone signal. Receptor proteins may be localized to the cell membrane, cytoplasm, and nucleus. Broadly, polypeptide hormone receptors are cell-membrane associated, whereas soluble intracellular proteins selectively bind to steroid hormones (Fig. 1-2) (Figure Not Available). This idea of selective localization has recently been challenged, however, because related sequences can be found in multiple cellular compartments.

Membrane-associated receptor proteins usually consist of extracellular sequences that recognize and bind ligand, transmembrane anchoring hydrophobic sequences, and intracellular sequences, which initiate intracellular signaling. Intracellular signaling is mediated by soluble second messengers (e.g., cyclic AMP) or by activation of intracellular signaling molecules (e.g., signal transduces and activates of transcription [STAT] proteins). Receptor-dependent activation of heterotrimeric G-proteins, comprising , , and subunits, may either induce or suppress effector enzymes or ion channels.

Figure 1-2 (Figure Not Available) Hormonal signaling by cell-surface and intracellular receptors. The receptors for the watersoluble polypeptide hormones, LH, and IGF-I; are integral membrane proteins located at the cell surface. They bind the hormone-utilizing extracellular sequences and transduce a signal by the generation of second messengers, cAMP for the LH receptor, and tyrosine-phosphorylated substrates for the IGF-I receptor. Although effects on gene expression are indicated, direct effects on cellular proteins, for example, ion channels, are also observed. In contrast, the receptor for the lipophilic steroid hormone progesterone resides in the cell nucleus. It binds the hormone and becomes activated and capable of directly modulating target gene transcription. (TF = transcription factor; R = receptor molecule.) (Reproduced from Mayo K. In Conn PM, Melmed S (eds). Endocrinology: Basic and Clinical Principles. Totowa, NJ, Humana Press, 1997, p. 11.)

Several growth factors and hormone receptors (e.g., for insulin) behave as intrinsic tyrosine kinases or activate intracellular protein tyrosine kinases. Ligand activation may cause receptor dimerization (e.g., growth hormone [GH]) or heterodimerization (e.g., interleukin-6 [IL-6]), followed by activation of intracellular phosphorylation cascades. These activated proteins ultimately determine specific nuclear gene expression.

Both the number of receptors expressed per cell and their responses are also regulated, thus providing a further level of control for hormone action. Several mechanisms account for altered receptor function. Receptor endocytosis causes internalization of cell surface receptors; the hormonereceptor complex is subsequently dissociated, resulting in abrogation of the hormone signal. Receptor trafficking may then result in recycling back to the cell surface (e.g., as for insulin), or the internalized receptor may undergo lysosomal degradation. Both these mechanisms triggered by activation of receptors effectively lead to impaired hormone signaling by down-regulation of these receptors. The hormone signaling pathway may also be down-regulated by receptor desensitization (e.g., as for epinephrine); ligand-mediated receptor phosphorylation leads to a reversible deactivation of the receptor. Desensitization mechanisms can be activated by a receptor’s ligand (homologous desensitization) or by another signal (heterologous desensitization), thereby attenuating receptor signaling in the continued presence of ligand. Receptor function may also be limited by the action of specific phosphatases (e.g., Src homology 2 domain-containing protein tyrosine phosphatase [SHP]) or by intracellular negative regulation of the signaling cascade (e.g., suppressor of cytokine signaling [SOCS] proteins inhibiting Janus kinase [JAK]-STAT signaling).

Mutational changes in receptor structure can also determine hormone action. Constitutive receptor activation may be induced by activating mutations (e.g., TSH receptor), leading to endocrine organ hyperfunction, even in the absence of hormone. Conversely, inactivating receptor mutations may lead to endocrine hypofunction (e.g., testosterone or vasopressin receptors). These syndromes are now well characterized and are well described in this volume (Fig. 1-3).

The functional diversity of receptor signaling also results in overlapping or redundant intracellular pathways. For example, both GH as well as cytokines activate JAK-STAT signaling, whereas the distal effects of these stimuli clearly differ. Thus, despite common signaling pathways, hormones elicit highly specific cellular effects. Tissue or cell-type genetic programs or receptorreceptor interactions at the cell surface (e.g., dopamine D2 with somatostatin receptor hetero-oligization) may also confer specific cellular response to a hormone and provide an additive cellular effect.
CONTROL OF HORMONE SECRETION

Anatomically distinct endocrine glands are composed of highly differentiated cells that synthesize, store, and secrete hormones. Circulating hormone concentrations are a function of glanular secretory patterns and hormone clearance rates. Hormone secretion is tightly regulated to attain circulating levels that are most conducive to elicit the appropriate target tissue response. For example, longitudinal bone growth is initiated and maintained by exquisitely regulated levels of circulating hormones, whereas mild GH hyperssecretion results in gigantism and GH deficiency causes growth retardation. Ambient circulating hormone concentrations are not uniform, and secretion patterns determine appropriate physiologic function. Thus, insulin secretion occurs in short pulses elicited by nutrient and other signals and gonadotrophin secretion is episodic, determined by a hypothalamic pulse generator, whereas prolactin secretion appears to be relatively continuous, with secretory peaks elicited during suckling.

Hormone secretion also adheres to rhythmic patterns. Circadian rhythms serve as adaptive responses to environmental signals and are controlled by a circadian timing mechanism. Light is the major environmental cue adjusting the endogenous clock. The retinohypothalamic tract entrains circadian pulse generators situated within hypothalamic suprachiasmatic nuclei. These signals subserve timing mechanisms for the sleep-wake cycle and determine patterns of hormone secretion and action. Disturbed circadian timing results in hormonal dysfunction and may also be reflective of entrainment or pulse generator lesions. For example, adult GH deficiency due to a damaged hypothalamus or pituitary is associated with elevations in integrated 24-hour leptin concentrations, decreased leptin pulsatility, and yet preserved circadian rhythm of leptin. GH replacement restores leptin pulsatility, followed by loss of body fat mass. Sleep is also an important cue regulating hormone pulsatility. About 70% of overall GH secretion occurs during sleep, and increasing age is associated with declining slow-wave sleep and concomitant decline in GH and elevation of cortisol secretion. Most pituitary hormones are secreted in a circadian (day-night) rhythm, best exemplified by ACTH peaks before 9 AM, whereas ovarian steroids follow a 28-day menstrual rhythm. Disrupted episodic rhythms are often a hallmark of endocrine dysfunction. Thus, loss of circadian ACTH secretion with high midnight cortisol levels is a feature of Cushing's disease.

Hormone secretion is induced by multiple specific biochemical and neural signals. Integration of these stimuli results in the net temporal and quantitative secretion of the hormone (Fig. 1-4). Thus, signals elicited by hypothalamic hormones (GHRH, somatostatin), peripheral hormones (iGF-I, sex steroids, thyroid hormone), nutrients, adrenergic pathways, stress, and other neuropeptides, all converge on the somatotroph cell, resulting in the ultimate pattern and quantity of GH secretion. Networks of reciprocal interactions allow for dynamic adaptation and shifts in environmental signals. These regulatory systems embrace the hypothalamic, pituitary, and target endocrine glands, as well as the adipocyte and lymphocyte. Peripheral inflammation and stress elicit cytokine signals, which interface with the neuroendocrine system, resulting in hypothalamic-pituitary axis activation. The parathyroid and pancreatic secreting cells are less tightly controlled by the hypothalamus, but their functions are tightly regulated by the effects they elicit. Thus, parathyroid hormone (PTH) secretion is induced when serum calcium levels fall, and the signal for sustained PTH secretion is abrogated by rising calcium levels.

Several tiers of control subserve the ultimate net glandular secretion. First, central nervous system signals including stress, afferent stimuli, and neuropeptides signal the synthesis and secretion of hypothalamic hormones and neuropeptide (Fig. 1-5). Four hypothalamic releasing hormones (GHRH, corticotropin-releasing hormone [CRH], TRH, and gonadotrophin releasing hormone [GnRH]) traverse the hypothalamic portal vessels and impinge on their respective transmembrane tropic hormone-secreting cell receptors. These distinct cells express GH, ACTH, TSH, and gonadotrophins. In contrast, hypothalamic somatostatin and dopamine suppress GH, prolactin (PRL), or TSH secretion. Trophic hormones also maintain the structural-functional integrity of endocrine organs, including the thyroid and adrenal glands, and the gonads. Target hormones, in turn, serve as powerful negative feedback regulators of their respective tropic hormone and often also suppress secretion of hypothalamic releasing hormones. In certain circumstances, for example during puberty, peripheral sex steroids may positively induce the hypothalamic-pituitary-gland axis. Thus, luteinizing hormone (LH) induces ovarian estrogen.
A further level of secretion control occurs within the gland itself. Thus, intraglandular paracrine or autocrine growth peptides serve to autoregulate pituitary hormone secretion, as exemplified by epidermal growth factor (EGF) control of prolactin or IGF-I control of GH secretion. Molecules within the endocrine cell may also subserves an intracellular feedback loop. Thus, corticotrope SOCS-3 induction by gp 130-linked cytokines serves to abrogate the ligand-induced JAK-STAT cascade and to block pro-opiomelanocortin gene transcription and ACTH secretion. This rapid on-off regulation of ACTH secretion provides a plastic endocrine response to changes in environmental signaling and serves to maintain homeostatic integrity.\(^1\)

In addition to the central-neuroendocrine interface mediated by hypothalamic chemical signal transduction, the CNS directly controls several hormonal secretory processes. Posterior pituitary hormone secretion occurs as direct efferent neural extensions. Postganglionic sympathetic nerves also regulate rapid changes in renin, insulin, and glucagon secretion, and preganglionic sympathetic nerves signal to adrenal medullary cells, eliciting adrenaline release.
HORMONE MEASUREMENT

Endocrine function can be assessed by measuring levels of basal circulating hormone, evoked or suppressed hormone, or hormone-binding proteins. Alternatively, peripheral hormone receptor function can be assessed. Meaningful strategies for timing hormonal measurements vary from system to system. In some cases, circulating hormone concentrations can be measured in randomly collected serum samples. This measurement, when standardized for fasting, environmental stress, age, and gender, is reflective of true hormone concentrations only when levels do not fluctuate appreciably. For example, thyroid hormone, prolactin, and IGF-I levels can be accurately assessed in fasting morning serum samples. On the other hand, when hormone secretion is clearly episodic, timed samples may be required over a defined time course to reflect hormone bioavailability. Thus, early morning and late evening cortisol measurements are most appropriate. Although 24-hour sampling for GH measurements, with samples collected every 2, 10, or 20 minutes, are expensive and cumbersome, they may yield valuable diagnostic information. Random sampling may also reflect secretion peaks or nadirs, thus confounding adequate interpretation of results.

In general, confirmation of failed glandular function is made by attempting to evoke hormone secretion by recognized stimuli. Thus, testing of pituitary hormone reserve may be accomplished by injecting appropriate hypothalamic releasing hormones. Injection of tropic hormones, including TSH and ACTH, evokes specific target gland hormone secretion. Pharmacologic stimuli, for example metoclopramide for induction of prolactin secretion, may also be useful tests of hormone reserve. In contrast, hormone hypersecretion can be diagnosed by suppressing glandular function. Thus, failure to appropriately suppress GH levels after a standardized glucose load implies inappropriate GH hypersecretion.

Radioimmunoassays utilize highly specific antibodies unique to the hormone, or a hormone fragment, to quantify hormone levels. Enzyme-linked immunoabsorbent assays (ELISA) employ enzymes instead of radioactive hormone markers, and enzyme activity is reflective of hormone concentration. This sensitive technique has allowed ultrasensitive measurements of physiologic hormone concentrations. Hormone-specific receptors may be employed in place of the antibody in a radioreceptor assay.
ENOCRINE DISEASES

Endocrine diseases fall into four broad categories: (1) hormone overproduction; (2) hormone underproduction; (3) altered tissue responses to hormones; and (4) tumors of endocrine glands.

Hormone Overproduction

Occasionally, hormones are secreted in increased amounts because of genetic abnormalities that cause abnormal regulation of hormone synthesis or release. In glucocorticoid-remediable hyperaldosteronism, for example, an abnormal chromosomal crossing-over event puts the aldosterone synthetase gene under the control of the ACTH-regulated 11α-hydroxylase gene. More often, diseases of hormone overproduction are associated with an increase in the total number of hormone-producing cells. For example, the hyperthyroidism of Graves' disease, in which antibodies mimic TSH and activate the TSH receptors on thyroid cells, is associated with dramatic increase in thyroid cell proliferation, as well as with increased synthesis and release of thyroid hormone from each thyroid cell. In this example, the increase in thyroid cell number represents a polyclonal expansion of thyroid cells, in which large numbers of thyroid cells proliferate in response to an abnormal stimulus. Most endocrine tumors are not polyclonal expansions, however, but instead represent monoclonal expansions of one mutated cell. Pituitary and parathyroid tumors, for example, are usually monoclonal expansions in which somatic mutations occur in multiple tumor suppressor genes and proto-oncogenes. These mutations lead to an increase in proliferation and/or survival of the mutant cells. Sometimes this proliferation is associated with abnormal secretion of hormone from each tumor cell as well. For example, mutant G proteins in somatotrophs can lead to both increased cellular proliferation and increased secretion of growth hormone from each tumor cell.
Hormone Underproduction

Underproduction of hormone can result from a wide variety of processes, ranging from surgical removal of parathyroid glands during neck surgery, to tuberculous destruction of adrenal glands, or to iron deposition in β-cells in hemochromatosis. A frequent cause of destruction of hormone-producing cells is autoimmunity. Autoimmune destruction of beta cells in type 1 diabetes mellitus and autoimmune destruction of thyroid cells in Hashimoto’s thyroiditis are two of the most common disorders treated by endocrinologists. More uncommonly, a host of genetic abnormalities can also lead to decreased hormone production. These disorders can result from abnormal development of hormone-producing cells (e.g., hypogonadotrophic hypogonadism caused by KAL gene mutations), from abnormal synthesis of hormones (e.g., deletion of the growth hormone gene), or from abnormal regulation of hormone secretion (e.g., the hypoparathyroidism associated with activating mutations of the parathyroid cell’s calcium-sensing receptor).
Altered Tissue Responses

Resistance to hormones can be caused by a variety of genetic disorders. Examples include mutations in the growth hormone receptor in Laron dwarfism and mutations in the G gene in the hypoparathyroidism of pseudohypoparathyroidism, type 1a. The insulin resistance in muscle and liver central to the etiology of type 2 diabetes mellitus appears to be polygenic in origin. Type 2 diabetes is also an example of a disease in which end organ insensitivity is worsened by signals from other organs, in this case by signals originating in fat cells. In other cases, the target organ of hormone action is more directly abnormal, as in the parathyroid hormone (PTH) resistance of renal failure.

Increased end organ function can be caused by mutations in signal reception and propagation. For example, activating mutations in TSH, LH, and PTH receptors can cause increased activity of thyroid cells, Leydig cells, and osteoblasts, even in the absence of ligand. Similarly, activating mutations in the $G_s$ protein can cause precocious puberty, hyperthyroidism, and acromegaly in McCune-Albright syndrome.
Tumors of Endocrine Glands

Tumors of endocrine glands, as noted above, often result in hormone overproduction. Some tumors of endocrine glands produce little if any hormone but cause disease by their local compressive symptoms or by metastatic spread. Examples include so-called nonfunctioning pituitary tumors, which are usually benign but can cause a variety of symptoms due to compression on adjacent structures, and thyroid cancer, which can spread throughout the body without causing hyperthyroidism.
THERAPEUTIC STRATEGIES

In general, hormones are employed pharmacologically for both their replacement and their suppressive effects. Hormones may also be used for diagnostic stimulatory effects (e.g., hypothalamic hormones) to evoke target organ responses or to diagnose endocrine hyperfunction by suppressing hormone hypersecretion (e.g., T_{3}). Ablation of endocrine gland function due to genetic or acquired causes can be restored by hormone replacement therapy. In general, steroid and thyroid hormones are replaced orally, whereas peptide hormones (e.g., insulin, DH) require injection. Gastrointestinal absorption and first-pass kinetics determine oral hormone dosage and availability. Physiologic replacement can achieve both appropriate hormone levels (e.g., thyroid), as well as approximate hormone secretory patterns (e.g., GnRH delivered intermittently via a pump). Hormones can also be used to treat diseases associated with glandular hyperfunction. Long-acting depot preparations of somatostatin analogs suppress GH hypersecretion in acromegaly or 5-HIAA hypersecretion in carcinoid syndrome. Estrogen receptor antagonists (e.g., tarmoxifen) are useful for some patients with breast cancer, and GnRH analogs may downregulate the gonadotrophin axis and benefit patients with prostate cancer.

Novel formulations of receptor-specific hormone ligands are now being clinically developed (e.g., estrogen agonists/antagonists, somatostatin receptor subtype ligands), resulting in more selective therapeutic targeting. Modes of hormone injection (e.g., for PTH) may also determine therapeutic specificity and efficacy. Improved hormone delivery systems, including computerized minipumps, intranasal sprays (e.g., for 1-desamino-8-D-arginine vasopression [DDAVP]), pulmonary inhalations, and depot intramuscular injections, will also allow added patient compliance and ease of administration.

Despite this tremendous progress, some therapies, such as insulin delivery to rigorously control blood sugar, still require tremendous patient involvement and await innovative approaches.
References


Chapter 2 - The Endocrine Patient

Daniel D. Federman

A textbook of medicine is inevitably about disease, but the practice of medicine deals with illness, that is, a person experiencing a disease. It is for that reason that the present chapter has been entitled "The Endocrine Patient." It is my intention to lay out the general issues and approaches applicable to caring for patients with endocrine disorders. The topics to be discussed include initial evaluation and the nature of referral, the fact finding required in clinical evaluation, the use of the laboratory and imaging, the formulation of a differential diagnosis, decision making, and management. In each case, the steps are portrayed from the patient's point of view.

It is worth noting that, except for acute adrenal insufficiency, endocrine disorders are seldom life-threatening. They have enormous effect on the quality of life, however, and successful intervention can be extremely important to both patient and family.
GENERAL CONSIDERATIONS

Many features of being an endocrine patient are common to all experiences of illness. Most often, a perceived change in bodily function, a symptom, gets one to the doctor. Although generations of medical students have described new patients as being “in no acute distress,” most patients are, in fact, worried and anxious when they see a physician, the more so when the physician is not known to them. A few minutes spent in getting to know the patient can pay enormous dividends in the accuracy of the history obtained and in setting the stage for further cooperation with testing and treatment.

Inasmuch as most endocrine consultation is elective rather than emergent, I favor asking a few simple questions, such as “Where are you from?” “What do you do?” “How did you come to us?” “Were you referred?” and so on. Almost always, some common experience or acquaintance is discovered that provides the basis for a rapport that does not emerge from formal medical questioning. This step also immediately conveys that you are interested in the patient as a person and not just as a disease.
SPECIAL FEATURES OF ENDOCRINE ILLNESS

Discovery through Screening

Numerous special features of endocrine disease make patient presentation quite different from that seen in general medicine. One is the discovery of abnormality through screening of asymptomatic individuals, for example, a high serum calcium level discovered through multiphasic screening or a high blood glucose level discovered in a shopping mall kiosk. The very absence of symptoms lends an unreality to the moment and should become an explicit topic of the patient-doctor interaction. In this circumstance, it is worth emphasizing the value of early discovery and prevention of greater morbidity.
Quantitative Rather Than Qualitative Abnormalities

A second special feature of endocrine disorders is that they are all quantitative, rather than qualitative, departures from normal. No endocrine disorder is due to a novel hormone. Everyone has cortisol circulating as a determining feature of his or her life. Hypercorticism and adrenal insufficiency represent just more or less of the hormone. Similarly, all hormones found in excess or in deficiency in disease are physiologic determinants of stature, weight, complexion, hairiness, temperament, and behavior. In contrast, no one has a little pneumonia or a little inflammatory bowel disease as a constitutive status. In addition, most endocrine glands have both a basal and a stimulable or reserve function. It is common to have partial diminution of capacity in which the basal function is adequate but a reserve called upon during part of each day or, more dramatically, in emergencies is not available.
Overlap with Other Diseases

The symptoms of endocrine disorders overlap a great range of normal characteristics, including body contour, facial configurations, weight distributions, skin and hair coloring, and muscular capacity. They also overlap with other conditions that are far more common, including depression and normal aging. The added adipose tissue of hyperadrenocorticism is more difficult to recognize in a person who is already obese. The nervousness associated with hyperthyroidism is less apparent in a thin, hyperkinetic man than in a person of moderate body weight. The effects of an androgen-producing adrenal tumor are less likely to be noticed in a family of swarthy, hirsute individuals.

Finally, most endocrine disorders evolve gradually over months to years instead of appearing suddenly, such as a heart attack or an acute infection. This combination of varied host background and slow evolution of disease leads to considerable delay in diagnosis: both the patient and primary care physician adapt to the changes as part of the person, and definitive evaluation, now relatively easy for most disorders, is not undertaken. Hypothyroidism and acromegaly are good examples of this phenomenon. All series show a remarkable delay in diagnosis despite sometimes disabling symptoms.

Hormones have more distant effects than local effects. This, of course, reflects their messenger status. Unlike an abscess, a myocardial infarction, or an esophageal cancer, endocrine disorders seldom produce symptoms near the gland of origin. (Subacute thyroiditis and large pituitary tumors, of course, are exceptions.) But because in most endocrinopathies the excess or missing hormone works on several or many systems, the resulting syndrome can be enigmatic.

Several endocrine disorders are important not because of their incidence but because of their curability: Cushing's disease, acromegaly, and pheochromocytoma are cases in point. Although these disorders enter the differential diagnosis of common problems such as diabetes, their occurrence is so rare that the primary care physician does not easily think of them.
Unique Features of Reproductive Disorders

Reproductive disorders have symptoms and signs that have no parallel in other areas. This is the one system in which sexual dimorphism is inherent rather than epiphenomenal; it is also the one with the greatest span of developmental change. Once the heart starts beating in the embryo, it goes on doing so until the last moment of life; but puberty, adult sexual functioning, and menopause establish time lines against which all symptoms are to be assessed. Thus, vaginal bleeding has entirely different meanings whether it occurs on the first day of life, as a natural appearance at age 12, between menstrual periods at age 25 years, as a harbinger of menopause at age 46 years, or as a highly probable symptom of cancer at age 66 years.

Physical appearance and function are important features of self-image. Thus, hirsutism, thinness, obesity, sexual arousal, and erectile capacity bear considerable psychological import to the endocrine patient. The clinician should be constantly aware of both spoken and unspoken thoughts that may be troubling the patient.

The Couple as a Clinical Unit

The ultimate goal of reproductive capacity is, of course, a fertile union. This means that the couple, rather than the individual, is the unit of clinical concern. It is thus the principal area in medicine whereby two people and their interaction, rather than a single person and her capacities, are studied and treated. In addition, there are dimensions of successful sexual function that are important at other times than when fertility is sought. Sex drive, erotic responsiveness, affection, and tenderness are all important aspects of life whether or not fertility is an issue. These areas are notoriously difficult to evaluate in a society that treats sexual function as such a special topic, particularly enshrouded in personal issues of such importance.
EVALUATION OF PATIENTS WITH ENDOCRINE DISORDERS

I have emphasized previously the belief that establishing an interested and warm relationship is the beginning of excellence in any elective medical interaction. In addition to its affective power, the relationship elicits a more informative history, establishes better cooperation in both testing and treatment, and provides a platform for informed decision making by the patient.

History

As in most areas of medicine, precision of diagnosis and economy of investigation begin with a carefully wrought history. An open-ended question, combined with an attentive silence, allows the patient to provide the background for the clinical moment. After the patient has spoken spontaneously, the physician provides a guided expansion of the information. Details of timing, sequence, changes of diet or activity, relationship to the menstrual cycle, changes in weight or size, and alterations in mood or sleep pattern all of these may provide clues to underlying endocrine abnormality.

A good example of the power of the history is the interpretation of irregular periods in a woman of reproductive age. The simple statement, “I’ve never been regular,” points to a presumptive diagnosis of polycystic ovary syndrome in a way that a very convoluted sequence of questions might actually fail to do. That statement is to be contrasted with this one: “I used to be regular, but in the last year or so, I never know when my period is going to come.” If the presenting symptom is irregular periods, the simple invitation, “Tell me about your periods,” is likely to be the key to the diagnosis.

Careful questioning about use of complementary and alternative medicines is an important and, occasionally, a very revealing step.

A thorough family history has become increasingly important as the genetic basis for more and more endocrine diseases becomes established. For practical purposes, I favor diagramming a pedigree of the first-order relatives—parents, siblings, children—of all patients, not just those for whom a genetic disorder is already suspected. Known disorders are readily revealed this way, and unknown conjunctions of clinical and genetic factors may also be disclosed.
Physical Examination

General Examination

It is said that the history is 80% or more of clinical diagnosis, and that is no less true in endocrine disorders than in general medicine. Yet the physical examination is a critical element in the process of arriving at a diagnosis, and here I want to call particular attention to the first impression.

The possibility of Cushing’s syndrome, Addison’s disease, hyperthyroidism and hypothyroidism, acromegaly, polycystic ovary syndrome, hypogonadism, and Turner’s syndrome and other endocrine disorders should be considered from the first moment one encounters a new patient. Otherwise, one risks accepting that the appearance of the patient is just that and no more. In other words, as soon as one accepts that the initial impression is what the person looks like naturally, the quantitative departure from normal that is the essence of endocrine disease fails to impress one. This, incidentally, is why both families and primary care physicians often miss a diagnosis that seems obvious to the consultant endocrinologist.

A quantification of this last point may be helpful. If the signs of hypothyroidism or acromegaly, for example, take 3 years to become striking, the person living with the patient is exposed to 1/1095th of fractional change per day, well below the threshold of just noticeable difference. Similarly, a primary care physician seeing the patient perhaps four times a year for a general checkup and management of hypertension is exposed to 91/1095 fractional change. This can sometimes lead to a diagnosis but often does not. When one sees the patient for the first time, however, the imprint of the disease catches attention and the constitutional appearance is in the background.

Although a consultant participates because of a special area of interest and expertise, he or she is a general physician first and should be alert to all dimensions of the physical examination:

- What is the height/weight ratio?
- What is the basic degree of muscularity?
- Is there evidence of heart disease to explain the chest pain and dyspnea one has heard about in the history?
- What is the degree of hirsutism?
- Are there signs of liver disease, malnutrition, or poor or excellent physical training?
- What is the blood pressure with the patient standing as well as lying or sitting?

These and many other points of a general examination begin to modify the thinking one has undertaken on the basis of history.

Targeted Examination

The targeted physical examination of any consultant is an interesting interplay of general and specific goals. Theoretically, any experienced clinician should undertake a general examination and come to all the findings pertinent to an underlying endocrine disorder. In fact, however, the physical examination is greatly influenced by the hypotheses generated in the history. Let us look at a few examples.

If a patient reports weight loss despite a good appetite, there is only a very restricted differential diagnosis, principally malabsorption or hypermetabolism. In doing a physical examination, therefore, I would pay particular attention to signs of malabsorption (muscular wasting, vitamin deficiencies, purpura) and to signs of thyroid disease with its generalized hypermetabolism and localized autoimmune phenomena, including ophthalmopathy.

Similarly, if a patient complains of hirsutism or other signs of androgen excess, one is immediately thrust into a consideration of ethnic hair distribution and quality. Is there temporal recession of the forehead? Does the hair on the abdomen come up over the umbilicus? Is hair present on the back (rare without marked hyperandrogenism)? How much acne is there? At the extreme, is there evidence of clitoral enlargement?

Finally, and most important, does the patient look like or unlike the other women in her family?

Direct Assessment of Endocrine Glands

Three endocrine glands are palpable: the thyroid, the testis, and the ovary. Specific attention should be given to each of these.

The thyroid gland should be approached first by inspection while the patient swallows for size, symmetry, or localized enlargement. Many thyroid nodules are visible, and inspection often calls attention to lesions that would be missed on palpation. The thyroid should then be felt while the patient swallows, from the front with your thumbs or from behind the patient with the index and third fingers. It is crucial to keep your own fingers from moving while the patient is swallowing. The principal observation is whether there is diffuse enlargement of the thyroid gland (most often Graves’ diffuse hyperplasia or Hashimoto’s thyroiditis) or one or more nodules. Although the consistency of the gland is to be noted, in fact it is often not concordant with the pathology.

Functioning tumors of the testis may be too small to be felt with the fingers, and most internists and general physicians are not skilled in palpation of the ovaries. For this reason, ultrasound and other forms of imaging have become key features of gonadal evaluation and are discussed later.

The size of one other endocrine gland, the pituitary, can be inferred from physical examination for what Cushing called “neighborhood signs.” As a pituitary tumor or adenoma increases in size, it pushes up on the optic apparatus, producing visual field changes and often to a degree to be easily detected. Many consequences of hormone action can be detected on physical examination; the results combine with the history to produce a highly reliable differential diagnosis and thus an informed basis for laboratory evaluation and imaging. Among the things to be looked for are the eye signs and dermopathy of Graves’ disease, acanthosis nigricans as a clue to insulin resistance, muscular wasting and tremor, changes in the voice due to hypothyroidism or acromegaly, and a general impression of nutrition and its adequacy or excess. Each of these findings is described in more detail with the specific disorder in subsequent chapters.

Indirect Assessment of Endocrine Status

Many consequences of hormone action can be detected on physical examination; the results combine with the history to produce a highly reliable differential diagnosis and thus an informed basis for laboratory evaluation and imaging. Among the things to be looked for are the eye signs and dermopathy of Graves’ disease, acanthosis nigricans as a clue to insulin resistance, muscular wasting and tremor, changes in the voice due to hypothyroidism or acromegaly, and a general impression of nutrition and its adequacy or excess. Each of these findings is described in more detail with the specific disorder in subsequent chapters.
LABORATORY TESTING OF ENDOCRINE FUNCTION

Modern endocrine laboratory evaluation began with the introduction of radioimmunoassay by Berson and Yalow. The precise measurement of hormone concentrations, determined by competitive displacement of specific antibodies, was soon succeeded by competitive binding assays and, more recently, by immunofluorescent and radioluminescent determinations of even greater sensitivity and specificity. It should therefore be possible to enter the name of a hormone on a laboratory slip and expect to get back a definitive reflection of the status of the patient for that gland.

For practical purposes, that has become true of thyroid-stimulating hormone (TSH). Reliable determinations of elevated, normal, and suppressed levels of this hormone by radioor chemiluminescent determination have made it the standard of care for thyroid disease and a model for all endocrine laboratory tests. However, it is an exception rather than the rule, and it is worthwhile reviewing why other testing is not as easy and why considerable judgment is required. The following examples illustrate this point.

Pulsatile Hormone Secretion

Many hormones are secreted in pulses rather than steadily. The peaks or valleys of hormones secreted in pulsatile fashion, such as luteinizing hormone or growth hormone, may fall above or below the ostensibly normal range. If such a value is obtained by chance, it can erroneously suggest hypofunction or hyperfunction. Repeating the test with three samples drawn at 30-minute intervals and pooled can clarify this type of problem.
Diurnal Variation

The hypothalamic-pituitary-adrenal axis of cortisol secretion is typically maximal during the day and lower in the evening and night. A plasma cortisol level of 12 µg/dL is normal at 8 AM, but the same value at 8 PM reflects a loss of diurnal rhythm resulting from either stress or hypercorticism. A plasma cortisol sample drawn at midnight is an excellent test for evaluation of overactive adrenal function.
Cyclic Variation

The menstrual cycle provides the most extreme "normal variation" of any hormone level. From the first day of a menstrual period, when estrogen levels may be indistinguishable from those of a normal man, the level rises extraordinarily rapidly and at the 14th day can be as high as in early pregnancy. As a consequence, an estrogen level must be evaluated in the light of the stage of the cycle at which it is drawn.
Age

All clinicians are aware that gonadal hormones show marked differences reflective of the individual's stage of life. It is not as widely known that the adrenal hormone dehydroepiandrosterone (DHEA) is barely secreted during childhood, is actively put out by the adrenal glands from age 8 or so to age 55, and then disappears as mysteriously as it came. At present, there is no clear understanding of the physiologic role of its presence or absence.
Sleep Entrainment

Both prolactin and growth hormone have a sleep-entrained secretory pulse shortly after sleep begins. In people who work at night and sleep during the day, this secretion is clearly related to sleep and not to clock time.
Hormone Antagonism

Certain hormones antagonize the effects of other hormones; it is thus necessary to know the value of each hormone to interpret the clinical phenomenon. The opposite effects of estrogen and androgen on the male breast are a good example. A normal testosterone level combined with an elevated estrogen level, or a normal estrogen level but a decreased androgen level, easily accounts for gynecomastia.
Dynamic Testing

Many endocrine glands have a basal secretory level and a reserve secretion elicited by either a tropic hormone or a change in metabolic or physiologic state. Cortisol secretion can increase fivefold to 10-fold in response to stress or adrenocorticotropic hormone (ACTH). Insulin release is stimulated by both glucose and amino acids and by distinct pathways.

Baseline hormone levels can be misleading. The test results in Table 2-1 were obtained on a 30-year-old woman who complained of fatigue and amenorrhea 6 months after a pregnancy during which she had been markedly anemic (hemoglobin, 9 g/dL); she had never been in shock and had received no transfusions.
Hormone and Metabolite Interaction

Insulin is a good example of a hormone whose absolute level is less meaningful than its relationship to the blood glucose level. A plasma insulin of 70 is a normal response to a meal, when the blood glucose level is rising. In contrast, an insulin value of 10 or 12 is abnormal (is not appropriately suppressed) if the glucose level is 40 mg/dL. Indeed, the lower insulin level in a hypoglycemic patient is distinct evidence of spontaneous hyperinsulinism, such as in an islet cell tumor.

Growth hormone represents another instance in which a single random sample cannot be given much meaning. During a day, plasma growth hormone levels vary from values that, if sustained, would be diagnostic of acromegaly to values that, again if sustained, would point to hypopituitarism. In normal people, growth hormone secretion is suppressed by glucose intake. A plasma growth hormone of 8 within an hour of a standard meal containing glucose would be pathologically elevated; it should be less than 2. Similarly, however, a growth hormone value of 2 in a fasting person who had run up a flight of stairs suggests deficient pituitary function.

Most hormones are part of a feedback loop in which an artificial increase, especially by ingestion of the hormone in a medication, decreases endogenous secretion. If a normal person takes 0.1 to 0.3 mg of thyroxine (T₄), hypothalamic secretion of thyrotropin-releasing hormone (TRH) and pituitary secretion of thyrotropin (TSH, or thyroid-stimulating hormone) are suppressed. Plasma levels of T₄ and triiodothyronine (T₃) may not change, but TSH levels would be decreased and reflections of TSH effect, such as radioactive iodine uptake, would similarly be suppressed. Although ultrasensitive TSH testing has replaced tests of suppressibility for the diagnosis of thyroid disease, tests of suppressibility are the standard approach to evaluating growth hormone and ACTH/cortisol regulation, respectively.
Protein Binding

Hormones such as T₄ and cortisol are compartmentalized into a fraction attached to a transport protein (and thus physiologically unavailable) and a free portion able to diffuse into cells and initiate a hormone effect. It is the free or unbound portion that is physiologically regulated; the level of the binding protein may be increased or decreased without physiologic consequence if the free portion is unchanged. The measurement of free T₄ or a free T₄ index (FT₄ I) (see Chapter 10) has become the standard second step if a screening TSH value is abnormally high or low.

Testosterone is even more complicated because it is trebly partitioned among sex hormone-binding globulin, albumin, and a free portion. Measurement of the free hormone level is often necessary, particularly when the binding protein level has been artifactually raised or lowered (see Chapter 6).

### TABLE 2-1 – Baseline Hormonal Values

<table>
<thead>
<tr>
<th>Time</th>
<th>Glucose</th>
<th>Cortisol</th>
<th>TSH</th>
<th>Prolactin</th>
<th>FSH</th>
<th>LH</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>80 mg/dL</td>
<td>7.7 µg/dL</td>
<td>3.2 mIU/L</td>
<td>6.6 ng/mL</td>
<td>8.9 mIU/mL</td>
<td>6.7 mIU/mL</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Insulin IV</td>
<td>TRH IV</td>
<td>GnRH IV</td>
</tr>
<tr>
<td>15</td>
<td>47</td>
<td>8.2</td>
<td>5.9</td>
<td></td>
<td>9.3</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>23</td>
<td>8.6</td>
<td>5.6</td>
<td>7.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>30</td>
<td>7.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>38</td>
<td>13.0</td>
<td></td>
<td>14.3</td>
<td>13.0</td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>50</td>
<td>9.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>63</td>
<td>9.6</td>
<td>15.4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*In this study, all basal values are within normal limits. Note, however, that intravenous insulin lowers the blood glucose levels but does not elicit an adequate release of cortisol. Thyrotropin-releasing hormone (TRH) does not induce a normal rise in thyrotropin (TSH) or prolactin. Gonadotropin-releasing hormone (GnRH) evokes a submaximal increase in follicle-stimulating hormone (FSH) and luteinizing hormone (LH).*
Laboratory Error

Laboratory error may seem too obvious a source of confusion to mention, but it provides a reminder for an important caution about laboratory testing. It is easy to be seduced by numbers and to consider the laboratory report the final arbiter. In fact, it is the history and physical examination, plus the clinician’s judgment, that establish the prior probability of a given diagnosis. Both in choosing and in interpreting laboratory tests, the endocrinologist should establish his or her own expectations before testing. If the physician feels strongly that a particular condition is present, discordant initial laboratory results should not be dissuasive. More detailed testing, as discussed in subsequent chapters, is then appropriate. The clinician’s judgment is still a key component of the process.
IMAGING

The extraordinary power of modern imaging, particularly ultrasound, computed tomography (CT), and magnetic resonance imaging (MRI), has enriched endocrinology as it has all of medicine. However, the role of imaging in endocrinology is, to my mind, different from its contribution elsewhere.

For one thing, several endocrine glands (the thyroid, the pituitary, and the adrenals in particular) frequently contain clinically insignificant, nonfunctioning adenomas and cysts. Second, functioning and nonfunctioning lesions other than in the thyroid gland can be very difficult to distinguish from each other. Thus, except in an emergency (e.g., suspected pituitary apoplexy), the clinician should define the functional state of the gland before requesting imaging. In other words, one should be clear from hormone measurements, including dynamic testing, whether the gland is overactive, underactive, or normal.

In addition, one should have a clear idea of how the radiologist can be expected to help. Such clarity reduces costs by targeting the selection of imaging and making the radiologic findings a truly complementary element of evaluation. The best imaging modalities for the various glands are discussed in their respective chapters. The approach suggested here, however, is broadly applicable.
CONVEYING RESULTS

Both sophisticated imaging and thorough laboratory testing produce results after the actual office visit. Patients are understandably anxious about the findings and deserve a prompt response. The best approach depends on the circumstance. A new patient with Cushing's syndrome or acromegaly should be given an early in-person visit. A patient with hypothyroidism who understands the disease well and just needs a slight change in $T_4$ dose can easily be informed with a telephone call. Someone with negative results can be left a message of reassurance and can be encouraged to call back, both to confirm receipt of the information and to get questions answered.
SOME NEW FEATURES OF CLINICAL ENDOCRINOLOGY

Genetics

The decoding of the human genome promises to change the face of medical practice. Ironically, the first human disease in which cancer was prevented by application of genetic testing in susceptible families was the screening for medullary thyroid carcinoma in pedigrees of multiple endocrine neoplasia type 2 (MEN-2). The screening at that time was done by pentagastrin or calcium provocation of calcitonin release. Now the screening for endocrine manifestations of MEN-2 is secondary to screening families for the \textit{RET} proto-oncogene defect that is the basis of the disease.

Endocrine testing, such as measurement of calcitonin or plasma catecholamines, is restricted to patients who have the genetic abnormality. In the dangerous variant of the MEN-2 syndrome in which medullary thyroid carcinomas appear in the first year of life, aggressive genetic screening is done during that year, and endocrine testing in patients at risk can justify surgical thyroidectomy before the first birthday. Hereditary predispositions will certainly emerge for other endocrine disorders and will make it crucial for the clinician to take a revealing family history and follow up even minor clues.
The Internet

Never in history has so much medical information been available to patients. I therefore now routinely ask patients what they already know about their condition or their symptoms. A bit sheepishly in some cases, many patients admit to looking up topics on the World Wide Web and are about to compare what I tell them with what they have already read. Much of that information is accurate, but some is nonsense, and it requires patience and clear explanation before such patients go away satisfied.
Electronic Mail

Although opinions differ widely, I find e-mail an extremely useful advance in communicating with patients who have computers. They have access to you between appointments and on a time frame of mutual convenience. The computer thus reduces anxiety on the patient's part, particularly regarding questions or findings for which they might otherwise hesitate to make an appointment. Reporting laboratory test results is expedited, and accompanying the report with a few sentences of interpretation can be as useful as a telephone call.

It is wise to keep copies of e-mails so that a clear record of the exchange is available. There are, however, several important caveats. Never let an e-mail exchange substitute for a true evaluation, including history and physical examination. I believe that one should not prescribe for a patient whom one has not seen, and one should not provide much interpretation of history or laboratory tests without very fundamental disclaimers. On the whole, however, e-mail can be used to initiate a new relationship and can certainly support an ongoing one.
Managed Care

The effort to control health care costs by limiting reimbursement for physician services, laboratory testing, and imaging has had a profound impact throughout medicine. Without taking on the whole issue, I want to comment on several practical consequences.

"Curbsiding" the request by a physician for patient guidance without being asked to see the patient has increased strikingly. Consultants can provide some general help to primary care physicians without seeing the patient; however, much hinges on the history and physical examination done by the primary care physician. The failure to realize that hyperthyroidism is due to a hot nodule, for example, totally distorts the picture and will lead to an erroneous recommendation for treatment. The failure to distinguish a recent onset of amenorrhea and virilization from a polycystic ovary-like syndrome may hide the presence of a readily curable virilizing tumor. The failure to recognize hypoglycemic unresponsiveness may perpetuate a dangerous degree of overinsulinization and elicit inappropriate advice from the consultant. Thus, the consultant must set bounds and at some point indicate that it is important for a formal consultation to take place.
Costs

Some endocrine work-ups can be expensive and invite challenge from third-party insurers. The best approach to this concern is a careful history and physical examination, clear establishment of the prior probabilities of certain diagnoses, and then effective use of screening tests before embarking on an unnecessarily extensive evaluation.

For example, in a patient with suspected Cushing's syndrome, it is mandatory to establish the presence of hypercorticism before embarking on a search for its cause. Once this lethal but curable disorder has been properly diagnosed, however, no cost should deter one from finding the cause and correcting it. One argument I make is that expensive tests and imaging should be amortized rather than considered an extravagant or unnecessary expense. If a young woman age 30, with a life expectancy of 80 years or more, has a husband and two children to whom her life matters, the $3000 evaluation breaks down to $20 per loved one per year of life expectancy. Any plan manager has to see that this is an appropriate cost.
MANAGEMENT

There are few more gratifying experiences in medicine than recognizing and correcting an endocrine disorder. Patients feel that they have been rescued from a mysterious overtaking of their identity. Body contour, facial appearance, temperament, and well-being are restored to the patient's constitutive status. Deterioration attributed to aging or depression or chronic disease is reversed. In brief, something almost miraculous takes place. Even when these goals cannot be achieved, as in diabetes, a major impact on mortality and morbidity can be.

Of course, these optimal outcomes require accuracy of diagnosis but that is only the beginning. A true sharing by patient and physician, based on a sound knowledge of normal physiology, provides the best foundation for choice of therapy and maintenance of a continuing program. The result can be, simply put, wonderful.
Chapter 3 - Genetic Control of Peptide Hormone Formation

Joel F. Habener

Advances in the fields of molecular and cellular biology have provided new insights into the mechanistic workings of cells. Recombinant deoxyribonucleic acid (DNA) technology and the sequencing (decoding) of the entire human and mouse genomes now make it possible to analyze the precise structure and function of DNA, the genetic substance that is the basis for life. The discovery of the unique biochemical and structural properties of DNA provided the conceptual framework with which to begin a systematic investigation of the origins, development, and organization of life forms.

The near completion of the entire sequences of the human and mouse genomes was accomplished in the years 2000 and 2001, 5 years ahead of the originally anticipated schedule. The availability of a complete blueprint of the structure and organization of all expressed genes now provides profound insights into the basis of genetically determined diseases. Within the next decades, genotyping of individuals shortly after birth will be possible. Therapeutic approaches for the correction of genetic defects by techniques of gene replacement are likely to become a reality.

The polypeptide hormones constitute a critically important and diverse set of regulatory molecules encoded by the genome whose functions are to convey specific information among cells and organs. This type of molecular communication arose early in the development of life and evolved into a complex system for the control of growth, development, and reproduction and for the maintenance of metabolic homeostasis. These hormones consist of approximately 400 or more small proteins ranging from as few as three amino acids (thyrotropin-releasing hormone, TRH) to 192 amino acids (growth hormone). In a broader sense, these polypeptides function both as hormones, whose actions on distant organs are mediated by way of their transport through the blood stream, and as local cell-to-cell communicators (Fig. 3-1). The latter function of the polypeptide hormones is exemplified by their elaboration and secretion within neurons of the central, autonomic, and peripheral nervous systems, where they act as neurotransmitters. These multiple modes of expression of the polypeptide hormone genes have aroused great interest in the specific functions of these peptides and the mechanisms of their synthesis and release.

This chapter reviews the diverse structures of genes encoding peptide hormones and the multiple mechanisms that govern their expression. The synthesis of nonpeptide hormones (e.g., catecholamines, thyroid hormones, steroid hormones) involves the action of multiple enzymes and hence the expression of multiple genes and is discussed in the individual chapters devoted to such hormones.
Peptide hormones arose early in the evolution of life. Indeed, polypeptides that are structurally similar to mammalian peptides are present in lower vertebrates, insects, yeasts, and bacteria. An example of the early evolution of regulatory peptides is the factor (matting pheromone) of yeast, which is similar in structure to mammalian luteinizing hormone-releasing hormone (also called gonadotropin-releasing hormone [GnRH]). The oldest member of the cholecystokinin-gastrin family of peptides appeared at least 500 million years ago in the protostome Ciona intestinalis.

Thus, the genes encoding polypeptide hormones, and particularly regulatory peptides, evolved early in the development of life and initially fulfilled the function of cell-to-cell communication to cope with problems concerning nourishment, growth, development, and reproduction. As specialized organs connected by a circulatory system developed during evolution, similar, if not identical, gene products became hormones for purposes of organ-to-organ communication.
STEPS IN EXPRESSION OF A PROTEIN-ENCODING GENE

The steps involved in transfer of information encoded in the polynucleotide language of DNA to the poly-amino acid language of biologically active proteins involve gene transcription, post-transcriptional processing of ribonucleic acids (RNAs), translation, and post-translational processing of the proteins. The expression of genes and protein synthesis can be considered in terms of several major processes, any one or more of which may serve as specific control points in the regulation of gene expression (Fig. 3-2):

1. **Rearrangements and transpositions of DNA segments.** These processes occur over many years (eons) in evolution, with the exception of uncommon mechanisms of somatic gene rearrangements such as the rearrangements in the immunoglobulin genes during the lifetime of an individual.

2. **Transcription.** Synthesis of RNA results in the formation of RNA copies of the two gene alleles and is catalyzed by the basal RNA polymerase II-associated transcription factors.

3. **Post-transcriptional processing.** Specific modifications of the RNA include the formation of messenger RNA (mRNA) from the precursor RNA by way of excision and rejoinder of RNA segments (introns and exons) and modifications of the 3’ end of the RNA by polyadenylation and of the 5’ end by addition of 7-methylguanine “cap.”

4. **Translation.** Amino acids are assembled by base pairing of the nucleotide triplets (anticodons) of the specific “carrier” aminoacylated transfer RNAs to the corresponding codons of the mRNA bound to polyribosomes and are polymerized into the polypeptide chains.

5. **Post-translational processing and modification.** Final steps in protein synthesis may involve one or more cleavages of peptide bonds, which result in the conversion of biosynthetic precursors (prohormones), to intermediate or final forms of the protein: derivation of amino acids (e.g., glycosylation, phosphorylation, acetylation, myristoylation); and the folding of the processed polypeptide chain into its native conformation.

Each of the specific steps of gene expression requires the integration of precise enzymatic and other biochemical reactions. These processes have developed to provide high fidelity in the reproduction of the encoded information and to provide control points for the expression of the specific phenotype of cells.

The post-translational processing of proteins creates diversity in gene expression through modifications of the protein. Although the functional information contained in a protein is ultimately encoded in the primary amino acid sequence, the specific biologic activities are a consequence of the higher order secondary, tertiary, and quaternary structures of the polypeptide. Given the wide range of possible specific modifications of the amino acids, such as glycosylation, phosphorylation, acetylation, and sulfation, any one of which may affect the conformation or function of the protein, a single gene may ultimately encode a wide variety of specific proteins as a result of post-translational processes.

Polypeptide hormones are synthesized in the form of larger precursors that appear to fulfill several functions in biologic systems (Fig. 3-3), including (1) intracellular trafficking, by which the cell distinguishes among specific classes of proteins and directs them to their sites of action, and (2) the generation of multiple biologic activities from a common genetically encoded protein by regulated or cell-specific variations in the post-translational modifications (Fig. 3-4).

All the peptide hormones and regulatory peptides studied thus far contain signal or leader sequences at the amino termini; these hydrophobic sequences recognize specific sites on the membranes of the rough endoplasmic reticulum, which results in the transport of nascent polypeptides into the secretory pathway of the cell (see Fig. 3-2 and Fig. 3-3). The consequence of the specialized signal sequences of the precursor proteins is that proteins destined for secretion are selected from a great many other cellular proteins for sequestration and subsequent packaging into secretory granules and export from the cell. In addition, most, if not all, of the smaller hormones and regulatory peptides are produced as a consequence of post-translational cleavages of the precursors within the Golgi complex of secretory cells.
SUBCELLULAR STRUCTURE OF CELLS THAT SECRETE PROTEIN HORMONES

Cells whose principal functions are the synthesis and export of proteins contain highly developed, specialized subcellular organelles for the translocation of secreted proteins and their packaging into secretory granules. The subcellular pathways utilized in protein secretion have been elucidated largely through the early efforts of Palade and colleagues (reviewed by Jamieson and colleagues (reviewed by Jamieson)). Secretory cells contain an abundance of endoplasmic reticulum, Golgi complexes, and secretory granules (Fig. 3-5). The proteins that are to be secreted from the cells are transferred during their synthesis into these subcellular organelles, which transport the proteins to the plasma membrane.

Protein secretion begins with translation of the mRNA encoding the precursor of the protein on the rough endoplasmic reticulum, which consists of polyribosomes attached to elaborate membranous saccules that contain cavities (cisternae). The newly synthesized, nascent proteins are discharged into the cisternae by transport across the lipid bilayer of the membrane. Within the cisternae of the endoplasmic reticulum, proteins are carried to the Golgi complex by mechanisms that are incompletely understood. The proteins gain access to the Golgi complex either by direct transfer from the cisternae, which are in continuity with the membranous channels of the Golgi complex, or by way of shuttling vesicles known as transition elements (see Fig. 3-5).

Within the Golgi complex, the proteins are packaged into secretory vesicles or secretory granules by their budding from the Golgi stacks in the form of immature granules. Immature granules undergo maturation through condensation of the proteinaceous material and application of a specific coat around the initial Golgi membrane. On receiving the appropriate extracellular stimuli (regulated pathway of secretion), the granules migrate to the cell surface and fuse to become continuous with the plasma membrane, which results in the release of proteins into the extracellular space, a process known as exocytosis.

The second pathway of intracellular transport and secretion involves the transport of proteins contained within secretory vesicles and immature secretory granules (see Fig. 3-5). Although the use of this alternative vesicle-mediated transport pathway remains to be demonstrated conclusively (it is generally considered to be a constitutive, or unregulated, pathway), different extracellular stimuli may modulate hormone secretion differently, depending on the pathway of secretion. For example, in the parathyroid gland and in the pituitary cell line derived from corticotropic cells (AIT-20), newly synthesized hormone is released more rapidly than hormone synthesized earlier. These findings suggest that the newly synthesized hormone may be transported by way of a vesicle-mediated pathway without incorporation into mature storage granules.
INTRACELLULAR SEGREGATION AND TRANSPORT OF POLYPEPTIDE HORMONES

Specific amino acid sequences encoded in the proteins serve as directional signals in the sorting of proteins within subcellular organelles. A typical eukaryotic cell synthesizes an estimated 5000 different proteins during its life span. These different proteins are synthesized by a common pool of polyribosomes. However, each of the different proteins is directed to a specific location within the cell, where its biologic function is expressed. For example, specific groups of proteins are transported into mitochondria, into membranes, into the nucleus, or into other subcellular organelles, where they serve as regulatory proteins, enzymes, or structural proteins. A subset of proteins is specifically designed for export from the cell (e.g., immunoglobulins, serum albumin, blood coagulation factors, and protein and polypeptide hormones).

This process of directional transport of proteins involves sophisticated informational signals. Because the information for these translocation processes must reside either wholly or in part within the primary structure or in the conformational properties of the protein, sequential post-translational modifications may be crucial for determining the specificity of protein function.

Signal Sequences in Peptide Prohormone Processing and Secretion

The early processes of protein secretion that result in the specific transport of exported proteins into the secretory pathway are now becoming better understood. Initial clues to this process came from determinations of the amino acid sequences of the proteins programmed by the cell-free translation of mRNAs encoding secreted polypeptides. Secreted proteins are synthesized as precursors that are extended at their NH₂ termini by sequences of 15 to 30 amino acids, called signal or leader sequences. Signal sequence extensions, or their functional equivalents, are required for targeting the ribosomal or nascent protein to specific membranes and for the vectorial transport of the protein across the membrane of the endoplasmic reticulum. On emergence of the signal sequence from the large ribosomal subunit, the ribosomal complex specifically makes contact with the membrane, which results in translocation of the nascent polypeptide across the endoplasmic reticulum membrane into the cisterna as the first step in the transport of the polypeptide within the secretory pathway. These observations initially left unanswered the question of how specific polyribosomes that translate mRNAs encoding secretory proteins recognize and attach to the endoplasmic reticulum (Fig. 3-6).

Figure 3-5 Schematic representation of subcellular organelles involved in transport and secretion of polypeptide hormones or other secreted proteins within a protein-secreting cell. (1) Synthesis of proteins on polyribosomes attached to endoplasmic reticulum (RER) and vectorial discharge of proteins through the membrane into the cisterna. (2) Formation of shutting vesicles (transition elements) from endoplasmic reticulum followed by their transport to and incorporation by the Golgi complex. (3) Formation of secretory granules in the Golgi complex. (4) Transport of secretory granules to the plasma membrane, fusion with the plasma membrane, and exocytosis resulting in the release of granule contents into the extracellular space. Note that secretion may occur by transport of secretory vesicles and immature granules as well as mature granules. Some granules are taken up and hydrolyzed by lysosomes (catabolism). Golgi, Golgi complex; RER, rough endoplasmic reticulum; SER, smooth endoplasmic reticulum. (From Habener JF. Hormone biosynthesis and secretion. In Felig P, Baxter JD, Broadus AE, et al. [eds]. Endocrinology and Metabolism. New York, McGraw-Hill, 1981, pp 2959. Copyright © 1981 by McGraw-Hill, Inc. Used by permission of McGraw-Hill Book Company.)

from the large ribosomal subunit, the ribosomal complex specifically makes contact with the membrane, which results in translocation of the nascent polypeptide across the endoplasmic reticulum membrane into the cisterna as the first step in the transport of the polypeptide within the secretory pathway. These observations initially left unanswered the question of how specific polyribosomes that translate mRNAs encoding secretory proteins recognize and attach to the endoplasmic reticulum (Fig. 3-6).

Because microsomal membranes in vitro reproduce the processing activity of intact cells, it was possible to identify macromolecules responsible for processing of the RNA precursor. Initial clues to this process came from determinations of the amino acid sequences of the proteins programmed by the cell-free translation of mRNAs encoding secreted polypeptides. Secreted proteins are synthesized as precursors that are extended at their NH₂ termini by sequences of 15 to 30 amino acids, called signal or leader sequences. Signal sequence extensions, or their functional equivalents, are required for targeting the ribosomal or nascent protein to specific membranes and for the vectorial transport of the protein across the membrane of the endoplasmic reticulum. On emergence of the signal sequence from the large ribosomal subunit, the ribosomal complex specifically makes contact with the membrane, which results in translocation of the nascent polypeptide across the endoplasmic reticulum membrane into the cisterna as the first step in the transport of the polypeptide within the secretory pathway. These observations initially left unanswered the question of how specific polyribosomes that translate mRNAs encoding secretory proteins recognize and attach to the endoplasmic reticulum (Fig. 3-6).

The specific interaction of the signal recognition particle with the nascent signal sequence and the polypeptide bears further translation of the mRNA. The nascent protein remains in a state of arrested translation until it finds a high-affinity binding protein on the endoplasmic reticulum membrane, the signal recognition particle receptor, or docking protein. On interaction with the specific docking protein, the translational block is released and protein synthesis resumes. The protein is then transferred across the membrane of the endoplasmic reticulum through a proteinaceous tunnel.

At some point, near the termination of synthesis of the polypeptide chain, the NH₂-terminal signal sequence is cleaved from the polypeptide by a specific signal peptidase located on the cisternal surface of the endoplasmic reticulum membrane. The removal of the hydrophobic signal sequence frees the protein (prohormone or hormone) so that it may assume its characteristic secondary structure during transport through the endoplasmic reticulum and the Golgi apparatus. Interestingly, after its cleavage from the protein by signal peptidase, the signal peptide may sometimes be further cleaved in the endoplasmic reticulum membrane to produce a biologically active peptide. The signal sequence of preprolactin of 30 amino acids, for example, is cleaved by a signal peptide peptidase to give a charged peptide of 20 amino acids that is released into the cytosol, where it binds to calmodulin and inhibits Ca²⁺-calmodulin-dependent phosphodiesterase.

This sequence in the directional transport of specific polypeptides ensures optimal conditional processing of secretory proteins, even when synthesis commences on free ribosomes. The presence of a proteinaceous form of the signal recognition particle complex that blocks translation guarantees that the synthesis of the presecretory proteins is not completed in the cytoplasm; the efficient transfer of proteins occurs only after contact has been made with the specific receptor or docking protein on the membrane. Although the identification of the signal recognition particle and the docking protein explains the specificity of the binding of ribosomes containing mRNAs encoding the secretory proteins, it does not explain the mode of translocation of the nascent polypeptide chain across the membrane bilayer. Further dissection and analysis of the membrane have identified other macromolecules that are responsible for the transport process.
Figure 3-6 Diagram depicting cellular events in initial stages of synthesis of a polypeptide hormone according to the signal hypothesis. In this schema, a signal recognition particle, consisting of a complex of six proteins and an RNA (7S RNA), interacts with the NH$_2$-terminal signal peptide of the nascent polypeptide chain after approximately 70 amino acids are polymerized, which results in the arrest of further growth of the polypeptide chain. The complex of the signal recognition particle and the polyribosome nascent chain remains in a state of translational arrest until it recognizes and binds to a docking protein, which is a receptor protein located on the cytoplasmic face of the endoplasmic reticular membrane. This interaction of the signal recognition particle complex with docking protein releases the translational block, and protein synthesis resumes. The nascent polypeptide chain is discharged across the membrane bilayer into the cisterna of the endoplasmic reticulum and is released from the signal peptide by cleavage with a signal peptidase located in the cisternal face of the membrane. In this model, the signal peptide is cleaved from the polypeptide chain by signal peptidase before the chain is completed (cotranslational cleavage). The configuration of the polypeptide during transport across the membrane and the forces and mechanisms responsible for its translocation are unknown. The loop, or hairpin, configuration of the chain that is shown is an arbitrary model; other models are equally possible.
Cellular Processing of Prohormones

The signal sequences of prehormones and pre-prohormones are involved in the transport of these molecules, but the function of the intermediate hormone precursors (prohormones) is not fully understood. The conversion of prohormones to their final products begins in the Golgi apparatus. For example, the time that elapses between the synthesis of pre-proparathyroid hormone and the first appearance of parathyroid hormone correlates closely with the time required for radioautographic grains to reach the Golgi apparatus. Similarly, the conversion of proinsulin to insulin takes place about an hour after the synthesis of proinsulin is complete, and processing of proinsulin to insulin and C peptide takes place during the transport within the secretory granule. The conversion of prohormones to hormones can also be blocked by inhibitors of cellular energy production such as antimycin A and dinitrophenol and by drugs that interfere with the functions of microtubules (vinblastine, colchicine). Thus, the translocation of the prohormone from the rough endoplasmic reticulum to the Golgi complex depends on metabolic energy and probably involves microtubules.

There is no evidence that sequences that are specific to the prohormone contribute to or are chemically involved in transport of the newly synthesized protein from the rough endoplasmic reticulum to the Golgi apparatus or that they are involved in the packaging of the hormone in the vesicles or granules. Analyses of the structures of the primary products of translation of mRNAs encoding secretory proteins indicate that many of these are not synthesized in the form of prohormone intermediates (see Fig. 3-3). It remains puzzling that some secretory proteins (e.g., parathyroid hormone, insulin, serum albumin) are formed by way of intermediate precursors, whereas others (e.g., growth hormone, prolactin, albumin) are not.

Size constraints may be placed on the length of a secretory polypeptide. When the bioactivity of peptides resides at the COOH termini of the precursors (e.g., somatostatin, calcitonin, gastrin), NH₂-terminal extensions may be required to provide a sufficient "spacer" sequence to allow the signal sequence on the growing nascent polypeptide chain to emerge from the large ribosome subunit for interaction with the signal recognition particle and to provide adequate polypeptide length to span the large ribosomal subunit and the membrane of the endoplasmic reticulum during vectorial transport of the nascent polypeptide across the membrane (see Fig. 3-6). When the final hormonal product is 100 amino acids long or longer (e.g., growth hormone, prolactin, or the α and β subunits of the glycoprotein hormones), there may be no requirement for a prohormone intermediate.

Although the exact functions of prohormones remain unknown, certain details of their cleavages have been established. Unlike

![Figure 3-7](image)

Figure 3-7 Regulatory feedback loops of the hypothalamic-pituitary-target organ axis. Being a combination of both stimulatory and inhibitory factors, hormones often act in concert to maintain homeostatic balance in the presence of physiologic or pathophysiologic perturbations. The concerted actions of hormones typically establish closed feedback loops by stimulatory and inhibitory effects coupled to maintain homeostasis.

Specific prohormone-converting enzymes (PCs) consist of a family of at least eight such enzymes. The most studied of the isozymes are PC2 and PC1/3, which are responsible for the cleavages of proinsulin between the A chain/C peptide and B chain/C peptide, respectively. A rare patient missing PC1 presented with childhood obesity, hypogonadotropic hypogonadism, and hypercortisolism and was found to have elevated proinsulin levels and presumably widespread abnormalities in neuropeptide modification. Targeted disruption of the PC2 gene in mice resulted in incomplete processing of proinsulin, leaving the A chain and C peptide intact. Notably, proglucagon in the pancreas remains completely unprocessed, indicating that PC2 is required for the formation of glucagon. As a consequence of defective PC2 activity and low levels of glucagon, the mice have severe chronic hypoglycemia.

After endopeptidase cleavage, the remaining basic residues are selectively removed by exopeptidases with activity resembling that of carboxypeptidase B. In the instances in which the COOH-terminal residue of the peptide hormone is amidated, a process that appears to enhance the stability of a peptide by conferring resistance to carboxypeptidase, specific amidation enzymes in the Golgi complex work in concert with the cleavage enzymes for modification of the COOH terminal of the bioactive peptides.

All proproteins and prohormones are cleaved by PC enzymatic processes within the Golgi complex of cells of diverse origins. The significance of specific cleavages of specific prohormones remains incompletely understood, as does the reason for the existence of prohormone intermediates in some but not all secretory proteins. As indicated earlier, precursor peptides removed from the prohormones may have intrinsic biologic activities that are as yet unrecognized.
 PROCESSES OF HORMONE SECRETION

Specific extracellular stimuli control the secretion of polypeptide hormones. The stimuli consist of changes in homeostatic balance; the hormonal products released in response to the stimuli act on the respective target organs to reestablish homeostasis [Fig. 3-7]. Endocrine systems typically consist of closed-loop feedback mechanisms such that, if hormones from organ A stimulate organ B, organ B in turn secretes hormones that inhibit the secretion of hormones from organ A. The concerted actions of both positive and negative hormonal influences thereby maintain homeostasis. For example, an increase in the concentration of plasma electrolytes as a consequence of dehydration stimulates the release of arginine vasopressin (also called antidiuretic hormone [ADH]) in the neural lobe of the pituitary gland, and vasopressin in turn acts on the kidney to increase the reabsorption of water from the renal tubule, thereby readjusting serum electrolyte concentrations toward normal levels.

These regulatory processes commonly include inhibitory feedback loops in which the products elaborated by the target organs in response to the actions of a hormone inhibit further endocrine secretion. An example of such negative feedback regulation is the control of the secretion of adrenocorticotropic hormone (ACTH) by the anterior pituitary gland. Increased ACTH stimulates the adrenal cortex to produce and secrete cortisol, which in turn feeds back to suppress further pituitary secretion of ACTH.

In many instances, endocrine regulation is complex and involves the responses of several endocrine glands and their respective target organs. After a meal, the release of a dozen or more hormones is triggered as a result of gastric distention, variations in the pH of the contents of the stomach and duodenum, and increased concentrations of glucose, fatty acids, and amino acids in the blood. The rise in plasma glucose and amino acid levels stimulates the release of insulin and the incretin hormones glucagon-like peptide 1 and glucose-dependent insulinotropic peptide and suppresses the release of glucagon from the pancreas. Both effects promote the net uptake of glucose by the liver; insulin increases cellular transport and uptake of glucose, and the lower blood levels of glucagon decrease the outflow of glucose because of diminished rates of glycogenolysis and gluconeogenesis.
STRUCTURE OF A GENE ENCODING A POLYPEPTIDE HORMONE

Structural analyses of gene sequences have resulted in at least three major discoveries that are important for understanding the expression of peptide-encoding genes. First, sequences of almost all the known biologically active hormonal peptides are contained within larger precursors that often encode other peptides, many of which are of unknown biologic activity. Second, the transcribed regions of genes (exons) are interrupted by sequences (introns) that are transcribed but subsequently cleaved from the initial RNA transcripts during their nuclear processing and assembly into specific mRNAs. Third, specific regulatory sequences reside in the regions of DNA flanking the 5' ends of structural genes, and these DNA sequences constitute specific targets for the interactions of DNA-binding proteins that determine the level of expression of the gene.

The DNA of higher organisms is wound into a tightly and regularly packed chromosomal structure in association with a number of different proteins organized into elements called nucleosomes. Nucleosomes are composed of four or five different histone subunits that form a core structure about which approximately 140 base pairs of genomic DNA are wound. The nucleosomes are arranged similarly to beads on a string, and coils of nucleosomes form the fundamental organizational units of the eucaryotic chromosome.

The nucleosomal structure serves several purposes. For example, nucleosomes enable the large amount of DNA (2 × 10^6 pairs) of the genome to be compacted into a small volume. Nucleosomes are involved in the replication of DNA and gene transcription. In addition to histones, other proteins are associated with DNA, and the complex nucleoprotein structure provides specific recognition sites for regulatory proteins and enzymes involved in DNA replication, rearrangements of DNA segments, and gene expression. The acetylation and deacetylation of histone-rich chromatin is involved in the regulation of gene transcription.

The topography of a typical protein-encoding gene consists of two functional units (Fig. 3-8):

- A transcriptional region
- A promoter or regulatory region

Transcriptional Regions

The transcriptional unit is the segment of the gene that is transcribed into an mRNA precursor. The sequences corresponding to the mature mRNA consist of the exon sequences that are spliced from the primary transcript during the posttranscriptional processing of the precursor RNA; these exons contain the code for the mRNA sequence that is translated into protein and for untranslated sequences at the 5' and 3'-flanking regions. The 5' sequence typically begins with a methylated guanine residue known as the cap site. The 3'-untranslated region contains within it a short sequence, AATAAA, that signals the site of cleavage of the 3' end of the RNA and the addition of a poly(A) tract of 100 to 200 nucleotides located approximately 20 bases from the AATAAA sequence. Although the functions of these modifications of the ends of mRNAs are not completely understood, they appear to provide signals for leaving the nucleus; enhance stability, perhaps through providing resistance to degradation by exonucleases; and stimulate initiation of mRNA translation. The protein-coding sequence of the mRNA begins with the codon AUG for methionine and ends with the codon immediately preceding one of the three nonsense, or stop, codons (UGA, UAA, and UAG).

The nature of the enzymatic splicing mechanisms that result in the excision of intron-coded sequences and the rejoining of exon-coded sequences is incompletely understood. Short "consensus" sequences of nucleotides reside at the splice junctions; for example, the bases GT and AG at the 5' and 3' ends of the introns, respectively, are invariant and a polypyrimidine stretch is found near the AG. Splicing involves a series of cleavage and ligation steps that remove the introns as a lariat structure with its 5' end ligated near the 3' end of the introns and ligate the two adjacent exons together. An elaborate machinery (the spliceosome) consisting of five small nuclear RNAs (snRNAs) and roughly 50 proteins direct these steps, guided by base pairing between three of the snRNAs and the mRNA precursor.
Regulatory Regions

The regulation of the expression of genes that encode polypeptides is beginning to be understood in some detail. As a result of experiments involving the deletion of 5' sequences upstream from structural genes, followed by analyses of the expression of the genes after introduction into cell lines, several insights have been obtained. These regulatory sequences, termed promoters and enhancers, consist of short polynucleotide sequences (see Fig. 3-8). They can be divided into at least four groups with respect to their functions and distances from the transcriptional initiation site.

First, the sequence TATAA (TATA, or Goldberg-Hogness, box) is usually present in the more proximal promoter within 25 to 30 nucleotides upstream from the point of transcriptional initiation. The TATA sequence is required to ensure the accuracy of initiation of transcription at a particular site. The TATA box directs the binding of a complex of several proteins, including RNA polymerase II. The proteins, referred to as TATA box transcription factors (TFs), number six or more basal factors (IIA, IIB, IID, IIE, IIF, IIH) and, along with RNA polymerase II, form the general or basal transcriptional machinery required for the initiation of RNA synthesis.

The other three groups of regulatory sequences consist of tissue-specific silencers (TSSs), which function by binding repressor proteins; tissue-specific enhancers (TSEs), which are activated by the binding of transcriptional activator proteins; and metabolic response elements (MREs), which are regulated by the binding of specialized proteins whose transcriptional activities (repressor or activator) are regulated by metabolic signaling, often involving changes in their phosphorylation.
Introns and Exons

Genes encoding proteins and ribosomal RNAs in eukaryotes are interrupted by intervening DNA sequences (introns) that separate them into coding blocks (exons). In bacterial genes the nucleotide sequences of the chromosomal genes match precisely the corresponding sequences in the mRNAs. Interruption of the continuity of genetic information appears to be unique to nucleated cells. The reasons for such interruption are not completely understood, but introns appear to separate exons into functional domains with respect to the proteins that they encode. An example is the gene for proglucagon, a precursor of glucagon in which five introns separate six exons, three of which encode glucagon and the two glucagon-related peptides contained within the precursor. A second example is the growth hormone gene, which is divided into five exons by four introns that separate the promoter region of the gene from the protein-coding region and the latter into three partly homologous repeated segments, two coding for the growth-promoting activity of the hormone and the third for its carbohydrate metabolic functions. As a rule, the genes for the precursors of hormones and regulatory peptides contain introns at or about the region where the signal peptides join the apoproteins or prohormones, thus separating the signal sequences from the components of the precursor that are exported from the cell as hormones or peptides.

There are exceptions to the one exon, one function theory in mammalian cells. The genes of several precursors of peptide hormones are not interrupted by introns in a manner that corresponds to the separation of the functional components of the precursor. Notable in this regard is the precursor proopiomelanocortin, from which the peptides ACTH, -melanocyte-stimulating hormone, and -endorphin are cleaved during the post-translational processing of the precursor. The protein-coding region of the pro-opiomelanocortin gene is devoid of introns. Likewise, no introns interrupt the protein-coding region of the gene for the proenkephalin precursor, which contains seven copies of the enkephalin sequences. It is possible that, in the past, introns separated each of these coding domains and were lost during the course of evolution.

A precedent for the selective loss of introns appears to be exemplified by the rat insulin genes. The rat genome harbors two nonallelic insulin genes: one containing two introns and the other containing a single intron. The most likely explanation is that an ancestral gene containing two introns was transcribed into RNA and spliced; then that RNA was copied back into DNA by a cellular reverse transcriptase and inserted back into the genome at a new site.
REGULATION OF GENE EXPRESSION

The regulation of expression of genes encoding polypeptide hormones can take place at one or more levels in the pathway of hormone biosynthesis (Fig. 3-10):

- DNA synthesis (cell growth and division)
- Transcription
- Post-transcriptional processing of mRNA
- Translation
- Post-translational processing

In different endocrine cells, one or more levels may serve as specific control points for regulation of production of a hormone (see also Generation of Biologic Diversification later).

Figure 3-10 Diagram of an endocrine cell showing potential control points for regulation of gene expression in hormone production. Specific effector substances bind either to plasma membrane receptors (peptide effectors) or to cytosolic or nuclear receptors (steroids), which leads to initiation of a series of events that couple the effector signal with gene expression. In the illustration shown, peptide effector-receptor complex interactions act initially through activation of adenylate cyclase (AC) coupled with a guanosine triphosphate-binding protein (G). Coupling factors and substances such as glucose, cyclic adenosine monophosphate, and cations activate protein kinases, resulting in a series of phosphorylations of macromolecules. As discussed in the text, specific effectors for various endocrine cells appear to act at one or more of the indicated five levels of gene expression, with the possible exception of post-translational processing of prohormones, for which no definite examples of metabolic regulation have yet been found.

Levels of Gene Control

Newly synthesized prolactin transcripts are formed within minutes after exposure of a prolactin-secreting cell line to TRH. Cortisol stimulates growth hormone synthesis in both somatotropic cell lines and pituitary slices through increases in rates of gene transcription and enhancement of the stability of mRNA. The time required for cortisol to enhance transcription of the growth hormone gene is 1 to 2 hours, which is considerably longer than the time required for the action of TRH on prolactin gene transcription. Regulation of proinsulin biosynthesis appears to take place primarily at the level of translation. Within minutes after raising the plasma glucose level, the rate of proinsulin biosynthesis increases fivefold to 10-fold. Glucose acts either directly or indirectly to enhance the efficiency of initiation of translation of proinsulin mRNA.

Rapid metabolic regulation at the level of post-transcriptional processing of mRNA precursors is not yet clearly established. However, alternative exon splicing plays a major role in the regulation of the formation of mRNAs during development (see next section in chapter on Generation of Biologic Diversification). For example, the primary RNA transcripts derived from the calcitonin gene are alternatively spliced to provide two or more tissue-specific mRNAs that encode chimeric protein precursors with both common and different amino acid sequences, suggesting that regulation might take place at the level of processing of the calcitonin gene transcripts.

In many instances, the level of gene expression under regulatory control is optimal for meeting the secretory and biosynthetic demands of the endocrine organ. For example, after a meal there is an immediate requirement for the release of large amounts of insulin. This release depletes insulin stores of the pancreatic beta cells within a few minutes, and increasing the translational efficiency of preformed proinsulin mRNA provides additional hormone rapidly.
Differentiated cells have a remarkable capacity for selective expression of specific genes. In one cell type, a single gene may account for a large fraction of the total gene expression, and in another cell type the same gene may be expressed at undetectable levels.

When a gene can be expressed in a particular cell type, the associated chromatin is loosely arranged; when the same gene is never expressed in a particular cell type, the chromatin organization is more compact. Thus, the DNA within the chromatin of expressed genes is more susceptible to cleavage by deoxyribonuclease than is the DNA in tissues in which the genes are quiescent. This looseness may facilitate access of RNA polymerase to the gene for purposes of transcription. In addition, inactive genes appear to have a higher content of methylated cytosine residues than the same genes in tissues in which they are expressed.

Determinants for the tissue-specific transcriptional expression of genes exist in control sequences usually residing within 1000 base pairs of the 5'-flanking region of the transcriptional sequence. Enhancer sequences in animal cell genes were first described for immunoglobulin genes, a finding that extended the earlier observations of enhancer control elements in viral genomes. However, the first clear demonstrations of these elements directing transcription to cells of distinct phenotypes came from studies of the comparative expression of two model genes, insulin and chymotrypsin, in the endocrine and exocrine pancreas, respectively. The restricted expression of genes in a cell-specific manner is determined by the assembly of specific combinations of DNA-binding proteins on a predetermined array of control elements of the promoter regions of genes to create a transcriptionally active complex of proteins that includes the components of the general or basal transcriptional apparatus.
Transcription Factors in Developmental Organogenesis of Endocrine Systems

Certain families of transcription factors are critical for organogenesis and the development of the body plan. Among these factors are the homeodomain proteins and the nuclear receptor proteins. The family of homeodomain selector, or homeodomain, proteins are highly conserved throughout the animal kingdom from flies to humans. The orchestrated spatial and temporal expression of these proteins and the target genes that they activate determine the orderly development of the body plan of specific tissues, limbs, and organs. Similarly, the actions of families of nuclear receptors (steroid and thyroid hormones, retinoic acid, and others) are critical for normal development to occur. Inactivating mutations in the genes encoding these essential transcription factors predictably result in loss or impairment of the development of the specific organ whose development they direct.

Three examples are described of impaired organogenesis attributable to mutations in essential transcription factors:

- Partial anterior pituitary agenesis (Pit-1)
- Adrenal and gonadal agenesis (SF-1, DAX-1)
- Pancreatic agenesis (IDX-1)

Partial Pituitary Agenesis

The transcription factor Pit-1 is a member of a family of pou-homeodomain proteins, which is a specialized subfamily of the larger family of homeodomain proteins. Pit-1 is a key transcriptional activator of the promoters of the growth hormone, prolactin, and thyroid-stimulating hormone genes, produced in the anterior pituitary somatotrophs, lactotrophs, and thyrotrophs, respectively. Pit-1 is also the major enhancer activating factor for the promoter of the growth hormone-releasing factor receptor gene. Mutations in Pit-1 that impair its DNA-binding and transcriptional activation functions are responsible for the phenotype of the Jackson and Snell dwarf mice.

Mutations in the gene encoding Pit-1 have been found in patients with combined pituitary hormone deficiency in which there is no production of growth hormone, prolactin, or thyroid-stimulating hormone, resulting in growth impairment and mental deficiency. Notably, the production of the other two of the five hormones secreted by the anterior pituitary gland, adrenocorticotropin and the gonadotropins luteinizing hormone (LH) and follicle-stimulating hormone (FSH), is unaffected.

In these human Pit-1 mutations, Pit-1 can bind to its cognate DNA control elements but is defective in trans-activating gene transcription. Furthermore, the mutated Pit-1 acts as a dominant negative inhibitor of Pit-1 actions on the unaffected allele.

Pancreatic Agenesis

The homeodomain protein islet duodenum homeobox 1 or IDX-1 (somatostatin transcription factor 1 [STF-1], insulin promoter factor 1 [IPF-1]) appears to be responsible for the development and growth of the pancreas. Targeted disruption of the IDX-1 gene in mice resulted in a phenotype of pancreatic agenesis. A child born without a pancreas was shown to be homozygous for inactivating mutations in the IDX-1 gene. Notably, the parents and their ancestors who are heterozygous for the affected allele have a high incidence of maturity-onset (type 2) diabetes mellitus, suggesting that a decrease in gene dosage of IDX-1 may predispose to the development of diabetes. The possibility that a mutated IDX-1 allele may be one of several "diabetes genes" is supported by the observation that IDX-1 and the helix-loop-helix transcription factors E47 and beta-2 appear to be key up-regulators of the transcription of the insulin gene.

Agensis of the Adrenal Gland and Gonads

Two nuclear receptor transcription factors have been identified as critical for the development of the adrenal gland, gonads, pituitary gonadotrophs, and the ventral medial hypothalamus. These nuclear receptors are SF-1 (steroidogenic factor 1) and DAX-1 (dosage-sensitive sex reversal, adrenal hypoplasia congenita, X chromosome). SF-1 binds to half-sites of estrogen response elements that bind estrogen receptors in the promoters of genes. DAX-1 binds to retinoic acid receptor (RAR) binding sites in promoters and inhibits RAR actions. Targeted disruption of SF-1 in mice results in a phenotype of adrenal and gonadal agenesis. In addition, pituitary gonadotrophs are absent and the ventral medial hypothalamus is severely underdeveloped.

X-linked adrenal hypoplasia congenita is an X-linked, developmental disorder of the human adrenal gland that is lethal if untreated. The gene responsible for adrenal hypoplasia congenita has been identified by positional cloning and encodes DAX-1, a member of the nuclear receptor proteins related to RAR. Several inactivating mutations identified in the DAX-1 gene result in the syndrome of adrenal hypoplasia congenita and hypogonadotropic hypogonadism. Thus, genetically defined and transmitted defects in the genes encoding the transcription factors SF-1 and DAX-1 result in profound arrest in the development of the target organs regulated by the hypothalamic-pituitary-adrenal axis involved in steroidogenesis (the adrenal gland (glucocorticoids, mineral corticoids) and the gonads (estrogens and androgens)).
Coupling of Effector Action to Cellular Response

Another mode of gene control consists of the induction and suppression of genes that are normally expressed in a specific tissue. These processes are at work in the minute-to-minute and day-to-day regulation of rates of production of the specific proteins produced by the cells (e.g., production of polypeptide hormones in response to extracellular stimuli).

At least two classes of signaling pathways protein phosphorylation and activation of steroid hormone receptors by hormone bindings appear to be involved in the physiologic regulation of hormone gene expression. These two pathways mediate the actions of peptide and steroid hormones, respectively. Peptide ligands bind to receptor complexes on the plasma membrane, which results in enzyme activation, mobilization of calcium, formation of phosphorylated nucleotide intermediates, activation of protein kinases, and phosphorylation of specific regulatory proteins such as transcription factors (see Chapter 5).

Steroidal compounds, because of their hydrophobic composition, readily diffuse through the plasma membrane, bind to specific receptor proteins, and interact with other macromolecules in the nucleus, including specific domains on the chromatin located in and around the gene that is activated (see Chapter 4). Phosphorylated nucleotides such as cyclic adenosine monophosphate (cAMP), adenosine triphosphate, and guanosine triphosphate, as well as calcium, appear to have important functions in secretory processes. In particular, fluxes of calcium from the extracellular fluid into the cell and from intracellular organelles (e.g., endoplasmic reticulum) into the cytosol are closely coupled to secretion.

The cellular signaling pathways that involve protein phosphorylations are multiple and complex. They typically consist of sequential phosphorylations and dephosphorylations of molecules referred to as protein kinase or phosphatase cascades. These cascades are initiated by hormones, sensor molecules known as ligands, that bind to and activate receptors located on the surface of cells, resulting in the generation of small second messenger molecules such as cAMP, diacylglycerol, or calcium ions. These second messengers then activate protein kinases that phosphorylate and thereby activate key target proteins. The final step in the signaling pathways is the phosphorylation and activation of important transcription factors, resulting in gene expression (or repression).

Insight has been gained into the identities of some of the phosphoproteins. As discussed earlier, a specific group of transcription factors, DNA-binding proteins, interacts with cAMP-responsive and phorbol ester-responsive DNA elements to stimulate gene transcription mediated by the cAMP-protein kinase A, diacylglycerol-protein kinase C, and calcium-calmodulin signal transduction pathways (see Fig. 3-11). These proteins are encoded by a complex family of genes and bind to the DNA elements in the form of heterodimers or homodimers through a coiled coil helical structure known as a leucine zipper motif. There is evidence that phosphorylation of these proteins modulates dimerization, DNA recognition and binding, and transcriptional trans-activation activities. Phosphorylation of the protein substrates might change their conformations and activate the proteins, which, in turn, interact with coactivator proteins such as the cAMP response element-binding protein (CREB) and the protein components of the basal transcriptional machinery, thereby allowing RNA polymerase to initiate gene transcription.

Generally, the second messengers activate serine/threonine kinases, which phosphorylate serine or threonine residues, or both, on proteins, whereas the receptor kinases are tyrosine-specific kinases that phosphorylate tyrosine residues. Examples of receptor tyrosine kinases are growth factor receptors such as those for insulin, insulin-like growth factor (IGF), and platelet-derived growth factor. Receptors in the cytokine receptor family, which include leptin, growth hormone, and prolactin, activate associated tyrosine kinases in a variation on the theme.

The different types of signal transduction pathways are described as more or less distinct pathways for semantic purposes. In reality, there is considerable cross-talk among the different pathways that occur developmentally and in cell typespecific settings. An active area of research in endocrine systems is attempting to understand these complex interactions among different signal transduction pathways. Although the growth factor and cytokine receptors are similar in some respects, they differ in other respects. For example, growth factor receptor tyrosine kinases activate transcription factors through cascades that involve both tyrosine phosphorylation and serine/threonine kinases such as mitogen-activated protein kinases, whereas the Janus kinases (JAKs) activated by cytokine receptors directly tyrosine phosphorylate the signal transducer and activator of transcription (STAT) factors.
GENERATION OF BIOLOGIC DIVERSIFICATION

In addition to providing control points for the regulation of gene expression, the various steps involved in transfer of information encoded in the DNA of the gene to the final bioactive protein are a means for diversification of information stored in the gene (Fig. 3-12). Five steps in gene expression can be arbitrarily described: (1) gene duplication and copy number, (2) transcription, (3) post-transcriptional RNA processing, (4) translation, and (5) post-translational processing.

Gene Duplications

At the level of DNA, diversification of genetic information comes about by way of gene duplication and amplification. Many of the polypeptide hormones are derived from families of multiple, structurally related genes. Examples include the growth hormone family, consisting of growth hormone, prolactin, and placental lactogen; the glucagon family, consisting of glucagon, vasoactive intestinal peptide, secretin, gastric inhibitory peptide, and growth hormonereleasing hormone; and the glycoprotein hormone family, thyrotropin, luteinizing hormone, follicle-stimulating hormone, and chorionic gonadotropin.

A remarkable example of diversification at the level of gene amplifications is the extraordinarily large number of genes encoding the pheromone and odorant receptors. It is estimated that as many as 1000 such receptor genes may exist in mouse and rat genomes, each receptive to a particular odorant ligand. Over the course of evolution, an ancestral gene encoding a prototypic polypeptide representative of each of these families was duplicated one or more times and, through mutation and selection, the progeny proteins of the ancestral gene assumed different biologic functions. The exonic-intronic structural organization of the genomes of higher animals lends itself to gene recombination and RNA copying of genetic sequences with subsequent reintegration of DNA reverse-transcribed sequences back into the genome, resulting in rearrangement of transcriptional units and regulatory sequences.
Transcription

In addition to duplication of genes and their promoters, another way to create diversity in expression is at the level of gene transcription by providing genes with alternative promoters and by utilizing a large array of cis-regulatory elements in the promoters regulated by complex combinations of transcription factors.

Alternative Promoters

Many of the genes encoding hormones and their receptors utilize more than one promoter during development or when expressed in different tissue types. The employment of alternative promoters results in the formation of multiple transcripts that differ at their 5' ends. It is presumed that some genes have multiple promoters because they provide flexibility in the control of expression of the genes. For example, in some cases, expression of genes in more than one tissue or developmental stage may require distinct combinations of tissue-specific transcription factors. This flexibility enables genes in different cell types to respond to the same signal transduction pathways or genes in the same cell type to respond to different signal transduction pathways. A single promoter may not be adequate to respond to a complex array of transcription factors and a changing environment of cellular signals.

The organization of alternative promoters in genes is manifested in several patterns within exons or introns in the 5' noncoding sequence or the coding sequence. The most common occurrence of alternative promoters is within the 5' noncoding or leader exons. The utilization of different promoters in the 5' untranslated region of a gene, often accompanied by alternative exon splicing, results in the formation of mRNAs with different 5' sequences. The alternative usage of promoters in 5' leader exons can affect gene expression and generate diversity in several different ways. These include the developmental stagespecific and temporal expression of genes, the tissue-type specificity of expression, the levels of expression, the responsivity of gene expression to specific metabolic signals conveyed through signal transduction pathways, the stability of the mRNAs, the efficiencies of translation, and the structures of the amino termini of proteins encoded by the genes.

Examples of genes that use alternative 5' leader promoters during development are those encoding IGF-I, IGF-II, the retinoic acid receptors, and glucokinase, all of which are regulated by multiple promoters that are active in a variety of embryonic and adult tissues and are subject to developmental and tissue-specific regulation. During fetal development, promoters P2, P3, and P4 of the IGF-II gene are active in the liver. These promoters are shut off after birth, at which time the P1 promoter is activated. The P1 and P2 promoters of the IGF-I gene are differentially responsive to growth hormone: P2 expressed in liver is responsive to growth hormone, whereas P1 expressed in muscle is not.

The retinoic acid receptor exists in three isoforms (RAR, RAR, and RAR) encoded by separate genes that give rise to at least 17 different mRNAs generated by a combination of multiple promoters and alternative splicing. The RAR isoforms appear to differ in their specificity for retinoic acid receptors, in their affinities for ligand isoforms, and in trans-activating capabilities. The different RAR isoforms are expressed at different times in different tissues during development. It has been proposed that the different RAR isoforms provide a means of achieving a diverse set of cellular responses to a single, simple ligand, retinoic acid.

Glucokinase is an example of the alternative use of 5' leader promoters that have different metabolic responsiveness. Expression of glucokinase in pancreatic beta cells and some other neuroendocrine cells utilizes an upstream promoter (1), whereas in liver a promoter (IL) 26 kb downstream of the 1 promoter is used exclusively. In beta cells, expression of the glucokinase gene is apparently not responsive to hormones. In contrast, in liver expression mediated by the IL promoter is intensely up-regulated by insulin and down-regulated by glucagon.

The amylase gene provides an example in which two alternative promoters in the 5' noncoding exons expressed in two different tissues have dramatically different strengths of expression. A strong upstream promoter directs expression within the parotid gland, contrasting with weak expression directed by an alternative downstream promoter in liver.

Examples of the alternative usage of promoters in the coding regions of genes are the progesterone receptor (PR) and the transcription factor CAMP response element modulator (CREM). In both of these examples, different protein isoforms are produced that have markedly different functional activities. The genes encoding the chicken and human progesterone receptors express two isoforms of the receptor (isoforms A and B). Isoform A initiates translation at a methionine residue located 164 amino acids downstream from the methionine that initiates the translation of the longer form B. Analyses of the mechanisms responsible for the synthesis of two different isoforms revealed that two promoters exist in the human PR gene: one upstream of the 5' leader exon and the other in the first protein coding exon. The two isoforms of the human PR differ markedly in their capabilities to trans-activate transcription from different progesterone responsive elements (PRE). Both human PR isoforms equivalently activate a canonical PRE. Isoform B is much more efficient than A at activating the PRE in the mouse mammary tumor virus promoter, whereas isoform A, but not B, activates transcription from the ovalbumin promoter.

The utilization of an alternative intronic promoter within the protein coding sequence of a gene is exemplified by the CREM gene. The CREM gene employs a constitutively active, unregulated promoter (P1) that encodes predominantly activator forms of CREM and an internal promoter (P2) located in the fourth intron that is regulated by cAMP signaling and encodes a repressor isoform, ICER (inducible cAMP early response). The remarkable complexity of the alternative mechanisms of expression of the CREM and CREB genes is discussed subsequently.

Diversity of Transcription Factors

Another mechanism to create diversity at the level of gene transcription is that of the interplay of multiple transcription factors on multiple cis-regulatory sequences. The promoters of typical genes may contain 20 or 30 or more cis-acting control elements, either enhancers or silencers. These control elements may respond to ubiquitous transcription factors found in all cell types and to cell type-specific factors.

Unique patterns of control of gene expression can be affected by several different mechanisms acting in concert. The spacing, relative locations, and juxtapositioning of control elements with respect to each other and to the basal transcriptional machinery can influence levels of expression. Transcription factors often act in the form of dimers or higher oligomers among factors of the same or different classes. A given transcription factor may act as either an activator or a repressor as a consequence of the existing circumstances. The ambient concentrations of transcription factors within the nucleus in conjunction with their relative DNA-binding affinities and trans-activation potencies may determine the levels of expression of genes.
Post-transcriptional Processing (Alternative Exon Splicing)

Identification of the mosaic structure of transcriptional units encoding polypeptide hormones and other proteins that consist of exons and introns raised the possibility that the use of alternative pathways in RNA splicing could provide informationally distinct molecules. Different proteins could arise either by inclusion or exclusion of specific exonic segments or by utilization of parts of introns in one mRNA as exons in another mRNA. In addition, differences in the splice sites would result in expression of new translational reading frames. Alternative splicing utilizes two distinct mechanisms (Fig. 3-14). One is that of exon skipping or switching in or out of exons. The other mechanism, known as intron slippage, is to include part of an intron in an exon, to splice out part of an exon along with the intron, or to include a "coding" intron.

There are many examples of both mechanisms used to generate diversity in endocrine systems. Included among the genes encoding prohormones in which the pre-mRNAs are alternatively spliced by exon skipping or switching are those for procalcitonin/calcitonin gene-related peptide, prosubstance P/K, and the prokininogens. Alternative processing of the RNA transcribed from the calcitonin gene results in production of an mRNA in neural tissues that is distinct from that formed in the C cells of the thyroid gland. The thyroid mRNA encodes a precursor to calcitonin, whereas the mRNA in the neural tissues generates a neuropeptide known as calcitonin gene-related peptide. Immunocytochemical analyses of the distribution of the peptide in brain and other tissues suggest functions for the peptide in perception of pain, ingestive behavior, and modulation of the autonomic and endocrine systems.

The splicing of the RNA precursor that encodes substance P can take place in at least two ways. One splicing pattern results in the mRNA that encodes substance P and another peptide, called substance K, in a common protein precursor. Other mRNAs are apparently spliced so as to exclude the coding sequence for substance K. An alternative RNA splicing pattern also occurs in the processing of transcripts arising from the gene encoding bradykinin.

Other examples of genetic diversification arise from the programmed flexibility in the choice of splice acceptor sites within coding regions (intron slippage), which allows an array of coding sequences (exons) to be put together in a number of useful combinations. For example, the coding sequences of the growth hormone, lutropin-choriogonadotropin, and leptin receptors can be brought together in two different ways, one to include, the other to exclude, an exonic coding sequence specifying the transmembrane spanning domains of the polypeptide chains that anchor the receptors to the surface of cells. If mRNA splicing excludes the anchor's peptide sequence, a secreted rather than a surface protein is produced.
Translation

The process of translation provides a fourth level for the creation of diversity of gene expression. As discussed earlier in the section on Regulation of Gene Expression, the rate of translational initiation can be regulated as typified by the proinsulin and prohormone convertase mRNAs, in which translation is augmented by glucose and cAMP. Molecular diversity of translation, however, is generated by the developmentally regulated utilization of alternative translation initiation (start) codons (methionine codons, AUGs). The mechanism of translation initiation involves the assembly of the 40S ribosome subunit on the 5' methyl guanosine cap of the mRNA. The ribosome subunit then scans 5' to 3' along the mRNA until it encounters an AUG sequence in a context of surrounding nucleotides favorable for the initiation of protein synthesis. Upon encountering such a favorable AUG, the subunit pauses and recruits the 60S subunit plus a number of other essential translation initiation factors, allowing the polymerization of amino acids.

The use of an alternative downstream start codon for translation can occur by mechanisms of loose scanning and reinitiation (Fig. 3-15). Loose scanning is believed to occur when the most 5' AUG codon is not in a strongly favorable context and allows the 40S ribosomal subunit to continue scanning until it encounters a more favorable AUG downstream. Thus, in the loose scanning mechanism, both translational start codons are used. In contrast, the mechanism of translational reinitiation involves the termination of translation followed by the reinitiation of translation at a downstream start codon. Thus, two proteins are encoded from the same mRNA by a start and stop mechanism.

This process of translational reinitiation can occur either by continued scanning of the 40S ribosomal subunit after termination of translation followed by reinitiation, or by complete dissociation of the ribosomal subunits at the time of termination followed by complete reassembly at a downstream start codon, referred to as an internal ribosomal entry site (IRES). Such utilization of alternative translation start codons occurs in mRNAs encoding certain classes of transcription factors illustrated by the basic leucine zipper (bZIP) proteins CREB, CREM, and certain of the CCAAT/enhancer binding proteins (C/EBPs), the C/EBP and C/EBP isoforms. In all four of these DNA-binding proteins, the alternative use of internal start codons results in a switch from activators to repressors.

The CREB gene uses translational reinitiation by the somewhat novel mechanism of alternative exon switching that occurs during spermatogenesis. At developmental stages IV and V of the seminiferous tubule of the rat, an exon (exon W) is spliced into the CREB mRNA. Exon W introduces an inframe stop codon, thereby terminating translation approximately 40 amino acids upstream of the DNA-binding domain. The termination of translation then permits reinitiation of translation at each of two downstream start codons, resulting in the synthesis of two repressor or inhibitor isoforms of CREB known as I-CREBs that are powerful dominant negative inhibitors of activator forms of CREB and CREM because they consist of the DNA-binding domain devoid of any trans-activation domains. The function, if any, of the amino-terminal truncated protein consisting of the activation domains devoid of the DNA-binding domain is unknown. It has been postulated that the role of the alternative splicing of exon W in the CREB pre-mRNA is to interrupt a forward positive feedback loop during spermatogenesis.

CREM, C/EBP, and C/EBP mRNAs utilize alternative downstream start codons to synthesize repressors during development. Like the I-CREBs, these repressors consist of the DNA-binding domains and lack trans-activation domains. The CREM repressor (S-CREM) is expressed during brain development. The C/EBP-30 and C/EBP-20 isoforms are expressed during the differentiation of adipoblasts to adipocytes, and the C/EBP repressor liver inhibitory protein (LIP) is expressed during the development of the liver.
Post-translational Processing

A fifth level of gene expression at which diversification of biologic information can take place is that of post-translational processing. Many precursors of polypeptide hormones, particularly those encoding small peptides, contain multiple peptides that are cleaved during post-translational processing of the prohormones. Certain polyprotein precursors, however, contain several copies of the peptide. Examples of prohormones that contain multiple identical peptides are the precursors encoding TRH and the mating factor of yeast, each of which contains four copies of the respective peptide. Polypeptides that contain several distinct peptides include proenkephalins, pro-opiomelanocortin, and proglucagon.

In many instances, biologic diversification at the level of post-translational processing occurs in a tissue-specific manner. The processing of pro-opiomelanocortin differs markedly in the anterior compared with the intermediate lobe of the pituitary gland. In the anterior pituitary the primary peptide products are ACTH and -endorphin, whereas in the intermediate lobe of the pituitary one of the primary products is -melanocyte-stimulating hormone. The smaller peptides produced are extensively modified by acetylation and phosphorylation of amino acid residues.

The processing of proglucagon in the pancreatic A cells and that in the intestinal L cells are also different. In the pancreatic A cells, the predominant bioactive product of the processing of proglucagon is glucagon itself; the two glucagon-like peptides are not processed efficiently from proglucagon in the A cells and are biologically inactive by virtue of having NH$_2$-terminal and COOH-terminal extensions. On the other hand, in the intestinal L cell the glucagon immunoreactive product is a molecule, called glicentin, that consists of the NH$_2$-terminal extension of the proglucagon plus glucagon and the small COOH-terminal peptide known as intervening peptide I.

Glicentin has no glucagon-like biologic activity, and therefore the bioactive peptide (or peptides) in the intestinal L cells must be one or both of the glucagon-like peptides. In fact, glucagon-like peptide I in its shortened form of 31 amino acids, GLP-I (737), is a potent insulinotropic hormone in its actions of stimulating insulin release from pancreatic beta cells. This peptide is released from the intestines into the blood stream in response to oral nutrients and appears to be a potent intestinal incretin factor implicated in the augmented release of insulin in response to oral compared with systemic (intravenous) nutrients. This potential for diversification of biologic information provided by the alternative pathways of gene expression is impressive when one considers that these pathways can occur in multiple combinations.
Unexpectedly Low Numbers of Expressed Genes in Genomes of Mammals (Humans and Mice)

A somewhat surprising initial conclusion, heralded in the lay press when the results of the sequencing of the human and mouse genomes were revealed, was that the number of genes in the human and mouse was approximately 30,000. This number was viewed as remarkably low because the number of genes in yeast (Saccharomyces cerevisiae), worm (Caenorhabditis elegans), and fly (Drosophila melanogaster) is about 20,000. However, it seems quite clear from the complexities of the mRNAs expressed in humans and mice, as exemplified by the growing database of expressed sequence tags, that tissue-specific alternative exon splicing and alternative promoter usage occur much more frequently in humans and mice than in yeast, worms, and flies. Considering the as yet incomplete database of expressed genes at the mRNA level, it seems reasonable to extrapolate that the human genome may actually express as many as 100,000 to 200,000 mRNAs that encode proteins with distinct, specific functions. This extrapolation is based on the observation that alternative exon splicing and promoter usage appear to be on the order of 5 to 10 times more frequent in higher vertebrate mammals than in yeasts and flies.
Acknowledgments

I am indebted to the members of the laboratory whose forbearance and helpful discussions of this chapter were invaluable. I thank Townley Budde for help in the preparation of the manuscript. J. F. H. is an Investigator with the Howard Hughes Medical Institute.
References


Chapter 4 - Mechanism of Action of Hormones That Act on Nuclear Receptors

Mitchell A. Lazar

Hormones can be divided into two groups on the basis of where they function in a target cell. The first group includes hormones that do not enter cells; instead, they signal via second messengers generated by interacting with receptors at the cell surface. All polypeptide hormones, as well as monoamines and prostaglandins, utilize cell surface receptors (see Chapter 5, “Mechanism of Action of Hormones That Act at the Cell Surface”). The second group, the focus of this chapter, includes hormones that can enter cells. These hormones bind to intracellular receptors that function in the nucleus of the target cell to regulate gene expression. Classical hormones that utilize intracellular receptors include thyroid and steroid hormones.

Hormones serve as a major form of communication between different organs and tissues that allows specialized cells in complex organisms to respond in a coordinated manner to changes in the internal and external environments. Classical endocrine hormones, such as thyroid and steroid hormones, are secreted by ductless glands and are distributed throughout the body via the blood stream. These hormones were discovered by purifying the biologically active substances from clearly definable glands.

It is now recognized that numerous other signaling molecules share with thyroid and steroid hormones the ability to function in the nucleus to convey intercellular and environmental signals. Not all of these molecules are produced in glandular tissues. Further, whereas some of these signaling molecules arrive at target tissues via the blood stream like classical endocrine hormones, others have paracrine functions (i.e., they act on adjacent cells) or autocrine functions (i.e., they act on the cell of origin).

Lipophilic signaling molecules that utilize nuclear receptors include the following:

- Derivatives of vitamins A and D
- Endogenous metabolites such as oxysterols and bile acids
- Non-natural chemicals encountered in the environment (xenobiotics)

These molecules are referred to generically as ligands for nuclear receptors. The nuclear receptors for all of these signaling molecules are structurally related and collectively referred to as the nuclear receptor superfamily.
LIGANDS THAT ACT VIA NUCLEAR RECEPTORS

General Features of Nuclear Receptor Ligands

Unlike polypeptide hormones that function via cell surface receptors, no ligands for nuclear receptors are directly encoded in the genome. To the contrary, all nuclear receptor ligands are small (molecular weight < 1000 daltons [d]) and lipophilic, enabling them to enter cells. Cellular uptake of nuclear receptor ligands may be a passive process, but in some cases a membrane transport protein is involved. For example, the oatp3 organic anion transporter mediates thyroid hormone entry into cells. The lipophilicity of nuclear receptor ligands also allows them to be absorbed from the gastrointestinal tract, thus facilitating their use in replacement or pharmacologic therapies of disease states.

Another common feature of nuclear receptor ligands is that all are derived from dietary, environmental, and metabolic precursors. In this sense, the function of these ligands and their receptors is to translate cues from the external and internal environments into changes in gene expression. Their critical role in maintaining homeostasis in multicellular organisms is highlighted by the fact that nuclear receptors are found in all vertebrates as well as insects but not in single-cell organisms such as yeast.
Subclasses of Nuclear Receptor Ligands

One classification of nuclear receptor ligands is outlined in Table 4-1 and is described next.

Classical Hormones

The classical hormones that utilize nuclear receptors for signaling are thyroid hormone and steroid hormones. Steroid hormones include receptors for cortisol, aldosterone, estrogen, progesterone, and testosterone. In some cases (e.g., thyroid hormone receptor [TR] and genes, estrogen receptor [ER] and ), there are multiple receptor genes, encoding multiple receptors. Multiple receptors for the same hormone can also derive from a single gene either by alternative promoter usage or alternative splicing (e.g., TR 1 and 2).

Finally, some receptors can mediate the signal of multiple hormones. For example, the mineralocorticoid (aldosterone) receptor (MR) has equal affinity for cortisol and probably functions as a glucocorticoid receptor in some tissues, such as the brain. The androgen receptor (AR) binds and responds to both testosterone and dihydrotestosterone (DHT).

Vitamins

Vitamins were discovered as essential constituents of a healthful diet. Two fat-soluble vitamins, A and D, are precursors of important signaling molecules that function as ligands for nuclear receptors.

<table>
<thead>
<tr>
<th>TABLE 4-1 – Nuclear Receptor Ligands and Their Receptors</th>
</tr>
</thead>
</table>

### Classical Hormones

- Thyroid hormone: thyroid hormone receptor (TR), subtypes
- Estrogen: estrogen receptor (ER), subtypes
- Testosterone: androgen receptor (AR)
- Progesterone: progesterone receptor (PR)
- Aldosterone: mineralocorticoid receptor (MR)
- Cortisol: glucocorticoid receptor (GR)

### Vitamins

- 1,25-(OH)2-vitamin D3: vitamin D receptor (VDR)
- All-trans-retinoic acid: retinoic acid receptor, subtypes
- 9-cis-retinoic acid: retinoid X receptor (RXR), subtypes

### Metabolic Intermediates and Products

- Oxysterols: liver X receptor (LXR), subtypes
- Bile acids: bile acid receptor (BAR)
- Fatty acids: peroxisome proliferator-activated receptor (PPAR), subtypes

### Xenobiotics

- Pregnancy X receptor (PXRx), constitutive androstane receptor (CAR)

Precursors of vitamin D are synthesized and stored in skin and activated by ultraviolet light; vitamin D can also be derived from dietary sources. Vitamin D is then converted in the liver to 25(OH) vitamin D and in the kidney to 1,25-(OH)2-vitamin D2, the most potent natural ligand of the vitamin D receptor (VDR). 1,25-(OH)2-vitamin D2 acts as a circulating endocrine hormone.

Vitamin A is stored in the liver and is activated by metabolism to all-trans-retinoic acid, which is a high affinity ligand for retinoic acid receptors (RARs). Retinoic acid is likely to function as a signaling molecule in paracrine as well as endocrine pathways. Retinoic acid is also converted to its 9-cis-isomer, which is a ligand for another nuclear receptor called the retinoid X receptor (RXR). These retinoids and their receptors are essential for normal life and development of multiple organs and tissues. They also have pharmaceutical utility for conditions ranging from skin diseases to leukemia.

Another "orphan receptor," bile acid receptor (BAR) also known as FXR, or "Farnesyl X receptor", is thus likely to play a role in regulation of bile synthesis and circulation in normal as well as disease states.

The peroxisome proliferator-activated receptors (PPARs) constitute another subfamily of nuclear receptors. There are three subtypes, and all are activated by polyunsaturated fatty acids. No single fatty acid has particularly high affinity for any PPAR, and it is possible that these receptors may function as integrators of the concentration of a number of fatty acids.

PPAR is expressed primarily in liver; to date, the natural ligand with highest affinity for PPAR is an eicosanoid, 8(S)-hydroxyeicosatetraenoic acid. The most potent PPAR ligands are the fibrate class of lipid-lowering pharmaceuticals. The name PPAR derives from the fact that compounds such as fibrates induce proliferation of hepatic peroxisomes, organelles involved in -oxidation of fatty acids.

The other PPARs ( and ) are structurally related but are not activated by peroxisome proliferators. PPAR- is ubiquitous, and its ligandother than fatty acidss are not well characterized. PPAR is expressed primarily in fat cells (adipocytes) and is necessary for differentiation along the adipocyte lineage. PPAR is also expressed in other cell types, including colonocytes, macrophages, and vascular endothelial cells, where it may play physiologic as well as pathologic roles. The natural ligand for PPAR is not known, although prostaglandin J derivatives have the highest affinity (in the micromolar range). It is exciting news that PPAR appears to be the target of thiazolidinedione antidiabetic drugs that improve insulin sensitivity. These pharmaceutical agents bind to PPAR with nanomolar affinities, and
non-thiazolidinedione PPAR ligands are also insulin sensitizers, further implicating PPAR in this physiologic role.

**Xenobiotics**

Other nuclear receptors appear to function as integrators of exogenous environmental signals, including natural **endobiotics** (e.g., medicinals and toxins found in plants) and xenobiotics (compounds that are not naturally occurring). In these cases, the role of the activated nuclear receptor is to induce cytochrome P450 enzymes that facilitate detoxification of potentially dangerous compounds in the liver. Receptors in this class include:

- SXR, or sterol and xenobiotic receptor
- CAR, or constitutive androstane receptor
- PPAR, which is also activated by certain environmental chemicals

Unlike other nuclear receptors that have high affinity for very specific ligands, xenobiotic receptors have low affinity for a large number of ligands, reflecting their function in defense from a varied and challenging environment. Although these xenobiotic compounds are clearly not "hormones" in the classical sense, the function of these nuclear receptors is consistent with the general theme of helping the organism to cope with environmental challenges.

**Orphan Receptors**

The nuclear receptor superfamily is one of the largest families of transcription factors. The hormones and vitamins just described account for the functions of only a fraction of the total number of nuclear receptors. The remainder have been designated as **orphan receptors** because their putative ligands are not known.

From analyses of mice and humans with mutations in various orphan receptors, it is clear that many of these receptors are required for life or development of specific organs ranging from brain nuclei to endocrine glands. Some orphan receptors appear to be active in the absence of any ligand ("constitutively active") and may not respond to a natural ligand. Nevertheless, all of the receptors now known to respond to metabolites and environmental compounds were originally discovered as orphans. Thus, it is likely that future research will find that additional orphan receptors function as receptors for physiologic, pharmacologic, or environmental ligands.

**Variant Receptors**

As to be discussed later, the carboxyl (C-) terminus of the nuclear receptors is responsible for hormone binding. In the case of a few nuclear receptors, including TR and the glucocorticoid receptor, alternative splicing leads to the production of variant receptors with unique C-termini that do not bind ligand. These variant receptors are normally expressed, but their biologic relevance is uncertain. It has been speculated that they modulate the action of the classical receptor to which they are related by inhibiting its function.

Another type of normally occurring variant nuclear receptors

Lacks a classical deoxyribonucleic acid (DNA) binding domain (see later). These include DAX-1, which is mutated in human disease, and SHP-1. Their ligands, if any, are not known, and it is likely that DAX-1 and SHP-1 bind to and repress the actions of other receptors.

Rare, naturally occurring mutations of hormone receptors can cause hormone resistance in affected patients, for instance:

1. Inheritance of the hormone resistance phenotype can be dominant if the mutant receptor inhibits the action of the normal receptor, as with generalized resistance to thyroid hormone.
2. Inheritance is recessive if the mutation results in a complete loss of receptor function, as with the syndrome of hereditary 1,25-dihydroxyvitamin D-resistant rickets.
3. Inheritance can be X-linked, as with the mutated androgen receptor in androgen insensitivity syndromes, including testicular feminization.
Regulation of Ligand Levels

Ligand levels can be regulated in a number of ways. A dietary precursor may not be available in required amounts, as occurs in hypothyroidism due to iodine deficiency. Pituitary hormones (e.g., thyroid-stimulating hormone) regulate the synthesis and secretion of classical thyroid and steroid hormones. When the glands that synthesize these hormones fail, hormone deficiency can occur.

Many of the nuclear receptor ligands are enzymatically converted from inactive prohormones to the biologically active hormone (e.g., 5' deiodination of thyroxine \( T_4 \) to triiodothyronine \( T_3 \)). In other cases, one hormone is precursor for another (e.g., aromatization of testosterone to estradiol). Biotransformation may occur in a specific tissue that is not the main target of the hormone (e.g., renal 1-hydroxylation of vitamin D) or may occur primarily in target tissues (e.g., 5-reduction of testosterone to DHT). Deficiency or pharmacologic inhibition of such an enzyme can also reduce hormone levels.

Hormones can be inactivated by standard hepatic or renal clearance mechanisms or by more specific enzymatic processes. In the latter case, reduction in enzyme activity due to gene mutations or pharmacologic agents can result in hormone excess syndromes, for example, the renal deactivation of cortisol by 11-hydroxysteroid dehydrogenase (11-OHSD). Since, as noted earlier, cortisol can activate the mineralocorticoid receptor, insufficient 11-OHSD activity due to licorice ingestion, gene mutation, or extremely high cortisol levels causes syndromes of apparent mineralocorticoid excess.
NUCLEAR RECEPTOR SIGNALING MECHANISMS

Nuclear receptors are multifunctional proteins that transduce the signals of their cognate ligands. General features of nuclear receptor signaling are illustrated in Figure 4-1.

First and foremost, the ligand and the nuclear receptor must get to the nucleus. The nuclear receptor must also bind its ligand with high affinity. Because a major function of the receptor is to selectively regulate target gene transcription, it must recognize and bind to promoter elements in appropriate target genes. One discriminatory mechanism is dimerization of a receptor with a second copy of itself or with another nuclear receptor. The DNA-bound receptor must also work in the context of chromatin to signal the basal transcription machinery to increase or decrease transcription of the target gene.

<table>
<thead>
<tr>
<th>TABLE 4-2 – Regulation of Nuclear Receptor Ligand Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precursor availability</td>
</tr>
<tr>
<td>Secretion</td>
</tr>
<tr>
<td>Deactivation (active hormone inactive hormone)</td>
</tr>
</tbody>
</table>

Throughout the following discussion on the mechanisms and regulation of signaling by nuclear receptors, it should be kept in mind that some basic mechanisms are generally used by many or all members of the nuclear receptor superfamily, whereas other mechanisms impart the specificity that is crucial to the vastly different biologic effects of the many hormones and ligands that utilize these related receptors.

Domain Structure of Nuclear Receptors

The nuclear receptors are proteins whose molecular weights are generally between 50,000 and 100,000 d. They all share a common series of domains, referred to as A to F (Fig. 4-2). This linear depiction of the receptors is useful for describing and comparing the receptors, but it does not capture the role of protein folding and tertiary structure in mediating the various receptor functions. As of this writing, no full-length nuclear hormone receptor has been crystallized, but structures of individual domains have been extremely revealing, as will be clear from the discussions of specific receptor functions that follow.
Nuclear Localization

The nuclear receptors, like all cellular proteins, are synthesized on ribosomes that reside outside the nucleus. Import of the nuclear receptors into the nucleus requires the nuclear localization signal (NLS), located near the border of the C and D domains (see Fig. 4-2). As a result of their nuclear localization signals, most of the nuclear receptors reside in the nucleus in the absence, as well as in presence, of ligand. A major exception is the glucocorticoid receptor (GR), which, in the absence of hormone, is tethered in the cytoplasm to a complex of chaperone molecules, including heat shock proteins (hsp90). Hormone binding to GR induces a conformational change that results in dissociation of the chaperone complex, thereby allowing the hormone-activated GR to translocate to the nucleus via its nuclear localization signal.
Hormone Binding

High-affinity binding of a lipophilic ligand is a shared characteristic of many nuclear receptors. This defining function of the receptor is mediated by the C-terminal ligand-binding domain (LBD), domains D and E in Figure 4-2. This region of the receptor also has many other functions, including dimerization and transcriptional regulation (see "Receptor Dimerization" and "Receptor Regulation of Gene Transcription" below).

The structure of the LBD has been solved for a number of receptors. All share a similar overall structure consisting of 12-helical segments in a highly folded tertiary structure (Fig. 4-3). The ligand binds within a hydrophobic pocket composed of amino acids in helix 3 (H3), H4, and H5. The major structural change induced by ligand binding is an internal folding of the most C-terminal helix (H12), which forms a cap on the ligand-binding pocket. Although the overall mechanism of ligand binding is similar for all receptors, the details are crucial in determining ligand specificity. Although the molecular details of ligand binding are beyond the scope of this chapter, this is the most critical determinant of receptor specificity.
Target Gene Recognition by Receptors

Another crucial specificity factor for nuclear receptors is their ability to recognize and bind to the subset of genes that are to be regulated by their cognate ligand. Target genes contain specific DNA sequences that are called hormone response elements (HREs). Binding to the HRE is mediated by the central C domain of the nuclear receptors (see Fig. 4-2). This region is typically composed of 66 to 68 amino acids, including two subdomains called zinc fingers because the structure of each subdomain is maintained by four cysteine residues that coordinate with a zinc atom.

The first of these zinc-ordered modules contains basic amino acids that contact DNA; as with the LBD, the overall structure of the DNA-binding domain (DBD) is very similar for all members of the nuclear receptor superfamily. The specificity of DNA binding is determined by multiple factors (Table 4-3). All steroid hormone receptors, except for the estrogen receptor (ER), bind to the double-stranded DNA sequence AGAACA (Fig. 4-4).

By convention, the double-stranded sequence is described by the sequence of one of the complementary strands, with the bases ordered from the 5’ to the 3’ end. Other nuclear receptors recognize the sequence AGGTCA. The primary determinant of this specificity is a group of amino acids residues in the so-called P-box of the DBD (see Fig. 4-4). These hexamer DNA sequences are referred to as half-sites. The only two differences between these hexameric half-sites are the central two base pairs (underlined). For some nuclear receptors, the C-terminal extension of the DBD contributes specificity for extended half-sites containing additional, highly specific DNA sequences 5’ to the hexamer (see Fig. 4-2).

Another source of specificity for target genes is the spacing and orientation of these half-sites, which in most cases are bound by receptor dimers.
Receptor Dimerization

As noted earlier, the nuclear receptor DBD has affinity for the hexameric half-site, or extended half-sites; many HREs, however, are composed of repeats of the half-site sequence, and most nuclear receptors bind such HREs as dimers. Steroid receptors, including ER, function primarily as homodimers, which preferentially bind to two half-sites oriented toward each other (inverted repeats) with three base pairs in between (IR3) (Fig. 4-4A). The major dimerization domain in steroid receptors is within the C-domain, although the LBD contributes. Ligand-binding facilitates dimerization and DNA binding of steroid hormone receptors. Most other receptors, including

\[ \text{Figure 4-3 Structural basis of nuclear receptor ligand binding and cofactor recruitment.} \]

TR, RAR, VDR, PPAR, LXR, and VDR, bind to DNA as heterodimers with RXR (Fig. 4-4B). Heterodimerization is mediated by two distinct interactions. The receptor LBD mediates the strongest interaction, which occurs even in the absence of DNA. These receptor heterodimers bind to two half-sites arranged as direct repeats (DRs) with variable numbers of base pairs in between.

The spacing of the half-sites is a major determinant of target gene specificity. This is due to the second receptor-receptor interaction, which involves the DBDs and is highly sensitive to the spacing of the half-sites. For example, VDR/RXR heterodimers bind preferentially to direct repeats separated by three bases (DR3 sites), TR/RXR binds DR4, and RAR/RXR binds DR5 with highest affinity. \[ \text{[30]} \]

The structural basis of this restriction on DNA binding is related to the fact that the RXR binds to the upstream half-site (farthest from the start of transcription). As a result of the periodicity of the DNA helix, each base pair separating the half-sites leads to a rotation of about 36° of one half-site relative to the other. Subtle differences in the structure of the receptor LBDs make the DBD interactions more or less favorable at the different degrees of rotation. \[ \text{[31]} \]
Nuclear receptors mediate a variety of effects on gene transcription. The most common modes of regulation (Table 4-4) are:

- Ligand-dependent gene activation
- Ligand-independent repression of transcription
- Ligand-dependent negative regulation of transcription

The remainder of this chapter describes these mechanisms.

**Ligand-Dependent Activation**

Ligand-dependent activation is the most well-understood function of nuclear receptors and their ligands. In this case, the ligand-bound receptor increases transcription of a target gene to which it is bound. The DBD serves to bring the receptor domains that mediate transcriptional activation to a specific gene. Transcriptional activation itself is mediated primarily by the LBD, which can function in the same way even when it is transferred to a DNA-binding protein that is not related to nuclear receptors. The *activation function* (AF) of the LBD is referred to as AF-2 (see Fig. 4-2).

Gene transcription is mediated by a large complex of factors that ultimately regulate the activity of ribonucleic acid (RNA) polymerase, the enzyme that uses the chromosomal DNA template to direct the synthesis of messenger RNA. Most mammalian genes are transcribed by RNA polymerase II, utilizing a large set of cofactor proteins, including *basal transcription factors*, and associated factors collectively referred to here as *general transcription factors* (GTFs). Details about GTFs are of fundamental importance and are available elsewhere. These GTFs bind to the nuclear receptor on DNA only when hormone or ligand is bound. Thus, these coactivators specifically recognize the ligand-bound conformation of the LBD.

The most important determinant of coactivator binding is the position of H12, which changes dramatically when ligands bind receptors (see Fig. 4-3). Along with H3, H4, and H5, H12 forms a hydrophobic cleft that is bound by short polypeptide regions of the coactivator molecules. These polypeptides, called *NR boxes*, have characteristic sequences of LxxLL, where L is leucine and xx can be any two amino acids. A number of coactivator proteins containing LxxLL

### Table 4-3: Determinants of Target Gene Specificity Of Nuclear Receptors

<table>
<thead>
<tr>
<th>Specificity</th>
<th>Region of Receptor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Binding to DNA</td>
<td>1. DNA-binding domain (DBD, C domain)</td>
</tr>
<tr>
<td>2. Binding to specific hexamer (AGGTCA vs. AGAACA)</td>
<td>2. P-box in C-domain</td>
</tr>
<tr>
<td>3. Binding to sequences 5' to hexamer</td>
<td>3. C-terminal extension of DBD</td>
</tr>
<tr>
<td>4. Binding to hexamer repeats</td>
<td>4. Dimerization domain (C domain for steroid receptors, D-E-F for others)</td>
</tr>
<tr>
<td>5. Recognition of hexamer spacing</td>
<td>5. Heterodimerization with retinoid X receptor (RXR) (non-steroid receptors, C domain)</td>
</tr>
</tbody>
</table>

### Table 4-4: Regulation of Gene Transcription By Nuclear Receptors

1. Ligand-dependent gene activation: DNA binding and recruitment of coactivators
2. Ligand-independent gene repression: DNA binding and recruitment of corepressors
3. Ligand-dependent negative regulation of gene expression: DNA binding and recruitment of corepressors or recruitment of coactivators off DNA

### Table 4-5: Nuclear Receptor Coactivators and Corepressors

**Coactivators**

1. Chromatin remodeling
   - Swi/Snf complex
2. Histone acetyl transferase
   - p160 family (SRC-1, GRIP-1, pCIP)
   - p300/CBP
   - pCAF (p300/CBP-associated factor)
3. Activation
   - TRAP/DRIP (thyroid receptor-associated proteins/D receptor interacting proteins)
negative response elements

The activity of the positively acting factors results in the observed negative regulation.

Interaction leads to removal of coactivators such as p300 and CBP from the other transcription factors that positively regulate the gene.

This leads to suppression of transcription.

Element of the gene for the subunit of thyroid-stimulating hormone (TSH), transcription is activated. Ligand binding recruits corepressors and HDAC to the TR and

Ligand-bound receptors recruit corepressors and HDAC activity to such binding sites.

This mechanism involves nuclear receptor binding to DNA binding sites that reverse the paradigm of ligand-dependent activation (negative response elements).

Ligand-bound receptors recruit corepressors and HDAC activity to such binding sites.

For example, when the unliganded TR binds to the negative response element of the gene for the subunit of thyroid-stimulating hormone (TSH), transcription is activated. Ligand binding recruits corepressors and HDAC to the TR and leads to suppression of transcription.

In other cases, it has been postulated that negative regulation may result from ligand binding to nuclear receptors that bind to other transcription factors without binding DNA. This

The receptor utilizes helices 3 to 5 to form the hydrophobic pocket, as in coactivator binding, but H12 does not promote and even hinders corepressor binding. This negative role of H12 highlights the role of the ligand-dependent change in the position of H12 as the switch that determines repression and activation by nuclear receptors. (see Fig. 4-5).

The transcriptional functions of N-CoR and SMRT are the opposite of those of the coactivators. The corepressors themselves do not possess enzyme activity but do interact directly with GTFs and further enhance their activities. An important complex that also links nuclear receptors to GTFs is the TRAP (TR-associated proteins) or DRIP (D receptor-associated proteins) complex.

It is possible that the recruitment of multiple HATs reflects different specificities for core histones and potentially other, nonhistone proteins. Some HATs also interact directly with GTFs and further enhance their activities. An important complex that also links nuclear receptors to GTFs is the TRAP (TR-associated proteins) or DRIP (D receptor-associated proteins) complex.

Repression of Gene Expression by Unliganded Receptor

Although DNA binding is ligand-dependent for steroid hormone receptors, other nuclear receptors are bound to DNA even in the absence of their cognate ligand. The unliganded DNA-bound receptor is not passively waiting for hormone; instead, it actively represses transcription of the target gene. This repression both "turns off" the target gene and amplifies the magnitude of the subsequent activation by hormone. For instance, if the level of gene transcription in the repressed state is 10% of the basal level in the absence of receptor, a hormone-activation to 10-fold above that basal level represents a 100-fold difference of transcription rate between hormone-deficient (repressed) genes and hormone-activated genes (Fig. 4-4).

In many ways, the molecular mechanism of repression is the mirror image of ligand-dependent activation. The unliganded nuclear receptor recruits negatively acting factors (corepressors) to the target gene. The two major corepressors are large (270 kd) proteins:

Nuclear receptor corepressor (N-CoR)
Silencing mediator for retinoid and thyroid receptors (SMRT)

N-CoR and SMRT specifically recognize the unliganded conformation of nuclear receptors and use an amphipathic helical sequence similar to the NR box of coactivators to bind to a hydrophobic pocket in the receptor.

For corepressors, the peptide responsible for receptor binding is called the CoRNR box and contains the sequence (I or L) xx (I or V)I (where I is isoleucine, L is leucine, V is valine, and xx represents any two amino acids). The receptor utilizes helices 3 to 5 to form the hydrophobic pocket, as in coactivator binding, but H12 does not promote and even hinders corepressor binding. This negative role of H12 highlights the role of the ligand-dependent change in the position of H12 as the switch that determines repression and activation by nuclear receptors.

The transcriptional functions of N-CoR and SMRT are the opposite of those of the coactivators. The corepressors themselves do not possess enzyme activity but do interact directly with GTFs to inhibit their transcriptional activities.

Repression and activation functions augmenting the dynamic range of transcriptional regulation by nuclear receptors. HRE, hormone response element.

The ligand-dependent switch between the repressed and activated receptor conformations explains how hormones activate gene expression. However, many of the most important gene targets of hormones are turned off in the presence of the ligand. This is referred to as ligand-dependent negative regulation of transcription, or transrepression, to distinguish it from the repression of basal transcription by unliganded receptors.

The mechanism of negative regulation is less well understood than ligand-dependent activation, and, indeed, there may be more than one mechanism. One mechanism involves nuclear receptor binding to DNA binding sites that reverse the paradigm of ligand-dependent activation (negative response elements).

Ligand-bound receptors recruit coactivators and HDAC activity to such binding sites.

For example, when the unliganded TR binds to the negative response element of the gene for the subunit of thyroid-stimulating hormone (TSH), transcription is activated. Ligand binding recruits corepressors and HDAC to the TR and leads to suppression of transcription.

In other cases, it has been postulated that negative regulation may result from ligand binding to nuclear receptors that bind to other transcription factors without binding DNA. This

TABLE 4-6 -- Factors Modulating Receptor Activity in Different Tissues

<table>
<thead>
<tr>
<th>Receptor concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ligand concentration</td>
</tr>
<tr>
<td>Ligand function (agonist, partial agonist, antagonist)</td>
</tr>
<tr>
<td>Concentrations and types of coactivators and corepressors</td>
</tr>
<tr>
<td>Phosphorylation state of nuclear receptor</td>
</tr>
</tbody>
</table>

interaction leads to removal of coactivators such as p300 and CBP from the other transcription factors that positively regulate the gene. In this model, inhibition of the activity of the positively acting factors results in the observed negative regulation.
Role of Other Nuclear Receptor Domains

The N-terminal A/B domain of the nuclear receptors is the most variable region among all members of the superfamily in terms of length and amino acid sequence. Even subtypes of the same receptor often have completely different A/B domains. The function of this domain is least well defined. It is not required for unliganded repression or ligand-dependent activation. In many receptors, the A/B domain contains a positive transcriptional activity, often referred to as AF-1 [see Fig. 4-2], that is ligand-independent but probably interacts with coactivators [56] and may influence the magnitude of activation by agonists or partial agonists (see later). This activation function is tissue-specific and tends to be more important for steroid hormone receptors, whose A/B domains are notably longer than those of other members of the superfamily. The F domain of the nuclear receptors is hypervariable in length and sequence, and its function is not known.

Cross-talk with Other Signaling Pathways

Hormones and cytokines that signal via cell surface receptors also regulate gene transcription, often by activating protein kinases that phosphorylate transcription factors such as cAMP-response element-binding protein (CREB). Such signals can also lead to phosphorylation of nuclear receptors. Multiple signal-dependent kinases can phosphorylate nuclear receptors, leading to conformational changes that regulate function. Phosphorylation can lead to changes in DNA binding, ligand binding, or coactivator binding; these variable consequences depend on the specific kinase, receptor, and domain of the receptor that is phosphorylated. The properties of coactivators and corepressor molecules are also regulated by phosphorylation.

Receptor Antagonists

Certain ligands function as receptor antagonists by competing with agonists for the ligand-binding site. In the case of steroid hormone receptors, the position of H12 in antagonist-bound conformation is not identical to that in the unliganded receptor or the agonist-bound receptor. H12, which itself has a sequence that resembles the NR box, binds to the coactivator-binding pocket and thereby prevents coactivator binding. This antagonist-bound conformation also favors corepressor binding to steroid hormone receptors.

Tissue-Selective Ligands

Some ligands function as antagonists in some tissues but as full or partial agonists in others. These selective receptor modulators include compounds such as tamoxifen, a selective estrogen receptor modulator (SERM). SERMs are estrogen receptor antagonists with respect to the functions of AF-2, including coactivator binding, and require the AF-1 function for their agonist activity. Such agonism, like AF-1 activity, tends to be tissue-specific and therefore has great therapeutic utility.

In addition to drugs, certain endogenous ligands (e.g., testosterone, DHT) also mediate tissue-specific effects. The molecular basis of tissue-specific activity is not well understood but is probably due to the expression or activity of transcriptional cofactors that differentiate between receptors bound to different ligands. Table 4-6 summarizes factors contributing to tissue-specificity of receptor activity.
References

Nuclear Receptors and Ligands


Nuclear Receptor Mutations


Structure and Function of Nuclear Receptors


Chapter 5 - Mechanism of Action of Hormones That Act at the Cell Surface

Allen Spiegel
Christin Carter-Su
Simeon Taylor

Hormones are secreted into the blood and act upon target cells at a distance from the secretory gland. In order to respond to a hormone, a target cell must contain the essential components of a signaling pathway. First, there must be a receptor to bind the hormone. Second, there must be an effector for example, an enzymatic activity that is regulated when the hormone binds to its receptor. Finally, there must be appropriate downstream signaling pathways to mediate the physiologic responses to the hormone. In fact, this type of mechanism involving receptors, effectors, and downstream signaling pathways is quite general and also functions in nonendocrine systems such as neurotransmitters, cytokines, and paracrine and autocrine factors. This chapter reviews several examples of endocrine signaling pathways, with particular attention to the molecular mechanisms that function in normal physiology and to the molecular pathology causing disease.
RECEPTORS

Definition and Classification

There are two essential functions that define hormone receptors: (1) the ability to bind the hormone and (2) the ability to couple hormone binding to hormone action. Both components of the definition are essential; for example, many hormones bind to binding proteins, which are distinct from receptors because the binding proteins do not trigger the signaling pathways that mediate hormone action.

Many classes of receptors are of interest in endocrinology. Some receptors are located within the cell and function as transcription factors (e.g., receptors for steroid and thyroid hormones). Other receptors are located on the cell surface and function primarily to transport their ligands into the cell by a process referred to as receptor-mediated endocytosis (e.g., low-density lipoprotein receptors). In this chapter, we focus upon cell-surface receptors that trigger intracellular signaling pathways. These cell-surface receptors can be classified according to the molecular mechanisms by which they accomplish their signaling function:

1. Ligand-gated ion channels (e.g., nicotinic acetylcholine receptor).
2. Receptor tyrosine kinases (e.g., receptors for insulin and insulin-like growth factor I).
3. Receptor serine/threonine kinases (e.g., receptors for activins and inhibins).
4. Receptor guanylate cyclase (e.g., atrial natriuretic factor receptor).
5. G protein-coupled receptors (e.g., receptors for adrenergic agents, muscarinic cholinergic agents, glycoprotein hormones, glucagon, and parathyroid hormone).
6. Cytokine receptors (e.g., receptors for growth hormone, prolactin, and leptin).

The receptors belonging to classes 1 to 4 are bifunctional molecules that can bind hormone and also serve as effectors by functioning either as ion channels or as enzymes. In contrast, the receptors belonging to classes 5 and 6 have the ability to bind the hormone but must recruit a separate molecule to catalyze the effector function. For example, as the name implies, G protein-coupled receptors utilize G proteins to regulate downstream effector molecules. Similarly, cytokine receptors recruit cytosolic tyrosine kinases (e.g., Janus family tyrosine kinases, JAKs) as effectors to trigger downstream signaling pathways.
Hormone Binding

As predicted by the fact that hormones circulate in relatively low concentrations in the plasma, the binding interaction between a hormone and its receptor is characterized by high binding affinity. Furthermore, hormone binding has a high degree of specificity. Generally, the receptor binds its cognate hormone more tightly than it binds other hormones. However, some receptors may bind structurally related hormones with lower affinity. For example, the insulin receptor binds insulin-like growth factors (IGFs) with approximately 100-fold lower affinity than it binds insulin. Similarly, the thyrotropin receptor binds human chorionic gonadotropin with lower affinity than it binds thyrotropin. This phenomenon has been referred to as specificity spillover and provides an explanation of several pathologic conditions, such as hypoglycemia caused by tumors secreting IGF-II and hyperthyroidism associated with choriocarcinoma.

Binding of a hormone (H) to its receptor (R) can be described mathematically as an equilibrium reaction:

\[ H + R \rightleftharpoons HR \]

At equilibrium, \( K_a = \frac{[HR]}{[H][R]} \), where \( K_a \) is the association constant for the formation of the hormone receptor complex (HR). As originally shown by Scatchard, it is possible to rearrange this equation in terms of the total concentration of receptor binding sites, \( R_0 = [R] + [RH] \), as follows:

\[ K_a = \frac{[RH]}{[R_0][RH][H]} \]

\[ \frac{[RH]}{[H]} = K_a \frac{R_0}{K_a} \]

A straight line is obtained when \( \frac{[RH]}{[H]} \) (i.e., the ratio of bound to free hormone) is plotted as a function of \( [RH] \) (the concentration of bound hormone). The slope of the line is \(-K_a\), and the line intercepts the horizontal axis at the point where \( [HR] = R_0 \), the total number of binding sites. This type of plot is referred to as a Scatchard plot and has been used as a graphic method to estimate the affinity with which a receptor binds its hormone. Although the binding properties of some receptors are described more or less accurately by these simple equations, other receptors exhibit more complex properties. This simple algebraic derivation of the Scatchard equation implicitly assumes that there is only one class of receptors and that the binding sites on the receptors do not interact with one another. If these assumptions do not apply to the interaction of a particular hormone with its receptor, the Scatchard plot may not be linear.

Several molecular mechanisms may contribute to nonlinearity of the Scatchard plot. For example, there may be more than one type of receptor that binds the hormone (e.g., a high-affinity, low-capacity site and a low-affinity, high-capacity site). Alternatively, some receptors have more than one binding site, and there may be cooperative interactions among the binding sites (e.g., the insulin receptor). In addition, the interaction between a G protein and a G protein-coupled receptor may affect the affinity with which the receptor binds its ligand; moreover, the effect on binding affinity depends on whether guanosine diphosphate (GDP) or guanosine triphosphate (GTP) is bound to the G protein. However, a detailed discussion of these complexities is beyond the scope of this chapter.
REGULATION OF HORMONE SENSITIVITY

Early in the history of endocrinology, attention was focused on the regulation of hormone secretion as the most important mechanism for regulating physiology. However, it has become apparent that the target cell is not passive. Rather, there are many influences that can alter the sensitivity of the target cell's response to a given concentration of hormone. For example, the number of receptors can be regulated. All things being equal, hormone sensitivity is directly related to the number of hormone receptors expressed on the cell surface. In addition, post-translational modifications of the receptor can modify either the affinity of hormone binding or the efficiency of coupling to downstream signaling pathways. Moreover, all of the downstream components in the hormone action pathway are subject to similar types of regulatory influences, which can have a significant impact on the ability of the target cell to respond to hormone.

Just as hormone sensitivity is subject to normal physiologic regulation, pathologic influences can cause disease by targeting components of the hormone action pathway. Multiple etiologic factors can impair the hormone action pathway, such as genetic, autoimmune, and exogenous toxins. For example, disease mechanisms can alter the functions of cell-surface receptors, effectors such as G proteins, and other components of the downstream signaling pathways. This chapter describes several examples illustrating these principles.
RECEPTOR TYROSINE KINASES

Receptor tyrosine kinases have several structural features in common: an extracellular domain containing the ligand-binding site, a single transmembrane domain, and an intracellular portion that includes the tyrosine kinase catalytic domain. Analysis of the sequence of the human genome suggests that there are approximately 100 receptor tyrosine kinases. The tyrosine kinase domain is the most highly conserved sequence among all the receptors in this family. In contrast, there is considerable variation among the sequences of the extracellular domains. Indeed, the family of receptor tyrosine kinases can be classified into 16 subfamilies, primarily on the basis of the differences in the structure of the extracellular domain. Furthermore, receptor tyrosine kinases mediate the biologic actions of a wide variety of ligands, including insulin, epidermal growth factor (EGF), platelet-derived growth factor (PDGF), and vascular endothelial cell-derived growth factor. The variation in the sequences of the extracellular domains enables the receptors to bind this structurally diverse collection of ligands.

The EGF receptor was the first cell-surface receptor demonstrated to possess tyrosine kinase activity and also the first receptor tyrosine kinase to be cloned. Like most receptor tyrosine kinases, the EGF receptor exists primarily as a monomer in the absence of ligand. However, binding of ligand induces receptor dimerization. As discussed later in this chapter, ligand-induced dimerization is central to the mechanism whereby the receptor mediates the biologic activity of EGF. In addition to the ability to form homodimers, the EGF receptor can form heterodimers with other members of the same subfamily of receptor tyrosine kinases. Because a small number of receptors can combine in a large number of pairings, heterodimers formation has the potential to fine-tune the specificity of receptors with respect to both ligand binding and downstream signaling.

The insulin receptor is of special interest to endocrinologists because diabetes is among the most common diseases of the endocrine system. Furthermore, the insulin receptor closely resembles the type 1 receptor for IGF1 and therefore also plays an important role in the physiology of growth hormone (GH) in vivo. Although the kinase domains of receptors for insulin and IGF-I closely resemble other receptor tyrosine kinases, at least two distinctive features set them apart. First, the receptors are synthesized as proreceptors that undergo proteolytic cleavage into two subunits and . The subunit contains the ligand-binding site; the subunit includes the transmembrane and tyrosine kinase domains. Second, both receptors exist as heterotetramers that are stabilized by intersubunit disulfide bonds. In contrast to other receptor tyrosine kinases, which are thought to dimerize in response to ligand binding, the insulin receptor exists as a dimer of monomers even in the absence of ligand. The remainder of this section reviews the molecular mechanisms whereby receptor tyrosine kinases mediate biologic action, with special emphasis on the insulin receptor as an illustrative example.

Receptor Activation: Role of Receptor Dimerization

Dimerization plays a central role in the mechanism whereby most receptor tyrosine kinases are activated by their cognate ligands. Although receptor dimerization is a common theme, the detailed molecular mechanisms differ from receptor to receptor. The following are three examples of the mechanisms of receptor dimerization.

### Dimeric Ligand

PDGF and vascular endothelial cell-derived growth factor are examples of dimeric ligands. Because each subunit of ligand can bind one receptor molecule, simultaneous binding of two receptor molecules drives receptor dimerization. Direct support for this type of mechanism is provided by the crystal structure of vascular endothelial cell-derived growth factor bound to its receptor (Fig. 5-1). As illustrated by the crystal structure of GH bound to its receptor, one molecule of ligand can bind two molecules of receptor. In fact, there are two distinct receptor-binding sites on each GH molecule, and this enables the ligand to promote receptor dimerization. This observation has an important implication for pharmacology. By abolishing one of the two receptor-binding sites, it is possible to design mutant ligands that lack the ability to promote receptor dimerization and therefore lack the ability to trigger hormone action. Nevertheless, by binding to receptors, the mutant ligand acquires the ability to inhibit the
action of the endogenous hormone. Such mutant GH molecules are being evaluated as potential therapeutic agents, for example, in conditions such as acromegaly.

Preexisting Receptor Dimers

The insulin receptor represents a paradox. The insulin receptor exists as a dimer even in the absence of ligand. (Actually, it is an \( \beta_2 \) heterotetramer, which is a dimer of monomers.) If the receptor is already dimerized, why is it not active? Although the molecular details remain to be elucidated, it seems likely that the two halves of the insulin receptor are not oriented in an optimal way to permit receptor activation in the absence of ligand. Perhaps, insulin binding triggers a conformational change that somehow mimics the effects of dimerization in other receptor tyrosine kinases. In any case, several studies have demonstrated that receptor dimerization is necessary for the ability of insulin to activate its receptor. For example, monomers retain the ability to bind insulin but are not activated in response to insulin binding.\(^{13}\)\(^{14}\) Furthermore, indirect evidence suggests that a single insulin molecule binds simultaneously to both subunits of the insulin receptor \(^{15}\)\(^{16}\); the ability to bind simultaneously to both halves of the dimeric receptor appears to be essential to the ability of insulin to activate its receptor.
Receptor Activation: Conformational Changes in the Kinase Domain

When ligand binds to the extracellular domain, it stimulates the tyrosine kinase activity of the intracellular domain. Although the detailed mechanisms of transmembrane signaling are not completely understood, considerable progress has been made in elucidating the molecular mechanisms of receptor activation. Investigations of the three-dimensional structure of the insulin receptor tyrosine kinase domain help to explain why the receptor is maintained in a low-activity state in the absence of insulin. In the inactive form of the insulin receptor kinase, Tyr1162 is located in a position so that it blocks protein substrates from binding to the active site. Furthermore, in the inactive state of the tyrosine kinase domain, the active site assumes a conformation that does not accommodate magnesium adenosine triphosphate (ATP). Thus, the tyrosine kinase is inactive because the active site cannot bind either of its substrates. How does insulin activate the receptor? Insulin binding triggers autophosphorylation of three tyrosine residues (Tyr1158, Tyr1162, and Tyr1163) in the "activation loop." When the three tyrosine residues in the activation loop become phosphorylated, an important conformational change occurs. As a result of the movement of the activation loop, the active site acquires the ability to bind both ATP and protein substrates. Thus, the conformational change induced by autophosphorylation activates the receptor to phosphorylate other substrates.

It remains unclear how this process is initiated. Because the inactive state of the tyrosine kinase cannot bind ATP, it seems unlikely that phosphorylation of Tyr1162 proceeds by a true autophosphorylation mechanism. Rather, it is likely that Tyr1162 in one subunit is transphosphorylated by the second subunit in the heterotetrameric molecule. However, this proposed mechanism poses a "chicken and egg" problem. It requires that at least one of the subunits is active before the Tyr residues in the activation loop become phosphorylated. Perhaps the activation loop is somewhat mobile so that some molecules of unphosphorylated tyrosine kinase can assume an active conformation and initiate a chain reaction of transphosphorylation and receptor activation.
Receptor Tyrosine Kinases Phosphorylate Other Intracellular Proteins

Once activated, tyrosine kinases are capable of phosphorylating other protein substrates. Several factors determine which proteins are phosphorylated under physiologic conditions within the cell.

Amino Acid Sequence Context of Tyr Residue

Tyrosine kinases do not exhibit strict specificity with respect to the amino acid sequence of the phosphorylation site. Nevertheless, most tyrosine phosphorylation sites are located in the vicinity of acidic amino acid residues (i.e., Glu or Asp). \[2\]

Binding to the Tyrosine Kinase

Some protein substrates bind directly to the intracellular domain of the receptor. The binding interaction brings the substrate into close proximity to the kinase, thereby promoting phosphorylation of the substrate. For example, the insulin receptor substrate (IRS) proteins are characterized by a highly conserved phosphotyrosine-binding (PTB) domain that binds to a conserved motif (Asn-Pro-Xaa-pTyr) in the juxtamembrane domain of the insulin receptor. \[23\] \[24\] \[25\] Binding of the PTB domain to the insulin receptor requires phosphorylation of the Tyr residue in the Asn-Pro-Xaa-pTyr motif. This provides another mechanism (in addition to activation of the intrinsic receptor tyrosine kinase) whereby autophosphorylation of the receptor enhances phosphorylation of IRS proteins. Similarly, substrates for some tyrosine kinases contain Src homology 2 (SH2) domains, highly conserved domains that bind phosphotyrosine residues (see later). For example, the activated PDGF receptor contains a phosphotyrosine residue near its C-terminus that binds the SH2 domain of phospholipase C. This enables the PDGF receptor to phosphorylate and activate phospholipase C. \[2\] \[26\]

Subcellular Localization

Because receptor tyrosine kinases are located in the plasma membrane, they are in close proximity to other plasma membrane proteins. This colocalization has the potential to promote phosphorylation. For example, the insulin receptor has been reported to phosphorylate pp120/hepatocyte antigen-4 (HA4). \[27\] \[28\] Like the insulin receptor, pp120/HA4 is an integral membrane glycoprotein associated with the plasma membrane of hepatocytes. Similarly, FGF receptor substrate-2 (FRS2), a substrate of the fibroblast-derived growth factor receptor, is targeted to the plasma membrane by an N-terminal myristoylation site. \[29\]
There are at least two distinct mechanisms whereby tyrosine phosphorylation regulates protein function. First, tyrosine phosphorylation can induce a conformational change in a protein, thereby altering its function. For example, as discussed earlier, phosphorylation of the three Tyr residues in the activation loop of the insulin receptor changes the conformation of the active site, thereby facilitating binding of substrates and activating the receptor tyrosine kinase. However, most of the effects of tyrosine phosphorylation on protein function are mediated indirectly by regulating protein-protein interactions. In order to understand how tyrosine phosphorylation regulates protein-protein interactions, it is useful to review the biochemistry of c-src, the prototype of a nonreceptor tyrosine kinase. When the amino acid sequence of c-src is analyzed, it is apparent that there are three highly conserved domains in the molecule: the kinase catalytic domain and two noncatalytic domains that are referred to as src homology domains 2 and 3 (SH2 and SH3, respectively).

**SH2 Domains**

SH2 domains consist of conserved sequences (approximately 100 amino acid residues) that are present in many proteins that function in signaling pathways. From a functional point of view, SH2 domains share the ability to bind pTyr residues. However, individual SH2 domains vary with respect to their binding specificity. The binding affinity of an SH2 is determined by the three amino acid residues downstream from the pTyr residue. For example, the SH2 domains of phosphatidylinositol (PI) 3-kinase exhibit a preference for pTyr-(Met/Xaa)-Xaa-Met, whereas the SH2 domain of growth factor receptor binding protein 2 (Grb-2) prefers to bind pTyr-Xaa-Asn-Xaa. Thus, a given SH2 domain binds to a tyrosine-phosphorylated protein if and only if the pTyr residue is located in a context that corresponds to the binding specificity of the SH2 domain.

**SH3 Domains**

SH3 domains consist of conserved sequences (approximately 50 amino acid residues) that bind to proline-rich sequences. Like SH2 domains, SH3 domains are found in many proteins that function in signaling pathways.
Downstream Signaling Pathways

Receptor tyrosine kinases mediate the action of a wide variety of ligands in a wide variety of cell types. The bewildering complexity of the downstream signaling pathways corresponds to the huge number of physiologic processes that are regulated by receptor tyrosine kinases. Although it is beyond the scope of this chapter to attempt an encyclopedic review of all the downstream signaling pathways, we have selected examples to illustrate general principles.

As discussed earlier, the activated insulin receptor phosphorylates multiple substrates including IRS-1, IRS-2, IRS-3, and IRS-4. Each of these substrates contains multiple tyrosine phosphorylation sites, many of which correspond to consensus sequences for SH2 domains in important signaling molecules. Thus, IRS proteins serve as docking proteins that bind SH2 domain-containing proteins. Among these, two of the most important are PI 3-kinase and Grb-2. As discussed subsequently, binding of SH2 domains triggers multiple downstream signaling pathways.

Phosphatidylinositol 3-Kinase

The catalytic subunit of PI 3-kinase (p110; molecular mass approximately 110,000) is bound to a regulatory subunit. The classical isoforms of the regulatory subunit (p85; molecular mass approximately 85,000) contain two SH2 domains, both of which bind to pTyr in the context of pTyr-(Met/Xaa)-Xaa-Met motifs. Binding of pTyr residues to both SH2 domains of p85 leads to maximal activation of PI 3-kinase catalytic activity. (Submaximal activation can be achieved with occupancy of a single SH2 domain in p85.) Because all four IRS molecules (IRS-1, IRS-2, IRS-3, and IRS-4) contain multiple tyrosine phosphorylation sites that conform to the Tyr-(Met/Xaa)-Xaa-Met consensus sequence, insulin-stimulated phosphorylation promotes binding of IRS proteins to the SH2 domains in the regulatory subunit PI 3-kinase, thereby increasing the enzymatic activity of the catalytic subunit. Activation of PI 3-kinase triggers activation of a cascade of downstream kinases, beginning with phosphoinositide-dependent kinases 1 and 2. These phosphoinositide-dependent kinases phosphorylate and activate multiple downstream protein kinases including protein kinase B and atypical isoforms of protein kinase C.

A large body of evidence demonstrates that the pathways downstream from PI 3-kinase mediate the metabolic activities of insulin (e.g., activation of glucose transport into skeletal muscle, activation of glycogen synthesis, and inhibition of transcription of the phosphoenolpyruvate carboxykinase gene). Among other lines of evidence, PI 3-kinase inhibitors (e.g., LY294002 and wortmannin) block the metabolic actions of insulin. Similarly, overexpression of dominant negative mutants of the p85 regulatory subunit of PI 3-kinase also inhibits the metabolic actions of insulin. Although it is generally agreed that activation of PI 3-kinase is necessary, it is controversial whether it is sufficient to trigger the metabolic actions of insulin. For example, a second parallel pathway may also be required. The latter pathway involves tyrosine phosphorylation of Cbl, another protein that can be phosphorylated by the insulin receptor in some cell types.

Grb-2 and the Activation of Ras

Grb-2 is a short adaptor molecule that contains an SH2 domain capable of binding to pTyr residues in several signaling molecules, for example, IRS-1 and Shc, another PTB domain-containing protein that is phosphorylated by several receptor tyrosine kinases including the insulin receptor. The SH2 domain of Grb-2 is flanked by two SH3 domains which bind to proline-containing sequences in mSos (the mammalian homologue of Drosophila son-of-sevenless). mSos is capable of activating Ras, a small G protein that plays an important role in intracellular signaling pathways. mSos activates Ras by catalyzing the exchange of GTP for GDP in the guanine nucleotide-binding site of Ras. This, in turn, triggers the activation of a cascade of serine/threonine-specific protein kinases including Raf, mitogen-activated protein/extracellular signal-regulated kinase (MEK), and mitogen-activated protein (MAP) kinase. These pathways downstream from Ras contribute to the ability of tyrosine kinases to promote cell growth and regulate the expression of various genes.

We have focused on the signaling pathways downstream from the insulin receptor because of the importance of insulin and IGF-I in endocrinology. In many ways, the molecular mechanisms closely resemble those downstream from other receptor tyrosine kinases. However, the insulin signaling pathway is atypical in at least one respect. The insulin receptor phosphorylates docking proteins (e.g., IRS-1), which bind SH2 domain-containing proteins (e.g., PI 3-kinase and Grb-2). In contrast, the intracellular domains of most receptor tyrosine kinases contain binding sites for SH2 domains. For example, the SH2 domain of Grb-2 binds to pTyr716 in the activated PDGF receptor. Similarly, the PDGF receptor contains two Tyr-(Met/Xaa)-Xaa-Met motifs in the kinase insert domain that bind to the two SH2 domains in the p85 subunit of PI 3-kinase. It is not clear why some tyrosine kinases (e.g., the PDGF receptor) activate PI 3-kinase through a direct binding interaction, whereas others (e.g., the insulin receptor) utilize an indirect mechanism involving docking proteins. However, in contrast to PDGF receptors, which are associated with the plasma membrane, IRS proteins appear to be associated with the cytoskeleton. Perhaps this differential subcellular localization contributes to signaling specificity. In other words, if insulin and PDGF receptors trigger translocation of PI 3-kinase to different locations within the cell, this compartmentation may permit two different receptors to elicit different biologic responses even though both responses are mediated by the same signaling molecule (i.e., PI 3-kinase).
Off Signals: Termination of Hormone Action

Just as there are complex biochemical pathways that mediate hormone action, there are also mechanisms to terminate the biologic response. The necessity for these mechanisms is illustrated by the following example. After we eat a meal, the concentration of plasma glucose increases. This elicits an increase in insulin secretion, which in turn leads to a decrease in plasma glucose levels. If these processes went on unchecked, the level of glucose in the plasma would eventually fall so low that it would lead to symptomatic hypoglycemia. How is insulin action terminated? The answers to this question are not yet entirely clear, but several mechanisms contribute to turning off the insulin signaling pathway.

Receptor-Mediated Endocytosis

Insulin binding to its receptor triggers endocytosis of the receptor. Although most of the internalized receptors are recycled to the plasma membrane, some receptors are transported to lysosomes, where they are degraded. As a result, insulin binding accelerates the rate of receptor degradation, thereby down-regulating the number of receptors on the cell surface. Furthermore, endosomes contain proton pumps, which acidify the lumen; the acidic pH within the endosome promotes dissociation of insulin from its receptor. Ultimately, insulin is transported to the lysosome for degradation. In fact, receptor-mediated endocytosis is the principal mechanism whereby insulin is cleared from the plasma. Binding of ligands to other receptor tyrosine kinases also triggers receptor-mediated endocytosis by similar mechanisms.

Protein Tyrosine Phosphatases

Protein phosphorylation is a dynamic process. Tyrosine kinases catalyze the phosphorylation of tyrosine residues, but there are also protein tyrosine phosphatases (PTPases) to remove the phosphates. Thus, PTPases antagonize the action of tyrosine kinases. Studies with knockout mice have demonstrated that the absence of PTPase-1B is associated with increased insulin sensitivity and also protects against weight gain. Nevertheless, the human genome encodes a large number of PTPases, and it is an important goal of research to elucidate their physiologic functions. If one could develop selective inhibitors of the PTPases that oppose the effects of the insulin receptor tyrosine kinase, it is possible that these inhibitors would provide novel therapies for diabetes.

![Figure 5-4](image-url) Simplified model of signaling pathways downstream from the insulin receptor. Insulin binds to the insulin receptor, thereby activating the receptor tyrosine kinase to phosphorylate tyrosine residues on insulin receptor substrates (IRSs) including IRS-1 and IRS-2. Consequently, phosphotyrosine residues in IRS molecules bind to Src homology 2 (SH2) domains in molecules such as growth factor receptor-binding protein 2 (Grb-2) and the p85 regulatory subunit of phosphatidylinositol (PI) 3-kinase. These SH2 domain-containing proteins initiate two distinct branches of the signaling pathway. Activation of PI 3-kinase leads to activation of phosphoinositide-dependent kinases (PDKs) 1 and 2, which activates multiple protein kinases including Akt/protein kinase B, atypical protein kinase C (PKC) isoforms, and serum/glucocorticoid-activated protein kinases (Sgk). Grb-2 interacts with m-SOS, a guanine nucleotide exchange factor that activates Ras. Activation of Ras triggers a cascade of protein kinases leading to the activation of mitogen-activated protein (MAP) kinase.

Serine/Threonine Kinases

Most receptor tyrosine kinases, including the insulin receptor, are substrates for phosphorylation by Ser/Thr-specific protein kinases. Interestingly, the Ser/Thr phosphorylation appears to inhibit the action of the tyrosine kinase. Similarly, other phosphorytyrosine-containing proteins are subject to inhibitory influences of Ser/Thr phosphorylation resistance. For example, it has been reported that Ser/Thr phosphorylation of IRS-1 may inhibit insulin action, thereby causing insulin resistance.
Mechanisms of Disease

The simplest forms of endocrine disease are caused by either a deficiency or an excess of a hormone. However, hormone resistance syndromes resulting from defects in the signaling pathways can masquerade as hormone deficiency states. Similarly, diseases associated with constitutively activated receptors can mimic a state of hormone excess. In some cases, the abnormality in hormone action is genetic in origin, resulting from a mutation in a gene encoding one of the proteins in the signaling pathway. Similar syndromes can also be caused by other mechanisms; for example, there are autoimmune syndromes caused by autoantibodies directed against cell-surface receptors. These clinical syndromes illustrate the principle that understanding the biochemical pathways of hormone action can provide important insights into the pathophysiology of human disease.

Genetic Defects in Receptor Function

At least two distinct major types of genetic defects can cause hormone resistance. First, mutations can lead to a decrease in the number of receptors. For example, in the case of the insulin receptor, mutations have been identified that decrease receptor number by at least three mechanisms: (1) impairing receptor biosynthesis, (2) inhibiting the transport of receptors to their normal location in the plasma membrane, and (3) accelerating the rate of receptor degradation. Second, mutations can impair the intrinsic activities of the receptor. In the case of the insulin receptor, mutations have been reported that decrease the affinity of insulin binding or inhibit receptor tyrosine kinase activity.

Receptor dimerization is known to play a central role in the mechanisms whereby ligands activate many cell-surface receptors. This role has been shown most convincingly in the case of the GH receptor (a member of the family of cytokine receptors) but has also been postulated for receptor tyrosine kinases. The syndromes of multiple endocrine neoplasia types 2A and 2B and familial medullary carcinoma of the thyroid are caused by mutations in the gene encoding the ret tyrosine kinase (a subunit of the receptor for glial cell-derived growth factor). Ordinarily, there are cysteine residues in the extracellular domain of Ret that participate in the formation of intramolecular disulfide bonds. Mutation of one of the cysteine residues leaves an unpaired cysteine residue that domain of Ret that participate in the formation of intramolecular disulfide bonds. Mutation of one of the cysteine residues leaves an unpaired cysteine residue that promotes dimerization of Ret molecules, thereby activating the Ret receptor tyrosine kinase. Activation of the Ret tyrosine kinase through this germ line mutation converts Ret into an oncogene.

Autoantibodies Directed against Cell-Surface Receptors

Inhibitory antireceptor autoantibodies were first identified in myasthenia gravis. In this neuromuscular disease, antibodies to the nicotinic acetylcholine receptor impair neuromuscular transmission, apparently by accelerating receptor degradation. Subsequently, autoantibodies to the insulin receptor were demonstrated to block insulin action in the syndrome of type B extreme insulin resistance. Insulin resistance is caused by at least two mechanisms: (1) the antireceptor antibodies inhibit insulin binding to the receptor, and (2) the antibodies accelerate receptor degradation.

Graves’ disease provided the first example of stimulatory antireceptor autoantibodies. In Graves’ disease, there are autoantibodies directed against the thyroid-stimulating hormone (TSH) receptor. These antireceptor antibodies activate the TSH receptor, thereby stimulating growth of the thyroid gland as well as hypersecretion of thyroid hormone. This “experiment of nature” demonstrates that the receptor can be activated by ligands other than the physiologic ligand and that the normal spectrum of biologic actions can be triggered by this unphysiologic ligand (i.e., the antireceptor antibody). Similarly, antibodies to the insulin receptor have been demonstrated to activate the insulin receptor by mimicking insulin action. Although it is more common for a patient with anti-insulin receptor autoantibodies to present with insulin resistance, patients with anti-insulin receptor autoantibodies have also been reported to experience fasting hypoglycemia.
RECEPTORS THAT SIGNAL THROUGH ASSOCIATED TYROSINE KINASES

Overview

Members of the cytokine family of receptors resemble receptor tyrosine kinases in their mechanism of action, with one important difference. Instead of the tyrosine kinase being intrinsic to the receptor, enzymatic activity resides in a protein that associates with the cytokine receptor. As with receptor tyrosine kinases, ligand binding to the cytokine receptor activates the associated kinase. The more than 25 known ligands that bind to members of the cytokine receptor family have diverse functions. Three of the ligands are hormones: GH, which is vital for normal body height; prolactin (PRL), which is required for reproduction and lactation; and leptin, which is a potent appetite suppressant and a regulator of rates of metabolism. Other ligands of cytokine receptors, for example, erythropoietin, most interleukins, and interferons, and, regulate hematopoiesis or the immune response. A number of genetic diseases can be traced to defects in cytokine receptors. For example, Laron dwarfism is caused by autosomal recessive mutations of the GH receptor and autosomal recessive mutations of the leptin receptor can cause morbid obesity.
Cytokine Receptors Are Composed of Multiple Subunits

Members of the cytokine family of receptors share homology in both the extracellular and cytoplasmic domains. Some cytokine receptors, including the receptors for GH, PRL, and leptin, are thought to be composed of dimers of a single receptor subunit (Fig. 5-6). One ligand is thought to bind to both receptor subunits, as discussed earlier for the GH receptor. However, most cytokine receptors are composed of two or more different subunits, with as many as six subunits constituting a single receptor. Some of these receptors are thought to bind ligand dimers. One or more of these receptor subunits is shared by receptors for other cytokines. This phenomenon of "mixing and matching" receptor subunits is an efficient way for the cell to fine-tune its cellular responses and increase the number of ligands a group of receptor subunits can bind. For example, a receptor composed of gp130 and leukemia inhibitory factor receptor subunit binds leukemia inhibitory factor, a pleiotropic cytokine with multiple functions that appears to serve as a molecular interface between the neuroimmune and endocrine systems. The same receptor subunits, when combined with a ciliary neurotrophic factor receptor subunit, show a preference for ciliary neurotrophic factor, a trophic factor for motor neurons in the ciliary ganglion and spinal cord and a potent appetite suppressor. Combine two gp130 subunits with an interleukin-6 (IL-6) receptor subunit, and the new receptor shows a preference for IL-6, an inducer of the acute phase response with additional anti-inflammatory properties.
Cytokine Receptors Activate Members of the Janus Family of Tyrosine Kinases

Members of the cytokine family of receptors do not themselves exhibit enzymatic activity. Rather, they bind members of the Janus family of tyrosine kinases (JAKs) through a proline-rich region (see Fig. 5-6). There are four known JAKs, designated JAK1, JAK2, JAK3, and TYK2. As do the cytokine receptors, the JAKs mix and match in that some receptors show a strong preference for a single JAK, some require two different JAKs, and others appear to activate multiple JAK family members. For example, GH, PRL, and leptin preferentially activate JAK2. Interferon- activates JAK1 and JAK2, and IL-2 activates JAK1 and JAK3.

Binding of ligand to a cytokine receptor activates the appropriate JAK family member or members. In some cases (e.g., PRL), the JAKs appear to be constitutively associated with the cytokine receptor and ligand binding increases their activity. In other cases (e.g., the GH receptor), ligand binding increases both the affinity of JAKs for the cytokine receptor and the activity of the associated JAKs. Activation of JAKs requires receptor oligomerization, presumably to bring two or more JAKs into sufficiently close proximity to transphosphorylate each other on the activating tyrosine in the kinase domain, as described previously in the chapter for the receptor tyrosine kinases. Although receptor dimerization appears to be required for receptor activation, a conformational change in receptor may also be required.

Transphosphorylation is believed to cause a conformational change that exposes the ATP- or substrate-binding site, or both. Once the JAKs are activated, they phosphorylate themselves and their associated receptor subunits on multiple tyrosines. JAKs appear to be vital for normal human function. Mutations in the JAK3 gene have been linked to an autosomal recessive form of severe combined immunodeficiency disease. Targeted disruption of the JAK2 gene in mice is embryonic lethal.
Signaling Pathways Initiated by Cytokine Receptor-JAK Complexes

Phosphorylated tyrosines within the cytokine receptor subunits and their associated JAKs form binding sites for various signaling proteins containing phosphotyrosine binding domains, such as SH2 and PTB domains. Each cytokine receptor-JAK complex would be expected to have some tyrosine-containing motifs shared with many other cytokine receptor-JAK complexes (e.g., tyrosines within JAKs) and some specific tyrosine-containing motifs (e.g., tyrosines within a specific combination of receptor subunits). Thus, ligand binding to cytokine receptors would be expected to initiate some signaling pathways that are shared by many cytokines and some that are more specialized to a particular cytokine receptor. The signaling proteins known to be recruited to subsets of cytokine receptor-JAK complexes are generally the same as those recruited to receptor tyrosine kinases. Examples include the IRS proteins, the adapter proteins Shc and Grb-2 that lead to activation of the Ras-MAP kinase pathway, phospholipase C, and PI 3-kinase. However, there is one family of signaling proteins that appears to be particularly important for the function of cytokine signal transducers and activators of transcription (STATs) (Fig. 5-7).

STAT proteins are latent cytoplasmic transcription factors. STATs bind, through their SH2 domains, to one or more phosphorylated tyrosines in activated receptor-JAK complexes. Once bound, they themselves are tyrosyl phosphorylated, presumably by the receptor-associated JAKs. STATs then dissociate from the receptor-JAK complexes, homodimerize or heterodimerize with other STAT proteins, move to the nucleus, and bind to gamma-activated sequence-like elements in the promoters of cytokine-responsive genes. The transcriptional response depends on how many STAT binding sites exist in the receptor-JAK complex, with which of the seven known STATs a particular STAT heterodimerizes, to what other proteins a particular STAT binds, the degree of serine or threonine phosphorylation of the STAT, and what other transcription factors are also activated. For example, leukemia inhibitory factor, whose receptor contains seven STAT3 binding motifs (YXXQ, where Y = tyrosine, X = any amino acid, and Q = glutamine) is a particularly potent activator of STAT3. The transcriptional activity of STAT5 is enhanced by its forming a complex with the glucocorticoid receptor.
Precise Regulation of the Cytokine Receptors Is Required for Normal Function

Ligand binding to cytokine receptors normally activates JAKs rapidly and transiently. Conversely, constitutively activated JAKs and STATs are associated with cellular transformation. For example, in cells transformed by the Abl oncoprotein v-Abl, JAK1 is constitutively activated and inhibition of JAK1 blocks the ability of v-Abl to transform bone marrow cells. Constitutively active JAKs and STATs are a common characteristic of leukemias, and both JAK2 and STAT5b have been identified as fusion partners in translocations in leukemias. The Tel-JAK2 fusion protein is constitutively active, leading to constitutively active STAT proteins. Thus, an understanding of what turns off cytokine receptor signaling is of utmost importance in understanding normal signaling through cytokine receptors.

As with the receptor tyrosine kinases, several steps have been hypothesized to serve as points of signal termination for cytokine signaling. These include receptor degradation (e.g., through a ubiquitination-proteosome pathway) and dephosphorylation of tyrosines within JAK or receptor (e.g., by an SH2 domain containing tyrosine phosphatase recruited to receptor-JAK complexes). The suppressors of cytokine-signaling (SOCSs) are thought to be particularly important players in the termination or suppression of cytokine-signaling pathways. SOCS proteins are an excellent example of an effective negative feedback loop. They are generally synthesized in response to cytokines. The newly synthesized SOCS proteins in turn bind, through their SH2 domain, to phosphorylated tyrosines within the cytokine receptor-JAK complex and inhibit further cytokine signaling. In some cases (i.e., SOCS1), SOCS proteins are thought to bind to phosphotyrosines in the kinase domain of JAK and inhibit kinase activity. In other cases (i.e., SOCS 3), SOCS proteins bind to phosphorylated tyrosines in the receptor and block STAT binding and activation. SOCS proteins can also be synthesized in response to noncytokine receptors, suggesting a mechanism whereby prior exposure to one ligand suppresses subsequent responses to another. For example, SOCS proteins have been implicated in the well-known ability of endotoxin to cause resistance to GH.

![Figure 5-6](image-url) Cytokine receptors are composed of multiple subunits and bind to one or more members of the Janus kinase (JAK) family of tyrosine kinases. A, Growth hormone (GH), like prolactin and leptin, binds to receptor homodimers and activates JAK2. B, Interferon (IFN) homodimers bind to their ligand-binding R1 subunits. The R2 subunits are then recruited, leading to activation of JAK1, which binds to R1 subunit, and JAK2, which binds to R2 subunit. Both subunits and both JAKs are necessary for responses to IFN. C, Interleukin-2 (IL-2) binds to receptors composed of three subunits: a common subunit shared with receptors for ILs 4, 7, 9, and 15; an IL-2R subunit shared with the IL-15 receptor; and a noncytokine receptor subunit, IL-2R. Extracellular regions of homology are indicated by the black lines and patterns. Intracellular regions of homology are indicated by the white boxed.
Hormones, growth factors, and cytokines that bind to members of the cytokine family of receptors activate JAK family tyrosine kinases. The activated kinases in turn phosphorylate tyrosines in themselves and associated receptors. The phosphorylated tyrosines form binding sites for other signaling proteins, including STAT proteins and a variety of other phosphotyrosine-binding proteins. STAT proteins promote the regulation of cytokine-sensitive genes, including SOCS proteins that serve a negative feedback function of terminating ligand activation of JAKs or STATs, or both.

Although this gives the general picture, it should be recognized that the picture is becoming much more complex every day. For example, there are reports that members of the Src family of tyrosine kinases can also be activated by some cytokine receptors (e.g., PRL receptor), that some JAK-binding proteins (e.g., SH2-B) are potent activators of JAK2, and that other proteins contribute to the down-regulation of cytokine-signaling pathways, including protein inhibitors of activated STAT (PIAS), that bind and inhibit specific STATs.
G PROTEIN-COUPLED RECEPTORS

Overview

G protein-coupled receptors (GPCRs) are an evolutionarily conserved gene superfamily with members in all eucaryotes from yeast to mammals. They transduce a wide variety of extracellular signals including photons of light; chemical odorants; divalent cations; monoamine, amino acid, and nucleoside neurotransmitters; lipids; and peptide and protein hormones. All members of the GPCR superfamily share a common structural feature, seven membrane-spanning helices, but various subfamilies diverge in primary amino acid sequence and in the domains that serve in ligand binding, G protein coupling, and interaction with other effector proteins (Fig. 5-8).

All GPCRs act as guanine nucleotide exchange factors. In their activated (agonist-bound) conformation, they catalyze exchange of GDP tightly bound to the subunit of heterotrimeric G proteins for GTP (Fig. 5-9). This in turn leads to activation of the subunit and its dissociation from the G protein dimer. Both G protein subunits are capable of regulating effector activity.

Effectors include enzymes of second messenger metabolism such as adenyl cyclase and phospholipase C- and a variety of ion channels. Agonist binding to GPCRs thus alters intracellular second messenger and ion concentrations with resultant rapid effects on hormone secretion, muscle contraction, and a variety of other physiologic functions. Long-term changes in gene expression are also seen as a result of second messenger-stimulated phosphorylation of transcription factors.

The G protein subunits are encoded by three distinct genes. The subunit binds guanine nucleotides with high affinity and specificity and has intrinsic guanosine triphosphatase (GTPase) activity. The and polypeptides are tightly but noncovalently associated in a functional dimer subunit. The three-dimensional structures of the individual and associated subunits have been determined.

There is considerable diversity in G protein subunits, with multiple genes encoding all three subunits and alternative gene splicing resulting in additional polypeptide products. There are at least 16 distinct subunit genes in mammals. These vary widely in range of expression. Some such as Gs- which couples many GPCRs to stimulation of adenyl cyclase, are ubiquitous; others such as Gt-, which couples the GPCR rhodopsin to cyclic guanosine monophosphate phosphodiesterase in retinal rod photoreceptor cells, are highly localized.

Because multiple distinct GPCRs, G proteins, and effectors are expressed within any given cell, the degree and basis for specificity in G protein coupling to GPCRs and to effectors are major subjects of investigation with implications for drug action and disease mechanisms. Since the pioneering work of Rodbell in discovering G proteins and showing that G protein-mediated signal transduction involves three separable components (receptor, G protein, and effector), additional complexity has emerged.

A large new gene family termed RGS (for regulators of G protein signaling) has been identified. RGS proteins bind to a transition state of the GTP-activated G protein subunit and accelerate its GTPase activity, thus helping deactivate the subunit. RGS domains have also been found in modular proteins with additional functions, in certain cases linking heterotrimeric G protein signaling with the function of low-molecular-weight GTP-binding proteins in the ras superfamily.

Lefkowitz has shown that a family of GPCR kinases and of arrestin proteins is involved in GPCR desensitization after agonist binding. In addition, it is now clear that GPCRs interact directly with a number of other proteins in addition to G proteins. Not only are GPCRs important targets for treatment of many diseases, but also mutations in genes encoding GPCRs have been identified as the cause of a number of endocrine as well as nonendocrine disorders.
G Protein-Coupled Receptor Structure and Function

Structure

Hydropathy analysis of the primary sequence of all GPCRs predicts seven membrane-spanning helices connected by three intracellular loops and three extracellular loops with an extracellular amino terminus and an intracellular carboxyl terminus (see Fig. 5-8). This basic structure has now been verified by x-ray crystallography for rhodopsin. Although there was already evidence that visual transduction in the retina and hormone activation of adenylyl cyclase share common features, the discovery that the adrenergic receptor has the same topographic structure as rhodopsin came as a surprise. Cloning of the complementary deoxyribonucleic acids (cDNAs) for a vast number of GPCRs followed elucidation of the primary sequence of the adrenergic receptor, and in every case the same core structure was predicted by hydropathy analysis. In addition to the predicted core structure, certain other common features (with exceptions in some subsets of the GPCR superfamily) were noted.

than for ligand binding. The disulfide bridge may help in proper arrangement of the transmembrane helices.

Superimposed on the basic structure of GPCRs are a number of variations relevant to differences in ligand binding, G protein coupling, and interaction with other proteins. First, there are major differences in amino acid sequence among members of the GPCR superfamily. Sequence alignment, especially of the transmembrane helices, allows one to divide the superfamily into subfamilies. Of these, family 1 is the largest and itself can be subdivided. The largest subset includes opsin, odorant receptors, and monoamine, purinergic, and opiate receptors. These are characterized by a short amino terminus. The next subset includes chemokines, protease-activated, and certain peptide hormone receptors characterized by a slightly longer amino terminus. The last subset comprises receptors for the large glycoprotein hormones, TSH, luteinizing hormone, and follicle-stimulating hormone. These have an approximately 400-residue extracellular amino terminus.

Family 2 shows essentially no sequence homology to family 1 even within the transmembrane helices and is characterized by an approximately 100-residue amino terminus. Members include receptors for peptide hormones such as parathyroid hormone (PTH) and secretin. Agonists (arrow) may bind to residues in the extracellular amino terminus and loops as well as transmembrane helices (arrow). Family 3 includes the extracellular Ca²⁺ sensing receptor and metaboletic glutamate receptors. Agonists (sphere) bind in a cleft of the Venus flytrap-like domain in the large extracellular amino terminus, thereby activating the receptor through as yet undefined interactions with the extracellular loops or transmembrane helices (arrow).

The G protein-coupled receptor (GPCR) superfamily: diversity in ligand binding and structure. Each panel depicts various members of the GPCR superfamily in cartoon form. The seven membrane-spanning helices are shown as cylinders with the extracellular amino terminus and three extracellular loops above and the intracellular carboxyl terminus and three intracellular loops below. The superfamily can be divided into three subfamilies on the basis of amino acid sequence conservation within the transmembrane helices. Family 1 includes (A) the opsins, in which light (jagged arrow) causes isomerization of retinal covalently bound within the pocket created by the transmembrane helices (bar); (B) monoamine receptors, in which agonists (arrow) bind noncovalently within the pocket created by the transmembrane helices (bar); (C) receptors for peptides such as vasopressin, in which agonist binding (arrow) may involve parts of the extracellular amino terminus and loops as well as the transmembrane helices (bar); and (D) glycoprotein hormone receptors, in which agonists (oval) bind to the large extracellular amino terminus, thereby activating the receptor through as yet undefined interactions with the extracellular loops or transmembrane helices (arrow). Family 2 includes receptors for peptide hormones such as parathyroid hormone (PTH) and secretin. Agonists (arrow) may bind to residues in the extracellular amino terminus and loops as well as transmembrane helices (bar). Family 3 includes the extracellular Ca²⁺ sensing receptor and metabotropic glutamate receptors. Agonists (sphere) bind in a cleft of the Venus flytrap-like domain in the large extracellular amino terminus, thereby activating the receptor through as yet undefined interactions with the extracellular loops or transmembrane helices (arrow)....
Because the number of potential G proteins to which GPCRs couple is much more limited than the number of ligands that bind GPCRs, more conservation of the domains involved in G protein coupling would be expected. Although GPCRs can be broadly divided into those that couple to Gs, those that couple to the Gq subfamily, and those that couple to the Gi-Go subfamily, the situation is probably more complicated. Specificity of coupling to the most recently identified G proteins, G12 and G13, is still uncertain. Also, some GPCRs evidently can couple to both Gs and Gq.

A vast number of studies have been performed to define the sites of ligand binding and G protein coupling of GPCRs. Considerable evidence points to the third intracellular loop (particularly its membrane-proximal portions) and to the membrane-proximal portion of the carboxyl terminus as key determinants of G protein coupling specificity. For example, exchanging only the third intracellular loop between different GPCRs confers the G protein coupling specificity of the exchanged loop upon the recipient GPCR. In contrast, the second intracellular loop, although important for G protein coupling, appears to play a role in the activation mechanism rather than in determining specificity of coupling. A tripeptide motif (D/E, R, Y/W) at the start of the second intracellular loop that is highly conserved in family 1 GPCRs is critical for G protein activation.

Mechanism of Activation

The precise mechanism of activation after agonist binding remains to be defined for most GPCRs, but studies of rhodopsin provide the clearest picture available. In the ground state, retinal covalently bound to the seventh transmembrane helix in rhodopsin holds the transmembrane helices in an inactive conformation. Isomerization of retinal upon absorption of light of the appropriate wavelength converts an antagonist ligand into an agonist. The rhodopsin crystal structure identifies the residues in the transmembrane helices that interact with retinal and suggests a mechanism for movement of the helices upon photoactivation of retinal. Movement of the transmembrane helices in turn leads to changes in conformation of cytoplasmic loops that promote G protein activation.

For family 1 receptors related to rhodopsin, the determination of its three-dimensional structure validates the idea that a change in conformation of transmembrane helices is the direct result of agonist versus antagonist binding to residues within the helices. Further refinements in understanding the mechanism of activation for opsin-related GPCRs should come as additional three-dimensional structures are determined. Until then, molecular modeling by computer on the basis of the rhodopsin structure and then experimental testing offer a useful approach. For other GPCRs whose presumptive site of agonist binding does not involve direct contact with transmembrane helices (families 2 and 3 and the glycoprotein hormone receptors in family 1), much remains to be learned about the mechanism of activation. Specifically, determining how agonist binding to the extracellular domains of such GPCRs leads to presumptive changes in conformation of transmembrane helices requires further studies of structure and function.

A general hypothesis of GPCR activation postulates that GPCRs are in equilibrium between an activated state and an inactive state. These states presumably differ in the disposition of the transmembrane helices and, in turn, the cytoplasmic domains that determine G protein coupling. Agonists, according to this model, are viewed as stabilizing the activated state. Antagonists may be neutral, that is, they simply compete with agonist for receptor binding but their binding does not influence this equilibrium. Alternatively, they may be "inverse" agonists; that is, their binding stabilizes the inactive state of the receptor. Naturally occurring, activating mutations of the residues in the transmembrane helices that interact with retinal and suggests a mechanism for movement of the helices upon photoactivation of retinal.

Dimerization

Members of the tyrosine kinase receptor family have long been known to require dimerization as part of their activation mechanism. It is now apparent that many GPCRs likewise form homodimers and heterodimers. Residues within transmembrane helix 6 may foster dimerization of small family 1 GPCRs and intermolecular disulfide bonds in the extracellular amino-terminal domain are involved in homodimerization of most family 3 GPCRs. A coiled-coil interaction in the carboxyl terminus of -aminobutyric acid B receptor subtypes is responsible for heterodimerization, and this is critical for proper receptor function. Modifications of ligand binding, signaling, and receptor sequestration have been demonstrated upon heterodimerization of angiotensin with bradykinin receptors, of opioid receptors, and of opioid with -adrenergic receptors. Further studies are needed to elucidate the role of homodimerization and heterodimerization in GPCR function.

G Protein Coupling

Mechanism of Activation

The precise mechanism of activation after agonist binding remains to be defined for most GPCRs, but studies of rhodopsin provide the clearest picture available. In the ground state, retinal covalently bound to the seventh transmembrane helix in rhodopsin holds the transmembrane helices in an inactive conformation. Isomerization of retinal upon absorption of light of the appropriate wavelength converts an antagonist ligand into an agonist. The rhodopsin crystal structure identifies the residues in the transmembrane helices that interact with retinal and suggests a mechanism for movement of the helices upon photoactivation of retinal. Movement of the transmembrane helices in turn leads to changes in conformation of cytoplasmic loops that promote G protein activation.

For family 1 receptors related to rhodopsin, the determination of its three-dimensional structure validates the idea that a change in conformation of transmembrane helices is the direct result of agonist versus antagonist binding to residues within the helices. Further refinements in understanding the mechanism of activation for opsin-related GPCRs should come as additional three-dimensional structures are determined. Until then, molecular modeling by computer on the basis of the rhodopsin structure and then experimental testing offer a useful approach. For other GPCRs whose presumptive site of agonist binding does not involve direct contact with transmembrane helices (families 2 and 3 and the glycoprotein hormone receptors in family 1), much remains to be learned about the mechanism of activation. Specifically, determining how agonist binding to the extracellular domains of such GPCRs leads to presumptive changes in conformation of transmembrane helices requires further studies of structure and function.

A general hypothesis of GPCR activation postulates that GPCRs are in equilibrium between an activated state and an inactive state. These states presumably differ in the disposition of the transmembrane helices and, in turn, the cytoplasmic domains that determine G protein coupling. Agonists, according to this model, are viewed as stabilizing the activated state. Antagonists may be neutral, that is, they simply compete with agonist for receptor binding but their binding does not influence this equilibrium. Alternatively, they may be "inverse" agonists; that is, their binding stabilizes the inactive state of the receptor. Naturally occurring, activating mutations of the residues in the transmembrane helices that interact with retinal and suggests a mechanism for movement of the helices upon photoactivation of retinal.

Dimerization

Members of the tyrosine kinase receptor family have long been known to require dimerization as part of their activation mechanism. It is now apparent that many GPCRs likewise form homodimers and heterodimers. Residues within transmembrane helix 6 may foster dimerization of small family 1 GPCRs and intermolecular disulfide bonds in the extracellular amino-terminal domain are involved in homodimerization of most family 3 GPCRs. A coiled-coil interaction in the carboxyl terminus of -aminobutyric acid B receptor subtypes is responsible for heterodimerization, and this is critical for proper receptor function. Modifications of ligand binding, signaling, and receptor sequestration have been demonstrated upon heterodimerization of angiotensin with bradykinin receptors, of opioid receptors, and of opioid with -adrenergic receptors. Further studies are needed to elucidate the role of homodimerization and heterodimerization in GPCR function.
Pharmacologists long ago appreciated that continued exposure to agonist leads to a diminished response, so-called desensitization. This phenomenon has been extensively studied in GPCRs. Two forms are defined: heterologous, in which binding of agonist to one GPCR leads to a diminished response of a different GPCR to its agonist, and homologous, in which desensitization occurs only for the GPCR to which agonist is bound. Both forms of desensitization involve GPCR phosphorylation but by different kinases and at different sites. Stimulation of cyclic adenosine monophosphate formation by agonist binding to a Gs-coupled GPCR leads to activation of protein kinase A, which in turn can phosphorylate and desensitize the GPCR. Such phosphorylation may also alter G protein coupling specificity. Similarly, protein kinase C activation resulting from GPCR coupling to Gq family members may cause protein kinase C catalyzed phosphorylation of GPCRs with desensitization.

In retinal photoreceptors, a specific rhodopsin kinase and a protein termed arrestin were implicated in attenuation of the light response. Just as parallels were identified between rhodopsin and GPCR structure, so were parallels identified in this desensitization mechanism. Rhodopsin kinase is but one member of a family of GPCR kinases and arrestin only one of a family of related proteins that function in desensitization of many members of the GPCR superfamily. GPCR kinases preferentially phosphorylate the agonist-bound form of a GPCR, thus ensuring homologous desensitization. Upon GPCR phosphorylation by GPCR kinase, arrestins bind to the third intracellular loop and carboxyl-terminal tail of the GPCR, thereby blocking G protein binding. There is evidence that GPCR kinases and arrestins not only act to desensitize GPCRs but also mediate other functions including receptor internalization and interaction with other effectors (see next section).
G Protein-Coupled Receptor Interactions with Other Proteins

The initial paradigm of GPCR function postulated that G protein activation is the sole outcome of agonist binding to GPCRs. With the identification of GPCR interactions with GPCR kinases and arrestins, this concept was modified to include these proteins involved in GPCR desensitization. Later evidence, however, suggests that GPCR interaction with arrestins may also permit recruitment of other proteins to the GPCR. For example, the src tyrosine kinase may interact with the -adrenergic receptor with -arrestin serving as an adaptor. A arrestins may also recruit proteins involved in endocytosis. GPCR kinases may also serve to recruit additional signaling proteins to the GPCR.

Other classes of proteins may interact with specific GPCRs without recruitment by GPCR kinases and arrestins. These include SH2 domain-containing proteins, small GTP-binding proteins, and PDZ (for postsynaptic density protein-95/discs large/zona occluden-1) domain-containing proteins. Examples of the latter include binding of the Na+/H+ exchanger regulatory factor to the carboxyl terminus of the -adrenergic receptor. The long carboxyl terminus of family 3 GPCRs such as metabotropic glutamate receptors contains polyproline motifs involved in binding members of the Homer family. The latter can facilitate functional interactions with yet other proteins such as the inositol trisphosphate receptor. Receptor activity-modifying proteins (RAMPs), a new family of single-transmembrane-domain proteins, appear to heterodimerize with certain GPCRs, assisting them in proper folding and membrane trafficking. Interestingly, when the calcitonin receptorlike GPCR associates with RAMP1, it forms a calcitonin generelated peptide receptor, whereas when it associates with RAMP2, it becomes an adrenomedullin receptor. Clearly, this rapidly evolving aspect of GPCR function holds many further interesting developments in store.
Because of their diverse and critical roles in normal physiology, their accessibility on the cell surface, and the ability to synthesize selective agonists and antagonists, GPCRs have long been a major target for drug development. One estimate is that about 65% of prescription drugs are targeted against GPCRs. With the cloning of GPCR cDNAs, much greater diversity of receptor subclasses became evident than had been anticipated on the basis of pharmacologic studies. For example, five muscarinic receptor subtypes and an even greater number of serotonergic GPCRs were identified. This has allowed the development of highly specific, subtype-selective drugs that have fewer side effects than those produced by previously available agents.

Another result of the cloning of GPCR cDNAs by homology screening and polymerase chain reaction-based approaches is the identification of orphan GPCRs, that is, receptors with the canonical, predicted seven-transmembrane-domain structure of GPCRs but without knowledge of their physiologic agonist. There have been substantial efforts to identify the relevant ligands for such orphan receptors. An example of the success of such efforts is the identification of an orphan GPCR as the neuropeptide U receptor involved in regulation of feeding.

Beyond drug development, defects in GPCRs are an important cause of a wide variety of human diseases. GPCR mutations can cause loss of function by impairing any of several steps in the normal GPCR-GTPase cycle. These include failure to synthesize GPCR protein altogether, failure of synthesized GPCR to reach the plasma membrane, failure of GPCR to bind or be activated by agonist, and failure of GPCR to couple to or activate G protein. Because in most cases clinically significant impairment of signal transduction requires loss of both alleles of the GPCR gene, most such diseases are inherited in autosomal recessive fashion. Most of these diseases are manifested as resistance to the action of the normal agonist and thus mimic deficiency of the agonist. For example, TSH receptor loss-of-function mutations cause a form of hypothyroidism mimicking TSH deficiency, but serum TSH is actually elevated in such cases, reflecting resistance to the hormone's action caused by defective receptor function. Nephrogenic diabetes insipidus (renal vasopressin resistance) is caused by loss-of-function mutations in the V2 vasopressin receptor gene located on the X chromosome. Thus, males with a single copy of the gene experience the disease when they inherit a mutant gene, whereas most females do not show overt disease because random X inactivation leaves them with on average 50% of normal gene function. Most V2 vasopressin receptor mutations associated with nephrogenic diabetes insipidus cause loss of function by impairing normal synthesis or folding of the receptor, or both. A novel mechanism for receptor loss of function elucidated for a V2 vasopressin receptor missense mutation associated with nephrogenic diabetes insipidus involves constitutive arrestin-mediated desensitization.

The extracellular Ca²⁺-sensing receptor appears to be an interesting exception to the association between GPCR loss-of-function mutations and hormone resistance. Loss-of-function mutations of the Ca²⁺-sensing receptor mimic a hormone hypersecretion state, primary hyperparathyroidism. In fact, Ca²⁺-sensing receptor loss-of-function mutations do cause hormone resistance, but in this case extracellular Ca²⁺ is the hormonal agonist that acts through this receptor to inhibit PTH secretion. A loss-of-function mutation of one copy of the receptor gene typically causes mild resistance to extracellular Ca²⁺ manifested as familial hypocalciuric hypercalcemia. If two defective copies are inherited, extreme Ca²⁺-sensing causing neonatal severe primary hyperparathyroidism results. In some cases, a heterozygous receptor loss-of-function mutation may be associated with neonatal severe primary hyperparathyroidism, perhaps reflecting a dominant negative effect caused by dimerization of wild-type and mutant receptors.

GPCR gain-of-function mutations (Table 5-2) are also an important cause of disease. Given the dominant nature of activating mutations, most such diseases are inherited in an autosomal dominant manner. Activating TSH receptor mutations may be inherited in autosomal dominant fashion and cause diffuse thyroid enlargement in familial nonautoimmune hyperthyroidism, or they may occur as somatic mutations causing focal, sporadic hyperfunctional thyroid nodules. Unlike loss-of-function mutations, which may be missense as well as nonsense or frameshift mutations that truncate the normal receptor protein, GPCR gain-of-function mutations are almost always missense mutations. The location and nature of naturally occurring, disease-causing mutations offer important insights into GPCR structure and function. The basis for defective receptor function is clear with mutations that truncate receptor synthesis prematurely. More subtle missense mutations may impair function if they involve highly conserved residues in transmembrane helices critical for normal protein folding. Activating missense mutations often involve residues within or bordering transmembrane helices and are thought to disrupt normal inhibitory constraints that maintain the receptor in its inactive conformation. Mutations disrupting these constraints mimic the effects of agonist binding and shift the equilibrium toward the activated state of the receptor.
Clinically, diseases caused by activating GPCR mutations therefore mimic states of agonist excess, but direct measurement shows that agonist concentrations are actually low, reflecting normal negative feedback mechanisms. Again, the Ca\textsuperscript{2+}-sensing receptor is an apparent exception, with activating mutations causing functional hypoparathyroidism. For most GPCRs, disease-associated gain-of-function mutations cause constitutive, agonist-independent, activation but with rare exceptions, the Ca\textsuperscript{2+}-sensing receptor gain-of-function mutations cause increased sensitivity to extracellular Ca\textsuperscript{2+} rather than to Ca\textsuperscript{2+}-independent activation.

Naturally occurring animal models of human disease have revealed additional examples of etiologic GPCR mutations. For example, a loss-of-function mutation in the hypocretin (orexin) type 2 receptor gene was identified in canine narcolepsy. Dozens of mouse GPCR gene knockout models have been created, many revealing interesting and in some cases unexpected phenotypes. Characterization of the phenotype resulting from disruption of a mouse GPCR gene may accurately predict the clinical picture resulting from the corresponding mutation in humans, such as with disruption of the melanocortin-4 receptor gene resulting in obesity in mouse and human and disruption of the PTH/PTH-related protein receptor gene impairing normal bone growth and development in mouse and in the human disease Blomstrand chondrodysplasia. Further knockout models and further detailed studies of these models can be expected to increase substantially our understanding of GPCR function and to address questions such as the unique roles of multiple subtypes of various GPCR subclasses, for example, the 3-adrenergic receptor subtype. Availability of mouse knockout models of human diseases such as nephrogenic diabetes insipidus should also facilitate testing of novel therapies including gene transfer.

Screening of GPCR genes for mutations as the potential cause of additional human disorders may continue to turn up new examples, but it is also becoming clear that variations in GPCR gene sequence can have profound consequences beyond mutations causing diseases. One of the most striking examples is the discovery that homozygous loss-of-function mutations of the type 5 chemokine receptor (CCRS) confer resistance to human immunodeficiency virus (HIV) infection in individuals with this genotype. The reason is that CCR5 serves as a coreceptor for HIV entry into cells. In the roundworm, two isoforms of a neuropeptide receptor are associated with profound differences in feeding behavior. As more polymorphisms are discovered in the human genome, many examples of variations in GPCR gene sequence will be found and the challenge will be to elucidate their possible functional significance. In vitro studies may reveal functional differences, such as differences in G protein coupling seen with a four-amino-acid polymorphism in the third intracellular loop of the \( \beta_2 \)-adrenergic receptor, but further studies are required to determine whether such differences are important in individual variation in response to various drugs (pharmacogenomics) or in other subtle physiologic differences that could confer susceptibility to disease (complex disease genes). Given the high proportion of the human genome devoted to GPCR genes, it is clear that studies of this gene superfamily will play a prominent role in the postgenomic era.
References


Cancer Res 1997; 57:391395.


18. Cleverenger CV, Medaglia MV. The protein tyrosine kinase P59fyn is associated with proteolysis (PRL) receptor and is activated by PRL stimulation of T-lymphocytes. Mol Endocrinol 1994; 8:674681.


Chapter 6 - Laboratory Techniques for Recognition of Endocrine Disorders

George Klee

Endocrinology is a practice of medicine that is highly dependent on accurate laboratory measurements because small changes in hormone levels often may be more specific and more sensitive for early disease than the classic physical signs and symptoms. Because most endocrinologists currently do not have facilities to develop and validate laboratory assays, they rely on commercial analytic assays or send a patient’s specimen to specialized laboratories. Even most hospital and commercial laboratories have minimal expertise for developing analytic assays. This critical dependence on quality laboratory measurements, combined with minimal information about the performance of these tests, places endocrinologists in a potentially vulnerable position.

This chapter attempts to provide an overview of the strengths and weaknesses of the analytic techniques typically used for endocrine measurements in blood and urine. Concentrations of most hormones are much lower than those of general chemistry analytes, and specialized techniques are necessary to measure these low concentrations.

Three major types of assays for measuring hormones are described:
- Immunoassays (both competitive and sandwich)
- Chromatography
- Mass spectrometry

Nucleic acid measurements for evaluation of genetic alterations also are reviewed.

The minimal analytic performance validation required by the federal government for laboratories testing specimens of Medicare patients, along with explanations of these performance parameters, is outlined. This information should help endocrinologists better assess the performance of the analytic systems that they are using.

Techniques to investigate discordant laboratory test values also are presented to help clinicians work with their laboratories to reconcile test values that do not match clinical presentations.

Hormone concentrations are reported in molar units, mass units, or standardized units, such as World Health Organization (WHO) International Units (IU). When these measurements are expressed in molar units, most hormones in blood and urine are present in concentrations of 10^{-6} to 10^{-12} M/L (Fig. 6-1). The terms used to describe these concentrations are micromolar (10^{-6} M/L), nanomolar (10^{-9} M/L), and picomolar (10^{-12} M/L). The range from the lowest to highest concentrations is more than a million-fold difference. Therefore, laboratory techniques must be targeted to the levels of each given hormone.

The major techniques for measuring picomolar concentrations are immunoassay and mass spectrometry, whereas nanomolar and micromolar concentrations can be measured by these methods as well as chromatography and chemical detection systems. Some hormones, such as thyrotropin (TSH), have very low concentrations in the femtomolar (10^{-15} M/L) range in patients with diseases such as thyrotoxicosis. Exquisitely sensitive immunometric assays are usually needed to measure these very low concentrations.
TYPES OF ASSAYS

The four major techniques used for endocrine measurements are as follows:

Antibody-based immunologic assays, of which there are two subcategories: competitive immunoassays and immunometric (sandwich) assays

Chromatographic assays
Mass spectroscopy
Nucleic acid-based assays

Competitive Immunoassays

The term competitive radioimmunoassay refers to a measurement method in which an antigen (e.g., a hormone) in a specimen competes with radiolabeled reagent antigen for a limited number of binding sites on a reagent antibody. The three basic components of a competitive immunoassay are [14]:

1. Antiserum specific for a unique epitope on a hormone or antigen.
2. Labeled antigen that binds to this antiserum.
3. Unlabeled antigen in the specimen or standard that is to be measured.

The antisera is diluted to a concentration in which the number of binding sites available on the antibodies is fewer than the number of antigen molecules (labeled and unlabeled) in the reaction mixture. The labeled and unlabeled antigens compete for these limited number of binding sites on the antiserum. The competition is not always equal because the labeled antigen (tracer) may react differently with the antibody compared with the native antigen. This disparity in reactivity may be caused by alteration of the antigen due to the chemical attachment of the label or by differences in the endogenous antigen versus the form of the antigen used in the reagents. As long as the reactions are reproducible, these differences in reactivity are not important because the reaction can be calibrated with standard reference materials having known concentrations.

The precision of competitive immunoassays is related to the rate of change of the signal compared with the rate of change of concentration (i.e., the slope of the dose-response curve) [6]. In Figure 6-2, the slope is much lower at higher concentrations, causing the assay precision to be less at higher concentrations. Most competitive immunoassays also have a relatively flat dose-response curve at very low concentrations, causing poor precision at the low end of the assay. Consequently, the precision profile for most immunoassays is U-shaped, having the best coefficients of variation in the center of the dose-response curve.

As shown in Figure 6-2, the higher the concentration of the unlabeled antigen, the lower the amount of radiolabeled antigen that binds to the limited amount of antiserum. The signal decreases exponentially from the approximately half-maximum at zero concentration to a minimum value at high concentrations. This minimal binding, or nonspecific binding (NSB), is a valuable control parameter. Elevations in NSB usually signify impurities in the label that bind to the sides of the tubes and are not competitively displaced. Most assays add surfactants and proteins to minimize the NSB. Monitoring of changes in the NSB provides an early warning of potential assay problems.

Statistical data-processing techniques are needed to translate the assay signals into concentrations. As illustrated, because these reactions are not linear, numerous curve-fitting algorithms have been developed. Before the introduction of micro-processors, tedious error-prone, manual calculations were required to mathematically transform the data into linear models. A commonly used model was to cross-plot the logit of the normalized signal versus the logarithm of the concentration and to use linear regression lines to establish the dose-response curve. Fortunately, today this procedure of curve fitting usually is accomplished electronically by using programs that automatically test the robustness of fit of multivariable curves after statistically eliminating discordant data points. However, users of these systems must understand the limitations and should pay attention to any warnings presented by the programs during processing of the data.

In radioimmunoassays, radioactive iodine (¹²⁵I) is usually used to label the antigen. The immune complexes are separated from the unbound molecules by precipitation with centrifugation after reaction with secondary antisera and precipitating reagents (e.g., polyethylene glycol). These radioimmunoassays may require special handling and licensure to ensure safety of the radioisotopes and are labor-intensive. The statistical counting errors associated with the relatively low radioactive counts and the poor reproducibility associated with the multiple manual steps generally necessitate that most laboratories perform the measurements in
Immunoassays measure concentrations rather than biologic activity. For most hormones, there is a strong correlation between the concentration of the protein or measurement may provide more reliable information. Results and some disease states may require more analytic sensitivity to ensure sound clinical decisions. In these cases, extraction of specimens prior to commercial assays generally use reagents having adequate sensitivity and specificity to measure clinical laboratories. Although purification before immunoassay are seldom used in clinical assays. These techniques are difficult to automate and require skills and equipment not available in many solid-phase extraction with absorption and selective elution from resins such as silica gels, and (3) immunoaffinity chromatography.

Numerous extraction systems have been developed, including (1) organic-aqueous partitioning to remove water-soluble interferences seen with steroids, (2) extraction of hormones from serum and urine specimens prior to measurement is a technique that can enhance both sensitivity and specificity of immunoassays. The process of generating these antisera is a combination of art, science, and luck. Generally, a relatively pure form of the antigen is conjugated to a carrier protein (especially if the antigen is less than 10,000 d), mixed with adjuvant (e.g., Freund's complete adjuvant), and injected intradermally into the host animal. After several boosts with conjugated protein plus Freund's incomplete adjuvant, the host animal recognizes the material as foreign and develops immune responses. The antigen is then harvested from the animal's blood. Under optimal conditions, moderate quantities of high-affinity antisera, which react only with the specific target antigen, are developed. The analytic sensitivity of a competitive immunoassay is approximately inversely related to the affinity of the antigen, such that an antisera with an affinity constant of

10^9 L/M can be used to measure analytes in the nanomolar concentration range.

The polyclonal antisera developed by immunizing animals represents a composite of many immunologic clones, with each clone having a different affinity and different immunologic specificity. Most clones have affinities in the 10^-2 to 10^-4 L/M range, with only rarely having affinities above 10^-1 L/M. Various techniques are used to develop a specific antisera, including (1) altering the form of the antigen by blocking cross-reacting epitopes and (2) purifying the antisera using affinity chromatography to select antibodies directed toward the epitope of interest. Affinity-column purification can also be used for immunoextraction of higher-affinity antisera by selectively eluting antisera from the column by means of a series of buffers with increasing acidity. The major disadvantage of a polyclonal antisera is the limited quantity. The large quantities needed by commercial suppliers of immunoassay reagents often require them to use multiple sources of antisera. These changes in antisera can cause significant changes in assay performance. In many instances, laboratories and clinicians are not informed about these changes, which may cause problems in medical decisions.

Monoclonal antisera are used in many current immunoassays. These antisera are made by immunizing animals (usually mice) using techniques similar to those used for polyclonal antisera; instead of harvesting the antisera from the blood, however, the animal is killed and the spleen is removed. The lymphocytes in the spleen are fused with myeloma cells to make cells that will grow in culture and produce antisera. These fused cells are separated into clones by means of serial plating techniques similar to those used in subculturing bacteria. The supernantant of these monoclonal cell lines (or ascites fluid if the cells are transplanted into carrier mice) contains monoclonal antisera. The selection processes used to separate the initial clones can be targeted to identify specific clones producing antisera with high affinities and low cross-reactivity to related compounds.

The high specificity of monoclonal antisera can cause problems for some endocrine assays. Many hormones circulate in the blood as heterogeneous mixtures of multiple forms. Some of these forms are caused by genetic differences in patients, whereas other forms are related to metabolic precursors and degradation products of the hormone. Genetic differences cause some patients to produce variant forms of a hormone such as luteinizing hormone (LH). These genetic differences can cause marked variations in measurements made using assays with specific monoclonal antisera. Well-characterized monoclonal antisera can be mixed together to make an "engineered polyclonal antisem" with improved sensitivity and specificity. Cross-reactivity with precursor forms of the analytes and with metabolic degradation products can cause major differences in assays. For example, cross-reactivity with precursor forms causes differences in insulin assays, and cross-reactivity with metabolic fragments causes major differences in carbohydrate-terminal parathyroid hormone (PTH) assays. Extraction of hormones from serum and urine specimens prior to measurement is a technique that can enhance both sensitivity and specificity of immunoassays. Numerous extraction systems have been developed, including (1) organic-aqueous partitioning to remove water-soluble interferences seen with steroids, (2) solid-phase extraction with absorption and selective elution from resins such as silica gels, and (3) immunoaffinity chromatography. Unfortunately, extraction and purification before immunoassay are seldom used in clinical assays. These techniques are difficult to automate and require skills and equipment not available in many clinical laboratories. Although commercial assays generally use reagents having adequate sensitivity and specificity to measure most patient specimens, some patient specimens may give spurious results and some disease states may require more analytic sensitivity to ensure sound clinical decisions. In these cases, extraction of specimens prior to measurement may provide more reliable information.

Immunnoassays measure concentrations rather than biologic activity. For most hormones, there is a strong correlation between the protein or steroid being measured and the biologic activity, but this is not universally true. The reactive site for most antibodies is relatively small, about 5 to 10 amino acids for linear peptides. Some antisera reactions are specific for the tertiary structure that corresponds to unique molecular configurations, but immunoassays seldom react...
with the exact antigenic structure that confers biologic activity.

Figure 6-3 presents a schematic illustration of the difference between immunologic binding site and biologic receptor binding site on a hormone. Indirect immunoassays have been developed using cultured cells that synthesize second messengers such as cyclic adenosine monophosphate (cAMP) at rates proportional to the concentration of hormone in the specimen. An example of this technique is the immunoassay measurement of cAMP produced by osteosarcoma cells to quantitate PTH bioactivity in serum. Unfortunately, these assays are tedious and generally are not reproducible. More recent techniques using recombinant receptors as immunoassay binders may provide improved specificity with good reliability.
Immunometric (Sandwich) Assays

A second immunologic technique used to measure hormones is the immunometric (sandwich) assay. The three basic components of a sandwich assay are:

1. An antigen large enough to allow two antibodies to bind concurrently on different binding sites.
2. A capture antiserum directed to one of the antigenic sites on the antigen. This antiserum is attached to a solid phase to permit immunologic extraction of the immune complex.
3. A signal antiserum directed to a second antigenic site on the antigen. This antiserum is attached to an assay signal system.

In contrast to competitive immunoassays, these assays use a large excess of antiserum binding sites compared with the concentration of antigen. The capture antibody immunoextracts the antigen from the sample and the signal antibody binds to the capture antibody-antigen complex to form a tertiary complex. As the antigen concentration increases, the signal increases progressively.

Figure 6-4 schematically illustrates these concepts. The capture antiserum (ATB1) is attached to biotin (see solid circles). The signal antiserum (ATB2) is labeled with a detection system (see asterisks). The ATB1-antigen-ATB2 complexes are immunologically extracted using a streptavidin solid phase (see horizontal cups). After the complex is bound to the solid phase, most of the unbound signal antibody is washed away.

As shown in Figure 6-4, the signal increases progressively with the concentration. For lower concentrations, the signal generally increases proportional to the assay concentrations (after the offset caused by the NSB). At higher concentrations, the signal generally is less than proportional, so that nonlinear curve-fitting techniques are used to generate the dose-response curves. Again, the relative imprecision, expressed as a coefficient of variation, depends on the slope of the dose-response curve; consequently, the relative precision is less at higher concentrations.

In immunometric assays, the background level of signal is associated with very low concentrations. This background signal is caused by the NSB. The analytic sensitivity of immunometric assays is related to the ratio of the true signal to the NSB signal. Therefore, assays can be made more sensitive either by increasing the response signal or by decreasing NSB. Inadvertent increases in NSB caused by specimen interference or reagent deterioration can significantly alter the assay performance.

In immunometric assays, it is also important that a large excess of capture antibody be used. When the antigen concentration approaches the effective binding capacity of the capture antibody system, the signal no longer increases. If the antigen concentration exceeds the binding capacity of the capture antibody, the signal may actually decrease.

Figure 6-5 illustrates this high-dose hook effect for immunometric assays caused by insufficient amounts of capture or signal antiserum. The signal increases progressively until the hormone concentration exceeds the binding capacity; the signal then decreases, apparently as a result of the removal of some of the weaker binding antigen-antibody complexes during the wash cycle on the assay. This is a potentially dangerous phenomenon because very high concentrations can give the same “answer” as lower concentrations. If this artifact is suspected, the specimen can be diluted and reanalyzed. If the answer for the diluted specimen is higher than the original answer, a high-dose hook effect probably is present.

Most manufacturers are aware of this potential problem and configure assays with relatively large amounts of capture or signal antiserum; however, some patients produce high concentrations of hormones or antigens that may exceed assay limits. Laboratories are able to detect this phenomenon by analyzing specimens at two dilutions, but this practice generally is not cost-effective. Therefore, feedback to the laboratory about results that are inconsistent with clinical findings is essential.

Another potential problem for immunometric assays consists of endogenous heterophile antibodies that cross-react with reagent antiserum. Normally, the signal antibody does not form a “sandwich” with the capture antibody unless the specific antigen is present; however, divalent heterophile antibodies may mimic the antigen by simultaneously binding to the signal and capture reagent antibodies.

Figure 6-6 schematically illustrates this situation. The problem is most common with monoclonal antibodies but may also occur with polyclonal antibodies. Immunoglobulins contain both a constant (Fc) region and a variable (Fab) region. As implied in the name, the Fc region is constant, or similar, for all immunoglobulins from that species. Therefore, if a patient receives immunotherapy or imaging reagents containing mouse immunoglobulin, they are likely to develop human antimouse antibodies (HAMAs) directed to the Fc fragment. Some patients may develop heterophile antibodies after exposure to foreign proteins from domestic pets or food contaminants. When these endogenous antibodies are present in a patient’s specimen, they may bridge across the reagent antibodies used for immunometric assays.
in immunometric assays and may cause falsely high values. These antibodies also may bind to sites on the reagent antibodies, which sterically block the binding of the specific antigen and give falsely low test values. Most manufacturers include nonimmune immunoglobulin in the assays to help block these interferences; as with the high-dose hook effect, however, the amounts added are not always adequate and some patients with high titer antibodies thus may still show in vitro assay interference.

The combined specificity of the two antibodies used in an immunometric assay can produce exquisitely sensitive and specific immunoassays. In the past, a common problem with early competitive immunoassays was cross-reactivity among the structurally similar gonadotropins: LH, follicle-stimulating hormone (FSH), TSH, and human chorionic gonadotropin (hCG). The subunits of each of these hormones are almost identical, and the subunits have considerable structural homology. Many individual antisera (especially polyclonal antisera) used for measuring one of these hormones may have cross-reactivity for the other gonadotropins. The cross-reactivity of a pair of antibodies is less than the cross-reactivity of each of the individual antibodies because any cross-reacting substance must contain both of the binding epitopes in order to simultaneously bind to both antibodies.

For example, consider two antibodies for LH, each having 1% cross-reactivity with hCG. The cross-reactivity of the pair is less than the product of the two cross-reactivities or, in this case, less than 0.01%. Most current immunoassays for LH have cross-reactivity less than 0.01% because even this relatively low percentage of cross-reactivity would still cause significant assay interference in pregnant patients and patients with choriocarcinoma who have high hCG concentrations.

Multiple forms of most hormones circulate in the blood. Some hormones (e.g., prolactin, growth hormone) circulate with macro forms, which can cause difficulty in their analysis if specimens are not pretreated. For hormones composed of subunits (e.g., the gonadotropins), both the intact and the free subunits circulate in blood. Immunometric assays can be made specific for intact molecules by pairing an antibody specific for the - bridge site of the subunits with a second antibody specific for the subunit. Assays using these antibody pairs retain the two-antibody, low cross-reactivity needed for measuring gonadotropins and do not react with the free subunit forms of the hormones.

The heterogeneous specificity characteristics of immunoassays make calibration and harmonization difficult. Two immunoassays calibrated with the same reference preparation can give widely varying measurements on patient specimens. Consider the example of hCG in Table 6-1. The three assays are calibrated with a pure preparation of intact hCG, such as the WHO Third International Reference Preparation. The three assays differ in their cross-reactivity with free hCG (0, 100%, and 200%, respectively). These assays give identical measurements for a specimen containing only intact hCG but progressively disparate values as the percentage of free hCG in the specimen increases. In reality, the standardization issue is much more complex because multiple forms of hormones (i.e., intact, free subunits, nicked forms, glycosylated forms, degradation products) circulate in patients and each assay has different cross-reactivities for these forms.

<table>
<thead>
<tr>
<th>TABLE 6-1 -- Effect of Immunoassay Specificity on Calibration of Human Chorionic Gonadotropin (hCG) Assay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specificity for intact hCG standard</td>
</tr>
<tr>
<td>Cross-reactivity with free hCG</td>
</tr>
<tr>
<td>Value of specimen with no free hCG, IU/L</td>
</tr>
<tr>
<td>Value of specimen with 10% free hCG, IU/L</td>
</tr>
<tr>
<td>Value of specimen with 50% free hCG, IU/L</td>
</tr>
</tbody>
</table>
Free (Unbound) Hormone Assays

Many hormones are tightly bound to specific plasma-binding proteins and loosely bound to albumin. The unbound (free) forms as well as some of the loosely bound forms are biologically active.

Multiple methods are available to measure these free or biologically active forms of a hormone. Theoretically, the best procedure is direct measurement of the free hormone concentration after physical separation of free-form bound hormone by equilibrium dialysis, ultrafiltration, or gel filtration. Unfortunately, this method is difficult to perform and is thus not readily available and is subject to technical errors.

The two major clinical applications for free hormone measurements are for thyroid hormones (thyroxine [FT₄] and triiodothyronine [FT₃]) and steroids (testosterone and estradiol). Four techniques are commonly used to estimate free thyroid hormone concentrations:

1. **Indirect index methods.** The indirect indices involve two measurements: one for total hormone concentration and another for the thyroxine-binding globulin (TBG), followed by calculation of the ratio or a normalized index (FT₄I or FT₃I). These methods correct for routine changes in TBG associated with estrogen levels, but they may produce inappropriately abnormal values in patients with extreme variations in TBG levels found in patients with congenital disorders of the TBG gene, familial dysalbuminemic hyperthyroxinemia, thyroid hormone autoantibodies, and nonthyroidal illnesses.

2. **Two-step labeled hormone methods.** These methods immunologically bind the free and loosely bound thyroid hormone to a solid phase. The other serum components are washed away, and the residual binding sites are back-titrated with labeled hormone. When calibrated with appropriate serum standards, these methods are thought to pose fewer problems with binding protein abnormalities.

3. **One-step labeled hormone analogue methods.** These methods use synthetic analogues of T₄ and T₃ that bind to the measurement antibody but do not bind to normal TBG. These methods are seldom used because performance has been poor in patients with abnormal albumin concentrations, abnormal free fatty-acid concentrations, and all conditions that interfere with the indirect indices.

4. **Labeled antibody methods.** These methods use kinetic reactions of antibodies with selected affinities that bind preferentially with the free form of the hormone. These methods work best for automated testing instruments and have become popular.

Each of these methods works well for correcting for minor changes in TBG levels, but each has problems with some patient sera, especially those containing interfering substances such as inhibitors and heterophilic antibodies. Unfortunately, most manufacturers have not fully validated their methods in patients with these abnormalities.

Multiple methods are also available for measuring both the free and the biologically active forms of steroid hormones. The preferred method for measurement of free hormones consists of direct physical separation and high-sensitivity assays similar to those recommended for the thyroid hormones. One-step labeled hormone-analogue methods also have been developed, but these are associated with interference problems similar to the problems with free thyroid hormone assays.

Another complexity in regard to steroid hormones is that in addition to the free hormones, testosterone and estrogen bound to albumin also are biologically active. The concentration of the biologically active forms can be estimated using indirect indices calculated from measurements of the total hormones and sex hormone-binding globulin (SHBG) or by measurement of the residual free and albumin-bound steroids after separation of the SHBG-bound forms after differential precipitation with ammonium sulfate.
Chromatographic Assays

Another major method of measuring hormone concentrations involves chromatographically separating the various biochemical forms and quantitating specific characteristics of the molecules. High-performance liquid chromatography (HPLC) systems utilize multiple forms of detection, including light absorption, fluorescence, electrochemical properties, and mass spectrometry.\[^{37,38}\]

There are two major advantages of these techniques: (1) they can be used to simultaneously measure multiple forms of an analyte, and (2) they are not dependent on unique immunologic reagents. Therefore, harmonization of measurements made with different assays is more feasible. The major disadvantages of these methods are their complexity and their limited availability.

Many chemical separation techniques are based on chromatography, but the two most commonly used for liquid chromatography are (1) normal-phase HPLC and (2) reverse-phase HPLC.\[^{25}\] In both systems, a bonded solid-phase column is made that interacts with the analytes as they flow past in a liquid solvent. In normal-phase HPLC, the functional groups of the stationary phase are polar (e.g., amino or nitrile ions) relative to the nonpolar stationary phase (e.g., hexane); in reverse-phase HPLC, a nonpolar stationary phase (e.g., C-18 octadecylsilane molecules bonded to silica) is used.

More recently, polymeric packings made of mixed copolymers have been made with C4, C8, and C18 functional groups directly incorporated so that they are more stable over a wide pH range. The mobile and stationary phases are selected to optimize adherence of the analytes to the stationary phase. The adhered molecules can be eluted differentially from the solid phase after washing to separate specific forms of the analyte from interfering substances as follows:

1. When the composition of the mobile phase remains constant throughout the run, the process is called an **isocratic elution**.
2. If the mobile-phase composition is abruptly changed, a **step elution** occurs.
3. If the composition is gradually changed throughout the run, a **gradient elution** occurs.

The efficiency of separation in a chromatography system is a function of the flow rates of the different substances.\[^{39}\] The resolution of the system is a measure of the separation of the two solute bands in terms of their relative retention volumes ($V_r$) and their bandwidths ($W$). Resolution ($R_s$) of solutes A and B is shown as:

$$ R_s = \frac{2(V_r(B) - V_r(A))}{W(A) + W(B)} $$

Values of $R_s$ less than 0.8 result in inadequate separation, and values greater than 1.25 correspond to baseline separation. The resolution of a chromatography column is a function of flow rates and thermodynamic factors.

The simultaneous measurement of the three catecholamines (epinephrine, norepinephrine, and dopamine) can be performed with reverse-phase HPLC with a C-18 column and electrochemical detection system or fluorometric detection. Prior extraction by absorption on activated alumina and acid elution helps to improve specificity. Dihydroxybenzylamine, a molecule similar to endogenous catecholamines, can be used as an internal standard.
Mass Spectrometry

The technique of mass spectrometry involves fragmentation of target molecules, followed by separation and measurement of the mass to charge ratio of the components. When coupled with liquid chromatography, a mass spectrometer can function as a unique detector to provide structural information about the composition of individual solutes. Inclusion of internal standards in the specimens, which are molecularly similar to the measured compounds, allows precise quantitation of the concentration of the eluting analytes. The measurement of specific mass fragments makes possible the quantitation of multiple specific analytes in complex mixtures.

A fundamental step in mass spectrometry is the fragmentation of the target compound into charged ions. Multiple techniques are used to generate these charged ions, including chemical ionization and electron-impact ionization.

Chemical ionization uses reagent gas molecules, such as methane, ammonia, water, and isobutane, to transfer protons. This process produces less fragmentation than other techniques because the process is not highly excited.

The electron impact bombards gas molecules from the sample, with electrons emitted from a heated filament. The process occurs in a vacuum to prevent the filament from burning out. Electron-spray ionization is a process in which a solution containing the analyte is introduced into a gas phase and is sprayed across an ionizing potential. The charged droplets are desolvated and analyzed in a mass spectrometer.

A mass spectrum is a bar graph in which the heights of the bars correspond to the relative abundance of a particular ion plotted as a function of the mass/charge ratio. Modern mass spectrometers can measure molecular masses so accurately and precisely that the elemental composition of a compound can be predicted by comparison with stored spectral libraries. When these systems are used to measure only a few select compounds having known spectrums, the mass spectrometer can be programmed to focus only on these selected ions.

Stable isotopes of the compounds of interest can be used as internal standards through a technique called isotope dilution mass spectrometry. Stable isotopes generally perform the same as the native compounds in terms of extraction, chromatography, and mass spectrometry and are thus ideal internal standards. However, they must have a sufficient number of isotopic atoms to ensure that their mass is different from naturally occurring substances that may be in the specimen.

Tandem mass spectrometry (MS/MS) is a powerful new tool consisting of two mass analyzers separated by an ion-activation device. The first analyzer is used to isolate and dissociate the ion of interest by activation, and the second mass analyzer is used to analyze its dissociation products. This technique can be used to provide rapid, definitive measurements of multiple endocrine analytes. For example, liquid chromatography and tandem mass-spectrometry can be used to simultaneously quantitate multiple glucocorticoid-related compounds. In Figure 6-7, the chromatograph shows peaks for ten steroids that were first separated on reverse-phase liquid chromatography using a Supelcosil LC-18 column (Supelco, Bellefonte, Calif.) and a gradient elution of a 53% to 75% methanol/water mixture. The column eluate was fed directly into an electrospray ionization device in a triple-quadrupole mass spectrometer (API 3000, Perkin-Elmer Sciex, Foster City, Calif.). The stable isotopes were from Cambridge Isotope Laboratories (Andover, Mass.). A 10-minute analysis provides quantitation of the 10 compounds: cortisone, cortisol, 21-deoxycortisol, corticosterone, 11-deoxycortisol, androstenedione, deoxycorticosterone (DOC), 17-hydroxyprogesterone, progesterone, and pregnenolone. The sensitivity for cortisol using d4 cortisol calibration is 0.1 µg/dL.
Nucleic Acid-Based Assays

The decoding of the human genome has set the stage for an enormous increase in nucleic acid-based gene assays. The basic principles of nucleic acid-based assays have been known for several decades, but the identification of specific genes and the mapping of gene defects to clinical disease states have now made these measurements clinically useful.\textsuperscript{48,49} Four concepts important for nucleic acid measurements are (1) hybridization, (2) amplification, (3) restriction fragment length polymorphisms (RFLPs), and (4) electrophoretic separation.\textsuperscript{50}

Hybridization

Nucleic acid molecules have a unique ability to fuse with complementary base-pair sequences. When a fragment of a known sequence (probe) is mixed under specific conditions with a specimen containing a complementary sequence, hybridization occurs. This feature is analogous to the antibody-antigen binding used in immunoassays. Many of the formats used for immunoassay have been adopted to nucleic acid assays, including some of the same signal systems (e.g., radioactivity, fluorescence, chemiluminescence) and the same solid-phase capture systems (e.g., magnetic beads, biotin-streptavidin binding). In situ hybridization, which involves the binding of probes to intact tissue and cells, provides information about morphologic localization analogous to immunohistochemistry.

Amplification

Nucleic acid assays have an advantage that low concentrations can be amplified in vitro prior to quantitation. The best known amplification procedure is the polymerase chain reaction (PCR), first reported by Mullis and Faloona.\textsuperscript{51} The three steps in the process (denaturation, annealing, and elongation) occur rapidly at different temperatures. Each "cycle" of amplification can occur in less than 90 seconds by cycling the temperature. The target double-stranded DNA is denatured at high temperature to make two single-stranded DNA fragments. Oligonucleotide primers, which are specific for target region, are annealed to the DNA when the temperature is lowered. Addition of DNA polymerase allows the primer DNA to extend across the amplification region, thus doubling the number of DNA copies.

At 85% to 90% efficiency, this process can amplify the DNA by about 250,000-fold in 20 cycles. This huge amplification is subject to major problems with contamination if special precautions are not taken. In one control technique, a psoralen derivative is used to prevent subsequent copying by polymerase during exposure to ultraviolet light.

Restriction Fragment Length Polymorphisms

Some diseases (e.g., sickle cell anemia) are associated with a specific gene mutation; generally, however, a series of deletions and additions of DNA are involved with the disease. A number of restriction enzymes that cleave DNA at specific locations have been identified. Changes in the sequence of DNA result in different fragment lengths. This technique, or RFLP, is particularly helpful in family studies for disorders that have a unique genetic fingerprint.

Electrophoretic Separation

E. M. Southern invented an electrophoretic separation technique known as Southern blotting.\textsuperscript{52} Restriction enzymes are used to digest a sample of DNA into fragments, and the product is subjected to electrophoresis. The separated bands of DNA are then transferred to a solid support and hybridized. Northern blotting is a similar technique, in which RNA is used as the starting material. Western blotting refers to electrophoresis and transfer of proteins.
ANALYTIC VALIDATION

Clinicians generally assume that laboratory methods have been validated and that they function correctly. Although this assumption is generally true, it is helpful to understand the level of assay validation performed and the appropriateness of the validation criteria for each clinical application of a test. These regulations, published in the Federal Register, outline the validation requirements for both Food and Drug Administration (FDA)-approved instruments, kits, and test systems as well as methods developed in-house. Laboratories must document analytic accuracy, precision, reportable ranges, and reference ranges for all procedures. The regulations for in-house procedures and modifications of approved commercial procedures are more extensive and require laboratories to further document (1) analytic sensitivity; (2) analytic specificity, including interfering substances; and (3) other performance characteristics required for testing patient specimens.

Although the details of method validation may be unique to a specific procedure, the following analytic validation studies have proved valuable for most procedures: (1) method comparison, (2) precision, (3) linearity, (4) recovery, (5) detection limit, (6) reportable range, (7) analytic interference, (8) carry-over, (9) reference interval, (10) specimen stability, and (11) specimen type. Laboratories should have documentation for each of these performance characteristics, either from the diagnostics manufacturer or from direct studies.

Method Comparison

Ideally, the system should be compared with an established reference method; however, many endocrine tests do not have reference methods and many laboratories do not have the facilities to perform reference methods when they exist. As a minimum, the assay should be compared with an analytic system that has been clinically validated with specimens from healthy subjects and specimens from patients with the diseases being investigated. The system should be traceable to established reference standards, such as those from the WHO and the National Institute of Standards and Technology (NIST). Between 100 and 200 different specimens distributed over the assay range are recommended for method comparisons.

A cross-plot displaying the new method on the vertical axis versus the established method on the horizontal axis, along with the identity line, reference value lines, and regression statistics, is a useful way of displaying these comparisons. An alternative display method is the Bland-Altman difference plot, in which the difference between the test method and the reference method is plotted against the reference method values.

Although acceptable performance criteria for method comparisons are not well established, some important characteristics to examine are as follows:

1. Any grossly discordant test values.
2. The degree of scatter about the regression curve.
3. The size of the regression off-set on the vertical axis.
4. The number of points crossing between the low, normal, and high reference intervals for the two methods.

The European Union (EU) has enacted the In Vitro Diagnostics Directive, which requires manufacturers marketing in the EU after the year 2003 to establish that their products are "traceable to reference standards and reference procedures of a higher order" when these references exist. This directive should serve to harmonize many test methods worldwide because most diagnostic companies market internationally.
Precision

Precision is a measure of the replication of repeated measurements of the same specimen; it is a function of the time between repeats and the concentration of the analyte. Both short-term precision (within a run or within a day) and long-term precision (across calibrations and across batches of reagents) should be documented at clinically appropriate concentration levels.

In general, normal range, abnormally low range, and abnormally high range targets are chosen for precision studies; however, targets focused on critical medical decision limits may be more appropriate for some analytes. Twenty measurements are recommended at each level for both short-term and long-term precision validations. Precision generally is expressed as the coefficient of variation, calculated as 100 times the standard deviation divided by the average of the replicate measurements.

There is no universal agreement on the performance criteria for analytic precision, although numerous recommendations have been put forth. Two major approaches to defining these criteria have been (1) comparison with biologic variation and (2) expert opinion of clinicians based on their perceived impact of laboratory variation on clinical decisions.

The total variation clinically observed in test measurements is a combination of the analytic and biologic variations, for instance:

1. If the analytic standard deviation (SD) is less than one-fourth of the biologic SD, the analytic component increases the SD of the total error by less than 3%.
2. If the analytic precision is less than one-half of the biologic SD, the total error increases by only 12%.

These observations have led to recommendations for maintaining precision less than one quarter or one-half of the biologic variation. The expert opinion precision recommendations are based on estimates of the magnitude of change of a test value that would cause clinicians to alter their clinical decisions. Table 6-2 lists some precision recommendations for selected endocrine tests.
Linearity

Patient specimens commonly contain several different forms of the hormones to be measured compared with the pure form contained in the reference standards and calibrators used to establish the assay dose-response curve. When a patient specimen is diluted, the measured value for these dilutions should parallel the dose-response curve and give results proportional to the dilution. Linearity can be evaluated by measuring serial dilutions of patient specimens with high concentrations diluted in the appropriate assay diluent. The product of the measured value multiplied by the dilution factor should be approximately constant. There are no performance standards for linearity, but a reasonable expectation for most hormones is that dilutions are comparable within 10% of the undiluted value.
Recovery

Two methods of assessing the recovery of assays are (1) measuring the increase in test values after the reference analyte is added and (2) measuring the proportional changes caused by mixing high-concentration and low-concentration specimens. Some analytes circulate in the blood in multiple forms, and some of these forms may be bound to carrier proteins. The recovery rate of pure substances added to a specimen may be low if the assay does not measure some of the bound forms. Mixtures of patient specimens may not be measured correctly if one of the specimens contains cross-reacting substances such as autoantibodies. A thorough understanding of the chemical forms of the analyte and their cross-reactivities in the assay is important during assessment of recovery data.
Detection Limit

The minimal analytic detection limit is the smallest concentration that can be statistically differentiated from zero. This concentration is mathematically determined as the upper 95% limit of replicate measurements of the zero standard, calculated from the average signal plus 2.0 SD. This minimal detection limit is valid only for the average of multiple replicate measurements. When individual determinations are performed on a specimen having a true concentration exactly at the minimal detection limit, the probability that the measurement is above the noise level of the assay is only about 50%.

A second term for the lowest level of reliable measurement for an assay is the functional detection limit, or the limit of quantitation. For this parameter to be measured, multiple pools with low concentrations are made and analyzed in the replicate. A cross-plot of the coefficient of variation of the measurements versus the concentration allows one to generate a precision profile. The concentration corresponding to a coefficient of variation of 20% is the functional detection limit. This term generally applies to across-assay variation, but it also can be calculated using within-assay variation if one uses the tests to evaluate results measured within one run (e.g., provocative and suppression tests).
Reportable Range

The reportable range of an assay generally spans from the functional detection limit to the concentration of the highest standard. Values above the highest standard may be reported if they are diluted and the measured value is multiplied by the dilution factor. The validity of the analytic range is documented by the linearity and recovery studies. Some laboratories erroneously report the exact values displayed by the test systems even if they are outside of the analytic range. Therefore, it is important for clinicians to understand the limitations of valid measurements and not inappropriately use meaningless numbers that may be reported.

Another potential source of error is failure of the technologist to multiply the measured value of diluted specimens by the dilution factor to correct for the dilution. In addition, care should be taken to define the number of significant figures used for reporting test values and to establish an appropriate algorithm for rounding test values to the significant number of digits.
Analytic Interference

The cross-reactivity and potential interference of other analytes that may react in a test system should be documented. The choice of potential interfering substances that must be evaluated requires an understanding of both the analytic system and the pathophysiology of the analyte being evaluated. In immunoassays, for example, compounds with similar structures as well as precursor forms and degradation products should be tested. Drugs commonly prescribed for the diseases under evaluation should be assessed for interference both by addition of the drug to a specimen and by analysis of specimens from patients before and after receiving the drug. Most assays also are evaluated for the effects of hemolysis, lipemia, and icterus.
Carry-over Studies

Many diagnostic systems use automated sample-handling devices. If a specimen to be tested is preceded by a specimen with a very high concentration, a trace amount of the first specimen may significantly increase the reported concentration of the second specimen. The choice of the concentration that should be tested for carry-over depends on the pathophysiology of the disease, but high values may need to be tested because some endocrine disorders may produce these high values. A prudent procedure would be to retest all specimens following a specimen with an extraordinarily high value. One also should document that carry-over from the sampling probe has not inadvertently contaminated subsequent specimen vials, thereby invalidating subsequently repeated measurements.
Reference Intervals

The development and validation of reference intervals for endocrine tests can be a very complex task. The normal reference interval for most laboratory tests is based on estimates of the central 95 percentile limits of measurements in healthy subjects. A minimum of 120 subjects is needed to reliably define the 2.5 and 97.5 percentiles. The reference intervals for many endocrine tests depend on gender, age, developmental status, and other test values. Formal statistical consultation is recommended to determine the appropriate number of subjects to test and to develop statistical models for defining multivariate reference ranges.

Full evaluation of the adrenal, gonadal, and thyroid axes requires simultaneous measurement of the trophic and target hormones. Bivariate displays of these hormone concentrations along with their multivariate reference intervals facilitate the interpretation. Preanalytic conditions should be well defined and controlled during evaluation of both healthy reference subjects and patients.

Figure 6-8 shows a recommended protocol to control preanalytic conditions for collection of plasma catecholamine specimens.
Specimen Stability

Analyte stability is a function of storage conditions and specimen type. Although most hormones are relatively stable in serum or urine if they are rapidly frozen and stored in hermetically sealed vials at -70°C, multiple freeze/thaw cycles may damage analytes, and storage in frost-free freezers that repeatedly cycle through thawing temperatures can adversely affect stability. Blood specimens collected in edetate (EDTA) often are more stable than serum or heparinized specimens because edetate chelates calcium and magnesium ions, which function as coenzymes for some proteases. The addition of protease inhibitors (e.g., aprotinin) to blood specimens may also improve specimen stability.
Types of Specimens

Most hormones are measured in blood or urine, but alternate testing sources, such as saliva and transdermal membrane monitors, are also used.

Urine Specimens

The 24-hour urine specimen is used for many endocrine tests. Urine specimens represent a time average that integrates over the multiple pulsatile spikes of hormone secretion occurring throughout the day. The 24-hour urine specimen also has the advantage of better analytic sensitivity for some hormones. Urine often contains not only the original hormone but also key metabolites that may or may not have biologic activity.

Drawbacks include the inconvenience of and delays in collecting the 24-hour specimen. Another limitation of urine specimens is the uncertainty of the completeness of the collection. Measurement of urinary creatinine concentrations helps in monitoring collection completeness, especially when it is compared with the patient’s muscle mass. Many urinary hormones are conjugated to carrier proteins before excretion. Therefore, both hepatic function and, to a lesser degree, renal function may alter urinary hormone values.

Blood Specimens

Blood specimens have both the advantage and the limitation of time dependency. The ability to direct rapid changes to a provocative stimulus is a strong advantage, whereas the unsuspected changes due to pulsatile secretions may be a major limitation. Most hormones undergo significant biologic variations, including ultradian, diurnal, menstrual, and seasonal changes. Many urinary hormones have short half-lives and are thus rapidly cleared from the blood. The half-life is particularly important when one is attempting to measure the response to a provocative drug, such as the effect of gonadotropin-releasing hormone (GnRH). The development of rapid intraoperative methods for measuring PTH and growth hormone has highlighted the importance of plasma specimens, which do not require extra waiting time for the blood to clot to make serum.

Saliva Specimens

Saliva has been used to measure some hormones. Methods of stimulation, collection, and storage of saliva should be standardized in order to ensure that the measurements are reproducible and meaningful. Saliva measurements correlate with blood measurements in some hormones like cortisol, progesterone, estradiol and testosterone but do not correlate well for others. Unconjugated steroid hormones enter saliva by diffusion, and their concentrations are relatively independent of the rate of saliva production. The saliva concentration of conjugated steroids, thyroxine, chorionic gonadotropin, and many protein hormones generally do no correlate well with plasma concentrations.

Blood Drops

Blood drops collected on filter paper from punctures of a finger or heel are a convenient system for collecting, transporting, and measuring hormones. If standardized collection conditions and extraction techniques are used, these measurements correlate well with serum measurements. Integration of immunochemistry with computer chip technology has also led to immunochips that can measure multiple analytes using a single drop of blood.

Noninvasive Measurements

Noninvasive transcutaneous measurements also have been developed for some endocrine tests. Transcutaneous glucose measurements using near-infrared spectroscopy correlate well with blood measurements. The GlucoWatch device is also being marketed for noninvasive monitoring of glucose.
QUALITY ASSURANCE

Quality Control Systems

Laboratory quality control programs are intended to ensure that the test procedures are being performed within defined limits. A critical component of control systems is the definition of acceptable performance criteria. Unfortunately, these criteria often are not well defined and many laboratories use floating criteria that change when assays change. Control limits are often set at the mean ±2 or 3 SDs, where the mean and SD are arbitrarily assigned based on measurements made in that laboratory. When reagents or equipment change, new limits are assigned. These types of control systems provide some assurance that the laboratory is functioning at a level of performance similar to that of the recent past, but they provide little assurance that measurements are adequate for clinical decisions.

Statistically, there are two major forms of analytic errors: random and systematic. Random error relates to reproducibility; systematic error relates to the offset or bias of the test values from the target or reference value. Performance criteria can be defined for each of these parameters, and quality control systems can be programmed to monitor compliance with these criteria. Control systems must have low false-positive rates as well as high statistical power to detect assay deviations. The multirule algorithms developed by Westgard and colleagues use combinations of control rules, such as two consecutive controls outside of warning limits, one control outside of action limits, or moving average trend analyzers outside of limits to achieve good statistical error detection characteristics.

Traditionally, quality control programs have focused primarily on precision; however, analytic bias also can cause major clinical problems. When fixed decision levels are used to trigger clinical actions, such as therapy and additional investigations, changes in the analytic set-point of an assay can cause major changes in the number of follow-up cases. This concept is illustrated in Figure 6-9 for TSH measurements.

Under stable laboratory testing conditions, approximately 122 per 1000 patients tested have TSH values above 5.0 mIU/L. If the test shifts upward by 20%, the number of patients with TSH values above 5.0 mIU/L increases to 189, which equates to more than a 50% increase in the number of patients flagged as abnormal. These changes in test value distributions can often be sensed by clinicians who encounter multiple patients with unexpected elevated test values, causing them to call the laboratory and inquire whether the "test is running high today." Some modern quality control systems use moving averages of patient test values to help monitor changes in analytic bias.

Some medical facilities are linking together into networks to provide more integrated patient care. This crossover of both physicians and patients is increasing the importance of harmonized testing systems. For endocrine tests, harmonization is best achieved when all the laboratories in the network use the same test systems. Differences in analytic specificity may cause across-method differences in patient test distributions even when the methods use the same reference standard. Full harmonization of testing requires not only standardization of equipment but also standardization of reagents (including using the same lot numbers) and standardization of laboratory protocols. Real-time quality control monitors with peer group comparisons across the laboratories in the health care network are necessary to ensure uniformity of testing.
Investigation of Discordant Test Values

The practice of modern endocrinology depends extensively on reliable and accurate test values; even in the best laboratories, however, erroneous results sometimes are reported. Careful correlation of pathophysiology with test values can help to identify values that are "discordant." Some of these discordant test values may be analytically correct, but others may be erroneous. Clinicians can help investigate these suspicious test values by requesting laboratories to perform a few simple validation procedures.

Repeated testing of the same specimen is a valuable first step. If the specimen has been stored under stable conditions, the absolute value of the difference between the initial and the repeated measurements should be less than 3 analytic SDs 95% of the time. Normally, the 95% confidence range is associated with the mean ±2 SDs; with repeated laboratory tests, however, errors are associated with the first as well as the second measurement. The confidence interval for the uncertainty of the difference between two measurements can be calculated using the statistical rules for propagation of errors.

To better understand this propagation of error, consider that 

\[ D = X_1 - X_2 \]

where \( X_1 \) is the first measurement, \( X_2 \) is the repeated measurement, and \( D \) is the difference.

The variance of \( D \) is the sum of the variance of \( X_1 \) and the variance of \( X_2 \). The SD of \( D \) is the square root of the variance of \( D \), or the square root of twice the variance of \( X_1 \). The SD of \( D \) equates to 2 multiplied by the SDs of \( S \). Therefore, 95% of the absolute values for \( D \) should be within 22 SD(X), or approximately 3 SD(X). If a repeat measurement exceeds this 3 SD(X) limit, the initial (or reagent) measurement is probably in error.

Linearity and recovery are valuable techniques for evaluating test validity. If the initial test value is elevated, serially diluting the specimen in the assay diluent and reassaying should be considered. If the specimen dilutes nonproportionally (Fig. 6-10), no meaningful value can be reported with that assay.

In the example, the undiluted specimen reads 22, the twofold dilution multiplies back to 34 (2 × 17), and the fourfold dilution multiplies back to 60 (4 × 15). Therefore, the result depends on the dilution factor, so that no reliable answer can be reported.

If the initial value is low, one may consider adding known quantities of the analyte to part of the specimen. Analyzing these spiked or diluted specimens with the original specimen allows one to evaluate both reproducibility and recovery. It may be helpful to analyze the linearity or recovery of the assay standards at the same time to provide internal controls of the dilution or spiking procedures and the appropriateness of the diluent and spiking material.

If the replication, dilution, or recovery experiment appears successful, further analytic troubleshooting will vary according to the method used. Immunoassays may be affected by interference caused by heterophile antibodies. Addition of nonimmune mouse serum or heterophile antibody-blocking solutions may neutralize these effects. Chromatographic assays are usually more robust than immunoassays. Specimens with suspected interference on one type of assay can be reanalyzed by means of an alternative methodology.

Water-soluble interferences have been reported for some direct assays for steroid measurements. Extraction of the hormones into organic solvents, followed by drying down and reconstitution in the assay zero standard, removes these interferences. Similarly, interferences with cross-reacting drugs and metabolic products can be minimized with selective extraction.
Summary

The analytic methods of assessing endocrine problems in patients are continually expanding. The newer systems are often based on analytic techniques similar to those outlined in this chapter, but the configurations are generally more user-friendly. These advances make the systems more convenient, but they also become more of a "black box" that conceals most of the details of the system. The performance validation steps outlined in this chapter become important procedures for ensuring that these systems continue to provide the reliable measurements needed for quality medical care.
References


64. Fraser CG, Petersen PH. Analytical performance characteristics should be judged against objective quality specifications. Clin Chem 1999; 45:321333.
This chapter first presents the concepts of neural secretion, the neuroanatomy of the hypothalamic-pituitary unit, and the CNS structures most relevant to the control of endocrine, autonomic, and behavioral responses. Thus, many homeostatic systems exist in which the classical neuroendocrine axes are important but not the only contributors. These areas include studies of neuropeptide structure, function, and mechanism of action; neural secretion; hypothalamic neuroanatomy; G protein-coupled receptor function and signaling; transport of substances into the brain; and the action of hormones on the brain. Many homeostatic systems involve integrated neural, endocrine, and behavioral mechanisms, demonstrating that much remains to be learned regarding the regulation of the hypothalamic releasing factors.

This technology will allow detailed study of important neuroendocrine neurons in the more native context of slice preparations or organotypic cultures. For example, investigators have already used this method to characterize directly the electrophysiologic properties of individual GnRH neurons. The adipostatic hormone leptin, discovered in 1994, is an example of a humoral factor that has profound effects on multiple neuroendocrine circuits as the factor that suppresses the thyroid and reproductive axes during the starvation response. The subsequent discovery of ghrelin, a stomach peptide that regulates appetite and also acts on multiple neuroendocrine axes, demonstrates that much remains to be learned regarding the regulation of the hypothalamic releasing hormones. Traditionally, it has been extremely difficult to study the regulation of releasing factor gene expression or the specific regulation of the releasing factor binding proteins. Traditionally, this has been problematic due to the rapidity of the binding proteins. Additionally, there is evidence that several neuroendocrine axes are involved in the control of the pituitary gland, without damage to the overlying hypothalamus, resulting in morbid obesity and neuroendocrine derangements similar to those of the patients described by Fröhlich. This and subsequent studies clearly established that an intact hypothalamus is required for normal endocrine function. However, the mechanisms by which the hypothalamus was involved in endocrine regulation remained unsettled for years to come. We now know that the phenotypes of Fröhlich’s syndrome and the ventromedial hypothalamic lesion syndrome are probably due to destruction of key hypothalamic neurons that respond to key metabolic signals including leptin. We now appreciate the fundamental role of the hypothalamus in controlling anterior pituitary function. It is noteworthy that this concept is relatively recent, although the fundamental role of the hypothalamus in controlling anterior pituitary function was long recognized. Over the past two decades, work in the field of neuroendocrinology has continued to advance across several fronts. Cloning and characterization of the specific G protein-coupled receptors used by the hypothalamic releasing factors have helped define signaling mechanisms utilized by the releasing factors. Furthermore, characterization of the distribution of these receptors has, in every case, demonstrated receptor expression in the brain and in peripheral tissues other than the pituitary, arguing for multifactorial roles for these factors. Finally, the last two decades have also seen tremendous advances in our understanding of both regulatory neuronal and humoral inputs to the hypophyseotropic neurons.

The field of neuroendocrinology has been further expanded, however, because many areas of basic research have often been fundamental to understanding the neuroendocrine system and thus championed by scientists in the field. These areas include studies of neuropeptide structure, function, and mechanism of action; neural secretion; hypothalamic neuroanatomy; G protein-coupled receptor structure, function, and signaling; transport of substances into the brain; and the action of hormones on the brain. Many homeostatic systems involve integrated endocrine, autonomic, and behavioral responses. Thus, many homeostatic systems exist in which the classical neuroendocrine axes are important but not autonomous pathways, such as energy homeostasis and immune function, and these subjects are also often studied in the context of neuroendocrinology.

This chapter first presents the concepts of neural secretion, the neuroanatomy of the hypothalamic-pituitary unit, and the CNS structures most relevant to the control of endocrine, autonomic, and behavioral responses.
of the neurohypophysis and hypophysis. The chapter then covers each classical hypothalamic-pituitary axis, followed by two homeostatic systems, energy homeostasis and immune function, which are heavily integrated with neuroendocrine function. Finally, the chapter reviews the pathophysiology of disorders of neural regulation of endocrine function.
NEURAL CONTROL OF GLANDULAR SECRETION

A fundamental principle of neuroendocrinology is the concept of regulated secretion of hormones, neurotransmitters, or neuromodulators by secretory cells. Endocrine cells and neurons are prototypical secretory cells, and both are characterized by the ability to be stimulated to cause the release of their products. In addition, secretory cells exist that can be broadly classified by their mechanisms of secretion. For example, endocrine cells secrete their contents directly into the blood stream, allowing these substances to act globally as hormones. In contrast, secretory cells in exocrine glands secrete substances into ductal systems. Cells classified as paracrine secrete their contents and affect the function of cells in the immediate vicinity. Similarly, autocrine secretory cells affect their own function by the local actions of their own secretions.

Neurosecretion

Neurons are specialized secretory cells that send their axons throughout the nervous system to release their neurotransmitters and neuromodulators into chemical synapses. A specialized subset of neurons are the neurohumoral or neurosecretory cells. Two examples of neurosecretory cells are neurohypophyseal and hypothalamic cells. The prototypical neurohypophyseal cells are the magnocellular neurons of the paraventricular and supraoptic nuclei in the hypothalamus. Hypophyseotropic cells are neurons that secrete their products into the pituitary portal vessels at the median eminence (see later).

In the most basic sense, neurosecretory cells are neurons that secrete substances directly into the blood stream to act as hormones. This concept of release is often referred to as neurosecretion (Fig. 7-2). The theory of neurosecretion evolved from the seminal work of Scharrer and Scharrer, who used morphologic techniques to identify stained secretory granules in the supraoptic and paraventricular hypothalamic neurons. They found that cutting the pituitary stalk led to an accumulation of these granules in the hypothalamus. These findings led them to hypothesize that the source of substances secreted by the neural lobe (posterior pituitary) was hypothalamic neurons. Of course, we now know that the axon terminals in the neural lobe arise from the supraoptic and paraventricular magnocellular neurons that contain oxytocin and arginine vasopressin (AVP).

The modern definition of neurosecretion has evolved to include the release of any neuronal secretory product from a neuron. Indeed, a basic principle of neuroscience is that all neurons in the CNS, including neurons that secrete AVP and oxytocin in the neural lobe, receive multiple synaptic inputs largely onto their dendrites and cell bodies. In addition, neurons have the basic ability to respond and integrate input from multiple neurons through specific receptors. They in turn fire action potentials that result in the release of neurotransmitters and neuromodulators into synapses formed with postsynaptic neurons. The vast majority of communication between neurons is accomplished by "classical" neurotransmitters (e.g., glutamate, -aminobutyric acid [GABA], acetylcholine) and neuromodulators (e.g., neuropeptides) acting at chemical synapses (see Fig. 7-2). Thus, neurosecretion represents a fundamental concept in understanding the mechanisms used by the nervous system to control behavior and maintain homeostasis.

In the era of elucidation of the human genome, the importance of these early observations is often not fully appreciated. However, accounts of these early studies are illuminating. Moreover, it is not an overstatement that the confirmation of the neurosecretion hypothesis represented one of the major advances in the field of neuroscience and neuroendocrinology. Indeed, this and other early experiments, including the pioneering work of Geoffrey Harris, led to the fundamental concept that the hypothalamus releases hormones directly into the blood stream (neurohypophyseal cells). These observations provided the principles on which the modern discipline of neuroendocrinology is built.
The Autonomic Nervous System Contribution to Endocrine Control

One of these fundamental principles of neuroendocrinology is that the nervous system controls or modifies, or both, the function of both endocrine and exocrine glands. The exquisite control of the anterior pituitary gland is accomplished by the release of releasing factor hormones (see later). Other endocrine and exocrine organs (e.g., pancreas, adrenal, pineal, salivary glands) are also regulated through direct innervation from the cholinergic and noradrenergic inputs from the autonomic nervous system. Although it is beyond the scope of this chapter, an appreciation of the functional anatomy and pharmacology of the parasympathetic and sympathetic nervous systems is fundamental in understanding the neural control of endocrine function.

The efferent arms of the autonomic nervous system comprise the sympathetic and parasympathetic systems. Both limbs are a classical two-neuron chain. Both are characterized by a preganglionic neuron that innervates a postganglionic neuron that targets an end organ. Preganglionic and postganglionic parasympathetic neurons are cholinergic. In contrast, preganglionic sympathetic neurons are cholinergic and postganglionic neurons are noradrenergic (except for those innervating sweat glands, which are cholinergic). Another basic concept is that autonomic neurons coexpress several neuropeptides. This coexpression is a common feature in neurons in both the central and peripheral nervous systems. For example, postganglionic noradrenergic neurons coexpress somatostatin and neuropeptide Y (NPY). Postganglionic cholinergic neurons coexpress neuropeptides including vasoactive intestinal polypeptide and calcitonin gene-related peptide.

The majority of the sympathetic preganglionic neurons lie in the intermediolateral cell column in the thoracicolumbar regions of the spinal cord. Most postganglionic neurons are located in sympathetic ganglia lying near the vertebral column (e.g., sympathetic chain and superior cervical ganglia). Postganglionic fibers, in turn, innervate target organs. Thus, as a rule, sympathetic preganglionic fibers are relatively short and the postganglionic fibers are long. In contrast, the parasympathetic preganglionic neurons lie in the midbrain (Edinger-Westphal nucleus of the third cranial nerve), the medulla oblongata (e.g., dorsal motor nucleus of the vagus and nucleus ambiguus), and the sacral spinal cord. Postganglionic neurons that innervate the eye and salivary glands arise from the ciliary, pterygopalatine, submandibular, and otic ganglia. Postganglionic neurons in thorax and abdomen typically lie in the target organs including the gut wall and pancreas. Thus, preganglionic neurons are relatively long and the postganglionic fibers are short.

The importance of coordinated neural control of endocrine organs is illustrated by the innervation of the pancreas. The endocrine pancreas receives both parasympathetic (cholinergic) innervation and sympathetic (noradrenergic) innervation. The cholinergic innervation is provided by the vagus nerve (dorsal motor nucleus of the vagus). The activity in this innervation is an excellent example of neural modulation as it is clear that the secretory activity of insulin-producing beta cells is affected by the cholinergic tone of the beta cell. For example, vagal input is thought to modulate insulin secretion before (cephalic phase), during, and after ingestion of food. In addition, noradrenergic stimulation of the endocrine pancreas can alter the secretion of glucagon and inhibits insulin release. It should be noted, of course, that a major regulator of insulin secretion by beta cells is glucose concentrations. In fact, glucose can induce insulin secretion in the absence of neural input. However, the exquisite control by the nervous system is illustrated by the fact that populations of neurons in the brain stem and hypothalamus, like the beta cell, have the ability to sense glucose levels in the blood stream. This information is integrated by the hypothalamus and ultimately results in alterations in the activity of the autonomic nervous system innervating the pancreas. Thus, neural control of the endocrine pancreas probably contributes to the physiologic control of insulin secretion and may contribute to the pathophysiology of disorders such as diabetes mellitus. Certainly, an increased understanding of this complex interplay between the CNS and endocrine function is needed to diagnose and clinically manage endocrine disorders.
HYPOTHALAMIC-PITUITARY UNIT

The hypothalamus is one of the most evolutionarily conserved and essential regions of the mammalian brain. Indeed, the hypothalamus is the ultimate brain structure that allows mammals to maintain homeostasis, and destruction of the hypothalamus is not compatible with life. Hypothalamic control of homeostasis stems from the ability of this collection of neurons to orchestrate coordinated endocrine, autonomic, and behavioral responses. A key principle is that the hypothalamus receives sensory inputs from the external environment (e.g., light, pain, temperature, odorants) and information regarding the internal environment (e.g., blood pressure, blood osmolality, blood glucose levels). In addition, of particular relevance to neuroendocrine control, hormones (e.g., glucocorticoids, estrogen, testosterone, thyroid hormone) exert negative feedback directly on the hypothalamus.

These sensory and hormonal cues are examples of a fundamental concept of neuroendocrinology: the hypothalamus integrates sensory and hormonal inputs and provides coordinated responses through motor outputs to key regulatory sites. These include the anterior pituitary gland, the posterior pituitary gland, the cerebral cortex, premotor and motor neurons in the brain stem and spinal cord, and autonomic (parasympathetic and sympathetic) preganglionic neurons. The patterned hypothalamic outputs to these effector sites ultimately result in coordinated endocrine, behavioral, and autonomic responses that maintain homeostasis. The focus of this section, the hypothalamic control of the pituitary gland, is an exquisitely controlled system and underlies the ability of mammals to coordinate endocrine functions that are necessary for survival.

Anatomy of the Hypothalamic-Pituitary Unit

The pituitary gland is regulated by three interacting elements: hypothalamic inputs (releasing factors or hypophyseotropic hormones), feedback effects of circulating hormones, and paracrine and autocrine secretions of the pituitary itself. In humans, the pituitary gland (hypophysis) can be divided into two major parts, the adenohypophysis and the neurohypophysis. The adenohypophysis in turn can be subdivided into three distinct lobes, the pars distalis (anterior lobe), pars intermedia (intermediate lobe), and pars tuberalis (Fig. 7-3). Whereas a well-developed intermediate lobe is found in most mammals, only rudimentary vestiges of the intermediate lobe are detectable in adult humans with the bulk of intermediate lobe cells being dispersed in the anterior and posterior lobes.

The neurohypophysis is composed of the pars nervosa (also known as the neural or posterior lobe), the infundibular stalk, and the median eminence. The infundibular stalk is surrounded by the pars tuberalis, and together they constitute the hypothyalalal stalk. The pituitary gland lies in the sella turcica (the Turkish saddle) of the sphenoid bone and underlies the base of the hypothalamus. This anatomic location explains the hypothalamic damage described by Fröhlich. In humans, the base of the hypothalamus forms a mound called the tuber cinereum, the central region of which gives rise to the median eminence (Fig. 7-4).

The anterior and intermediate lobes of the pituitary are derived from an outgrowth of the pharyngeal cavity called Rathke's pouch and migrate during development to surround the neural lobe. The intermediate lobe is in contact with the neural lobe and is the least prominent of the three lobes. With age, the intermediate lobe in humans decreases in size and is represented in the adult as a relatively small collection of POMC cells. In some species, these cells are responsible for secreting the POMC-derived product melanocyte-stimulating hormone (MSH).

In a strict sense, the neurohypophysis is made up of the neural lobe, the infundibular stalk, and the median eminence. The major component of the neural lobe is a collection of axon terminals arising from magnocellular secretory neurons from the paraventricular and supraoptic nuclei of the hypothalamus (Fig. 7-5). The blood supply to the neurohypophysis arises from the inferior hypophyseal vessels and the general circulation (Table 7-1). The blood supply to the neurohypophysis arises from the internal carotid artery. Scattered among the nerve terminals are glial-like cells called pituicytes. As the source of AVP to the general circulation, the paraventricular and supraoptic nuclei and their axon terminals in the neural lobe are the effector arms of the central regulation of blood osmolality, fluid balance, and blood pressure. (see "Circumventricular Organs").

The secretion of oxytocin by magnocellular neurons is also well characterized and is critical at the time of parturition, resulting in uterine myometrial contraction. In addition, the secretion of oxytocin is regulated by the classical milk let-down reflex. The exact neuroanatomic substrate underlying the milk let-down response is still unclear. However, it is apparent that mechanosensory information from the nipple reaches the magnocellular neurons, directly or indirectly, from the dorsal horn of the spinal cord, resulting in release of oxytocin into the general circulation. Oxytocin acts on receptors on myoepithelial cells in the mammary gland acini, leading to release of milk into the ductal system and ultimately the release of milk from the mammary gland.
The Median Eminence and Hypophyseotropic Neuronal System

The median eminence lies in the center of the tuber cinereum; it is composed of an extensive array of blood vessels and nerve endings and is the functional link between the hypothalamus and the anterior pituitary gland. The median eminence can be considered the functional link between the hypothalamus and pituitary and the site of the hypothalamic releasing factors from which the pituitary portal vessels arise. The median eminence is characterized by an extremely rich blood supply that arises from the superior hypophyseal artery (from the internal carotid artery). The artery sends off many small branches that form capillary loops. These small capillary loops extend into the internal and external zones (see later), form anastomoses, and drain into sinusoids that become the pituitary portal veins that enter the vascular pool of the pituitary gland. The flow of blood in these short loops is thought to be predominantly (if not exclusively) in a hypothalamic-to-pituitary direction. This well-developed plexus results in a tremendous increase in the vascular surface area. In addition, the vessels are fenestrated, allowing diffusion of the peptide-releasing factors to their site of action in the anterior pituitary gland. The vascular complex in the base of the hypothalamus and its “arterialized” venous drainage to the pituitary compose a circulatory system analogous to the portal vein system of the liver, hence the term hypophyseal portal circulation. Typically, three zones of the median eminence are discussed, the ependymal layer, the internal zone, and the external zone (Fig. 7-8). The innermost zone is made up of ependymal cells that form the floor of the third ventricle. These ependymal cells are unique in that they have microvilli rather than cilia. The ependymal layer also contains specialized cells called tanycytes that send processes into the other layers of the median eminence. There are tight junctions between the ependymal lining of the third ventricle forming a barrier between the cerebrospinal fluid (CSF) and the blood in the median eminence. In addition, tight junctions exist between tanycytes at the lateral edges of the median eminence that are thought to prevent the diffusion of releasing factors back into the basal hypothalamus.

The internal zone of the median eminence is composed of axons of passage of the supraoptic and paraventricular magnocellular neurons en route to the posterior pituitary (see Fig. 7-3C).

and the axons of the hypophyseotropic neurons destined for the external layer of the median eminence (see Fig. 7-8A and B). In addition, supportive cells are found in this layer. Finally, the external zone represents the exchange point of the hypothalamic releasing factors and the pituitary portal vessels. The external zone contains terminals from two general types of releasing factors, peptides (discussed in detail later) and monoamines (dopamine and norepinephrine). This zone represents the site of convergence where the peptides come in contact with portal vessels.

Two general types of tuberhypophyseal neurons project to the external zone of the median eminence: peptide-secreting (peptidergic) neurons (i.e., thyrotropin-releasing hormone [TRH], corticotropin-releasing hormone [CRH], and luteinizing hormone-releasing hormone [LHRH]; see Fig. 7-7 and Fig. 7-8) and neurons containing bioamines (e.g., dopamine and serotonin). Although the secretion of these substances into the portal circulation is an important control mechanism, some peptides and neurotransmitters in nerve endings are not released into the hypophyseal-portal circulation but instead function to regulate the secretion of other nerve terminals. The anatomical relationships of nerve endings, basement membranes, interstitial spaces, fenestrated (windowed) capillary endothelia, and glia in the median eminence are similar to those in the neural lobe. As in the case of the neurohypophysis, the release of neuropeptides is mediated by the depolarization of hypothalamic cells leading to secretion at the median eminence.

Non-neuronal supporting cells in the hypothalamus also play a dynamic role in hypothyseotropic regulation. For example, nerve terminal cells in the neurohypophysis are enveloped by glia (in the neural lobe they are called pituicytes); when the gland is inactive they surround the nerve endings, whereas the nerve ending is exposed when vasopressin secretion is enhanced as in states of dehydration. Within the median eminence, LHRH nerve endings are enveloped by the specialized ependymal cells called tanycytes, which also cover or uncover neurons with changes in functional status. Thus, supporting elements, with their own sets of receptors, can change the neuroregulatory milieu within the hypothalamus, median eminence, and pituitary.

The site of production, the genetics, and the regulation of synthesis and release of peptide releasing factors are discussed in detail in the following. Briefly, the cell groups in the hypothalamus that contain releasing factors that are secreted into the pituitary portal circulation are located in several cell groups of the medial hypothalamus (Table 7-2). These cell groups include the arcuate (infundibular) nucleus (see Fig. 7-9C), the paraventricular nucleus (see Fig. 7-9A and C), the periventricular nucleus, and a group of cells in the medial preoptic area near the organum vasculosum of the lamina terminals (OVLT) (Fig. 7-10). As discussed, magnocellular neurons in the supraoptic and paraventricular nucleus send axon terminals that traverse the median eminence and make up the neural lobe of the pituitary. In addition, a projection from magnocellular neurons to the external zone of the median eminence has been described. However, its functional significance is not clear.

The third structure often mentioned as a component of the median eminence is the pars tuberalis. The pars tuberalis is a subdivision of the adenohypophysis and is a thin glandular sheet of tissue that lies around the infundibulum and pituitary stalk. In some animals, the epithelial component may make up as much as 10% of the total glandular tissue of the anterior pituitary. The pars tuberalis contains cells making pituitary tropic hormones including luteinizing hormone (LH) and thyrotropin. A definitive physiologic function of the pars tuberalis is not established, but melatonin receptors are expressed in the pars tuberalis.
CIRCUMVENTRICULAR ORGANS

A fundamental principle of physiology and pharmacology is that the brain, including the hypothalamus, resides in an environment that is protected from humoral signals. This exclusion of macromolecules is due to the structural

specializations that make up the blood-brain barrier. These include tight junctions of brain vascular endothelial cells that preclude the free passage of polarized macromolecules including peptides and hormones. In addition, astrocytic foot processes and perivascular microglial cells contribute to the integrity of the blood-brain barrier. However, to exert homeostatic control, the brain, especially the hypothalamus, must assess key sensory information from the blood stream including hormone levels, metabolites, and potential toxins. For example, to monitor key signals the brain has "windows on the circulation" or circumventricular organs (CVOs) that serve as a conduit of peripheral cues into key neuronal cell groups that maintain homeostasis.

As the name implies, CVOs are specialized structures that lie on the midline of the brain along the third and fourth ventricles. These structures include the OVLT, subcomical organ (SFO), median eminence, neurohypophysis (posterior pituitary), subcomissural organ, and the area postrema. Unlike thevasculature in the rest of the brain, the blood vessels in CVOs have fenestrated capillaries that allow relatively free passage of molecules such as proteins and peptide hormones. Thus, neurons and glial cells that reside within the CVOs have access to these macromolecules. In addition to the distinct nature of the vessels themselves, the CVOs have an unusually rich blood supply, allowing them to act as integrators at the interface of the blood-brain barrier. As discussed in more detail later, several of the CVOs have major projections to hypothalamic nuclear groups that regulate homeostasis. Thus, the CVOs serve as a critical link between peripheral metabolic cues, hormones, and potential toxins with cell groups within the brain that regulate coordinated endocrine, autonomic, and behavioral responses. Detailed discussion of the physiologic roles of individual CVOs is beyond the scope of this chapter, but several in-depth reviews have assessed the function of each.

Median Eminence

The median eminence and neurohypophysis contain the neurosecretory axons that control pituitary function. The role of the median eminence as a link between the hypothalamus and the pituitary gland is covered in greater detail in other sections of this chapter (see Fig. 7-9) and Fig. 7-9 and "Hypothalamic-Pituitary Unit"). However, it is important to understand that the anatomic location of the median eminence places it in a position to serve as an afferent sensory organ as well. Specifically, the median eminence is located adjacent to several neuroendocrine and autonomic regulatory nuclei at the tuber level of the hypothalamus (see Fig. 7-9). These nuclear groups include the arcuate, ventromedial, dorsomedial, and paraventricular nuclei.

A role of cell groups surrounding the median eminence as afferent sensory centers is supported by several observations. For example, toxins such as monosodium glutamate and gold

thioglucose damage neurons in cell groups overlying the median eminence, resulting in obesity and hyperphagia. Experimental evidence suggests that the median eminence is a portal of entry for hormones such as leptin. Indeed, administration of radiolabeled peptides or hormones, such as -MSH or leptin, led to their accumulation around the median eminence. Moreover, leptin receptor messenger ribonucleic acid (mRNA) and leptin-induced gene expression are densely localized in the arcuate, ventromedial, dorsomedial, and ventral premammillary hypothalamic nuclei. As discussed in detail in other sections of this chapter, leptin is an established mediator of body weight and neuroendocrine function that acts on several cells in the hypothalamus including POMC neurons that reside in the arcuate nucleus (see Fig. 8-9). Notably, POMC neurons are also found embedded within the median eminence. Thus, it is likely that the median eminence is involved in conveying information from humoral factors such as leptin to key hypothalamic regulatory neurons in the medial basal hypothalamus.

| TABLE 7-1 – Sequences of the Principal Peptides of the Neurohypophysis |
|---|---|---|---|---|---|---|---|---|---|
| **Mammals (except pig)** | Cys-Tyr-Ile-Gln-Asn-Cys-Pro-Leu-Gly-NH₂ | Cys-Tyr-Phe-Gln-Asn-Cys-Pro-Arg-Gly-NH₂, Arginine vasopressin |
| **Pig** | Cys-Tyr-Ile-Gln-Asn-Cys-Pro-Leu-Gly-NH₂ | Cys-Tyr-Phe-Gln-Asn-Cys-Pro-Lys-Gly-NH₂, Lysine vasopressin |
| **Birds, reptiles, amphibians, lungfishes** | Cys-Tyr-Ile-Gln-Asn-Cys-Pro-Ile-Gly-NH₂ | Cys-Tyr-Ile-Gln-Asn-Cys-Pro-Arg-Gly-NH₂, Vasotocin |
| **Bony fishes (palcopterygians and neopterygians)** | Cys-Tyr-Ile-Ser-Asn-Cys-Pro-Ile-Gly-NH₂, Isotocin | Cys-Tyr-Ile-Gln-Asn-Cys-Pro-Arg-Gly-NH₂, Vasotocin |

(From Bear MF, Connors BW, Paradiso MA. Neuroscience: Exploring the Brain. Baltimore, Williams & Wilkins, 1996, p 408.)
Thyrotropin-Releasing Hormone

Chemistry and Evolution

TRH, the smallest known peptide releasing hormone, is the tripeptide pyroGlu-His-Pro-NH₂. The TRH peptide sequence is repeated six times within the human TRH pre-prohormone gene (Fig. 7-13). The rat pro-TRH precursor contains five TRH peptide repeats flanked by dibasic residues (Lys-Arg or Arg-Arg), along with seven or more non-TRH peptides. Two prohormone convertases, PC1 and PC2, cleave the TRH tripeptides at the dibasic residues within the regulated secretory pathway. Carboxypeptidase E then removes the dibasic residues, leaving the sequence Gln-His-Pro-Gly. This peptide is then amidated at the C-terminus by peptidylglycine alpha-amidating monoxygenase, with Gly acting as the amide donor. The amino-terminal pyro-Glu residue results from cyclization of the Gin.

Although the TRH tripeptide is the only established hormone encoded within its large prohormone, the rat pro-TRH yields seven additional peptides that have unique tissue distributions.

Several biologic activities of these peptides have been observed: pre-pro-TRH (Fig. 7-13) may be a hypothalamic factor because it is released from hypothalamic slices and potentiates the thyrotropin-releasing effects of TRH.

TRH is a phylogenetically ancient peptide, being found in primitive vertebrates, such as the lamprey, and even invertebrates such as the snail. TRH is widely expressed in both CNS and periphery in amphibians, reptiles, and fishes but does not stimulate thyrotropin release in these poikilothermic vertebrates.

TRH has multiple peripheral and central activities and was co-opted as a hypophyseotropic factor midway during the evolution of vertebrates, perhaps specifically as a factor needed for coordinate regulation of temperature homeostasis.

Effects on the Pituitary Gland and Mechanism of Action

After intravenous injection of TRH in humans, serum thyrotropin levels rise within a few minutes (Fig. 7-14) followed by a rise in serum triiodothyronine (T₃) level. There is an increase in thyroxine (T₄) release as well, but a change in blood levels of T₄ is usually not demonstrable because the pool of circulating T₄ (most of which is bound to carrier proteins) is so large. The clinical applications of TRH testing are covered later in this chapter and in Chapter 10. TRH action on the pituitary is blocked by previous treatment with thyroid hormone, which is a crucial element in feedback control of pituitary thyrotropin secretion.

TRH is also a potent PRF (Fig. 7-14). The time course of response of blood PRL levels to TRH, the dose-response characteristics, and the suppression by thyroid hormone pretreatment (all of which parallel changes in thyrotropin secretion) suggest that TRH may be involved in the regulation of PRL secretion. Moreover, TRH is present in the hypophyseal-portal blood of lactating rats. However, it is unlikely to be a physiologic regulator of PRL secretion because the PRL response to nursing in humans is unaccompanied by changes in plasma thyrotropin levels. Nevertheless, TRH may occasionally cause hyperprolactinemia (with or without galactorrhea) in patients with hypothyroidism.

In normal individuals TRH has no influence on the secretion of pituitary hormones other than thyrotropin and PRL, but it enhances the release of human growth hormone (hGH) in acromegaly and of corticotropin in some patients with Cushing’s disease. Furthermore, prolonged stimulation of the pituitary by TRH also causes the release of hGH in some patients with uremia, hepatic disease, anorexia nervosa, and psychotic depression and in children with hypothyroidism. TRH inhibits sleep-induced hGH release through its actions in the CNS (see later in the section on extrapituitary actions of TRH).

Stimulatory effects of TRH are initiated by binding of the peptide to specific receptors on the plasma membrane of the thyrotroph. Neither thyroid hormone nor somatostatin, both of which antagonize the effects of TRH, interfere with its binding. TRH was originally thought to activate membrane adenylate cyclase to stimulate formation of cAMP and cAMP in turn was thought to stimulate thyrotropin secretion. However, cAMP does not increase under all conditions of TRH-induced thyrotropin release, and it is now clear that TRH action is mediated mainly through hydrolysis of phospholipidinositol, with phosphorylation of key protein kinases and an increase in intracellular free Ca²⁺ as the crucial step in postreceptor activation.

TRH is degraded to acid TRH and to the dipeptide histidylprolineamide, which cyclizes nonenzymatically to histidylproline diketopiperazine (cyclic His-Pro).

Figure 7-13 Structure of human thyrotropin-releasing hormone (TRH) gene and peptide, showing six repeating codons for the TRH sequence. CPE, carboxypeptidase E; PAM, peptidylglycine alpha-amidating monoxygenase; PC1, prohormone convertase 1. (From Yamada M, Radovick S, Wondisford FE, et al. Cloning and structure of human genomic DNA and hypothalamic cDNA encoding human preprothyrotropin-releasing hormone. Mol Endocrinol 1990; 4:551-556.)

Figure 7-14 Effect of intravenous injection of thyrotropin-releasing hormone on serum thyrotropin levels in humans. TRF, thyrotropin-releasing hormone; TSH, thyrotropin. (Replotted from data of Bowers C, Friesen HG, Hwang P, et al. Prolactin and thyrotropin release in man by synthetic pyroglutamylhistidyl-prolinamide. Biochem Biophys Res Commun 1971; 45:1033-1041.)
TRH has some behavioral effects in rats that are similar to those of TRH but no other proven actions. Cyclic His-Pro is reported to act as a PRF and to have other neural effects, including reversal of ethanol-induced sleep (TRH is also effective in this system), elevation of brain cyclic guanosine monophosphate levels, an increase in stereotyped behavior, modification of body temperature, and inhibition of eating behavior. Some of the effects of TRH may be mediated through cyclic His-Pro, but the fact that cyclic His-Pro is abundant in some areas and is not proportional to the amount of TRH suggests that the peptide may not be derived solely from TRH.

Extrahypothalamic Function

TRH is present in virtually all parts of the brain: cerebral cortex, circumventricular structures, neurohypophysis, pineal gland, and spinal cord. TRH is also found in pancreatic islet cells and in the gastrointestinal tract. Although it is present in low concentration, the total amount in extrahypothalamic tissues exceeds the amount in the hypothalamus.

The extensive extrahypothalamic distribution of TRH, its localization in nerve endings, and the presence of TRH receptors in brain tissue suggest that TRH serves as a neurotransmitter or neuromodulator outside the hypothalamus. TRH is a general stimulant of gonadal, pineal, neuroendocrine, and intestinal functions and induces hyperthermia on intracerebroventricular injection, suggesting a role in central thermoregulation.

Clinical Applications

The use of TRH for the diagnosis of hyperthyroidism is less common since the development of ultrasensitive assays for thyroid-stimulating hormone (TSH). see Chapter 10) Its use to discriminate between hypothalamic and pituitary causes of thyrotopin deficiency has also declined because of the test's poor specificity, but the application of ultrasensitive assays in conjunction with the TRH test has not been fully evaluated. TRH testing is also not of value in the differential diagnosis of causes of hyperprolactinemia but is useful for the demonstration of residual abnormal somatotropin-secreting cells in acromegalic patients who release hGH in response to TRH before treatment.

Studies of the effect of TRH on depression have shown inconsistent results, possibly because of poor blood-brain barrier penetration. Intrathecal administration of TRH may improve responses in depressed patients, but its clinical utility is unknown. Although a role for TRH in depression is not established, many depressed patients have a blunted thyrotropin response to TRH and changes in TRH responsiveness correlate with the clinical course. The mechanism by which the changes occur is unknown.

TRH has been proposed as a treatment for women with threatened premature labor to stimulate the production of lung surfactant in the preterm fetus. Despite encouraging results in early studies, several large-scale trials failed to show improvement in the survival of babies so treated.

TRH has been evaluated for the treatment of spinal muscle atrophy and amyotrophic lateral sclerosis; transient improvement in strength was reported in both disorders, but the combined experience at many centers using a variety of treatment protocols including long-term intrathecal administration failed to confirm efficacy. TRH administration also reduces the severity of experimentally induced spinal and ischemic shock. TRH has been used to treat children with neurologic disorders including West's syndrome, Lennox-Gastaut syndrome, early infantile epileptic encephalopathy, and intractable epilepsy. TRH has been proposed to be an analeptic agent. Sleeping or drug-sedated animals were awakened by the administration of TRH but its clinical utility is unknown. Although a role for TRH in depression is not established, many depressed patients have a blunted thyrotropin response to TRH and changes in TRH responsiveness correlate with the clinical course. The mechanism by which the changes occur is unknown.

TRH has been proposed as a treatment for women with threatened premature labor to stimulate the production of lung surfactant in the preterm fetus. Despite encouraging results in early studies, several large-scale trials failed to show improvement in the survival of babies so treated.

TRH has been evaluated for the treatment of paraparesis due to amyotrophic lateral sclerosis; transient improvement was reported in both disorders. Although these results were not confirmed by the combined experience at many centers using a variety of treatment protocols including long-term intrathecal administration, preliminary studies in humans suggest that TRH treatment may improve recovery after spinal cord injury and head trauma. TRH has been used to treat children with neurologic disorders including West's syndrome, Lennox-Gastaut syndrome, early infantile epileptic encephalopathy, and intractable epilepsy. TRH has been proposed to be an analeptic agent. Sleeping or drug-sedated animals were awakened by the administration of TRH but its clinical utility is unknown. Although a role for TRH in depression is not established, many depressed patients have a blunted thyrotropin response to TRH and changes in TRH responsiveness correlate with the clinical course. The mechanism by which the changes occur is unknown.

Regulation of Thyrotropin Release

The secretion of thyrotropin is regulated by two interacting elements: negative feedback by thyroid hormone and open-loop neural control by hypothalamic hypophyseotrophic factors.

![Figure 7-16 Regulation of the hypothalamic-pituitary-thyroid axis. AGRP, agouti-related protein; CART, cocaine and amphetamineregulated transcript; CRH, corticotropin-releasing hormone; NPY, neuropeptide Y; POMC, proopiomelanocortin; TSH, thyrotropin; T3, triiodothyronine; T4, thyroxine; TRH, thyrotropin-releasing hormone; TSH, thyrotropin; DB-R, leptin receptor.](image)

Feedback Control: Pituitary-Thyroid Axis

In the context of a feedback system, the level of thyroid hormone in blood or of its unbound fraction is the controlled variable and the set-point is the normal resting level of plasma thyroid hormone. Secretion of thyrotropin is inversely regulated by the level of thyroid hormone so that deviations from the set-point of control lead to appropriate changes in the rate of thyrotropin secretion. Factors that determine the rate of thyrotropin secretion required to maintain a given level of thyroid hormone include the rate at which thyrotropin and thyroid hormone disappear from the blood (turnover rate) and the rate at which T4 is converted to its more active form, T3.

![Figure 7-17 Relationship between plasma thyrotropin levels and thyroid hormone as determined by plasma protein-bound iodine (PBI) measurements in humans and rats. These curves illustrate, in the human (A) and the rat (B), that plasma thyrotropin levels are a curvilinear function of plasma thyroid hormone level. Human studies were carried out by giving myxodematous patients successive increments of thyrotropin T3 at approximately 10-day intervals. Each point represents simultaneous measurements of plasma PBI and plasma thyrotropin at various times in the six patients studied. The rat studies were performed by treating thyrotoxicated animals with various doses of T3 for 2 weeks before assay of plasma thyrotropin and plasma PBI. These curves illustrate that the secretion of thyrotropin is regulated over the entire range of thyroid hormone levels. At the normal set point for T3, the small changes above and below the control level are followed by appropriate increases or decreases in plasma thyrotopin. TSH, thyroid; T3, thyroxine. (From Reichlin S, Utiger RD. Regulation of the pituitary thyroid axis in man: relationship of TSH concentration to concentration of free and total thyroid hormone in plasma. J Clin Endocrinol Metab 1967; 27:25135, copyright by The Endocrine Society B from Reichlin S, Martin JJB, Bosshard RL, et al: Measurement of TSH in plasma and pituitary of the rat by a radioimmunoassay utilizing bovine TSH: effect of thyroidectomy or thyroxine administration on plasma TSH levels. Endocrinology 1970; 87:10221031, copyright by The Endocrine Society.)](image)

Thyroid hormones act on both the pituitary and the hypothalamus. Feedback control of the pituitary by thyroid hormone is remarkably precise. Administration of small doses of T3 and T4 inhibited the thyrotropin response to TRH and barely detectable decreases in plasma thyroid hormone levels were sufficient to sensitize the pituitary to TRH. TRH stimulates thyrotropin secretion within a few minutes through its action on a membrane receptor, whereas thyroid hormone actions,
mediated by intranuclear receptors, require several hours to take effect (see Chapter 10).

The secretion of hypothalamic TRH is also regulated by thyroid hormone feedback. Systemic injections of T₃ or implants of tiny T₃ pellets in the paraventricular nucleus of hypothyroid rats (Fig. 7-18) (Figure Not Available) reduced the concentration of TRH mRNA and TRH prohormone in TRH-secreting cells. Thyroid hormone also suppressed TRH secretion into hypophyseal-portal blood in sheep.

T₃ in the blood gains access to TRH-secreting neurons in the hypothalamus by way of the CSF. The hormone is taken up by epithelial cells of the choroid plexus of the lateral ventricle of the brain, bound within the cell to locally produced transthyretin (TTR, binding prealbumin), and then secreted across the blood-brain barrier. Within the brain, T₃ is converted by type II deiodinase, and T₃ interacts with subtypes of the thyroid hormone receptor, TR₁, TR₂, and TR₃ in the paraventricular nucleus and other brain cells (see Chapter 10). Hence the set-point of the pituitary-thyroid axis is determined by thyroid hormone levels within the brain. T₃ in the circulation is not transported into brain in this manner but presumably gains access to the paraventricular TRH neurons across the blood-brain barrier. The brain T₃ transport and deiodinase system account for the fact that the higher blood levels of T₃ are required to suppress pituitary-thyroid function after administration of T₃ than after administration of T₄.

Transthyretin is present in the brain of early reptiles and in addition is synthesized by the liver in warm-blooded animals. During embryogenesis in mammals, transthyretin is first detected when the blood-brain barrier appears, ensuring thyroid hormone access to the developing nervous system.

The hypothalamus determines the set-point of feedback control around which the usual feedback regulatory responses are elicited. Lesions of the thyrrotropic area lower basal thyroid hormone levels and make the pituitary more sensitive to inhibition by thyroid hormone, and high doses of TRH raise thyrotrophin and thyroid hormone levels. Synthesis of TRH in the paraventricular nucleus is regulated by feedback actions of thyroid hormones. The hypothalamus can override normal feedback control through an open-loop mechanism involving neuronal inputs to the hypophyseotropic TRH neurons (see Fig. 7-16). For example, cold exposure caused a sharp increase in thyrotropin release in animals and in human newborns. Circadian changes in thyrotropin secretion are another example of brain-directed changes in the set-point of feedback control, but if thyroid hormone levels are sufficiently elevated, as in hyperthyroidism, TRH cannot overcome the inhibition.

Hypothalamic regulation of thyrotropin secretion is also influenced by two inhibitory factors, somatostatin and dopamine. Antisomatostatin antibodies increase basal thyroid levels and potentiate the response to stimuli that normally induce thyrotropin release in the rat, as cold exposure and TRH administration. Thyroid hormone in turn inhibits the release of somatostatin, implying coordinated, reciprocal regulation of TRH and somatostatin by thyroid hormone. GH stimulates hypothalamic somatostatin synthesis and can inhibit thyrotropin secretion. The role of somatostatin in the regulation of thyrotropin secretion in humans is uncertain.

Dopamine has modest effects on thyrotropin secretion, and blockade of dopamine receptors (in the human) stimulates thyrotropin secretion slightly. Changes in the metabolism of thyroid hormone also influence T₃ homeostasis within the brain. In states of thyroid hormone deficiency, brain T₃ levels are maintained by an increase in the deiodinase that converts T₄ to T₃, and in both healthy and in human newborns. The pineal gland has been reported to inhibit thyroid function in some but not all studies. The pineal gland contains TRH, and in the frog its content changes with the season and with light and dark cycles independently of hypothalamic thyrotropin.

Circadian Rhythm

 Plasma thyrotropin in humans is characterized by a circadian periodicity, with a maximum between 9 pm and 5 am and a minimum between 4 pm and 7 pm. Smaller ultradian thyrotropin peaks occur every 90 to 150 minutes, probably because of bursts of TRH release from the hypothalamus, and are physiologically important in controlling the synthesis and glycosylation of thyrotropin. Glycosylation is a determinant of thyrotropin potency.

Temperature

External cold exposure activates and high ambient temperature inhibits the pituitary-thyroid axis in animals, and analogous changes occur in humans under certain conditions. Exposure of infants to cold at the time of delivery causes an increase in blood thyroid levels, possibly because of alterations in the turnover and degradation of the thyroid hormones. Blood thyroid hormone levels are higher in the winter than in the summer in individuals in cold climates but not in other climates. However, it is difficult to show that changes in environmental or body temperature in adults influence thyrotropin secretion. For example, exposure to cold ambient temperature or central hypothalamic cooling does not modify thyrotropin levels in young men. Behavioral changes, activation of the sympathetic nervous system, and shivering appear to be more important in temperature regulation in adults than the thyroid response.

The autonomic nervous system and the thyroid axis work together to maintain temperature homeostasis in mammals, and TRH plays a role in both pathways. Hypothalamic TRH release is rapidly (30 to 45 minutes) increased in rats exposed to cold. Rapid inhibition of somatostatin release in the median eminence has also been demonstrated, and both changes appear to play important roles in the rise in plasma TSH induced by cold exposure. TRH mRNA is elevated within an hour of cold exposure (see Fig. 7-18 C and D) (Figure Not Available). The regulation of hypothalamic TRH release and expression by cold is largely mediated by catecholamines. Noradrenergic and adrenergic fibers, originating in the brain stem, are found in close proximity to TRH nerve endings in the median eminence, and a rapid rise in TRH release was seen after norepinephrine treatment of hypothalamic fragments containing mainly median eminence. Brain stem adrenergic and noradrenergic fibers also make synaptic contacts with TRH neurons in the PVH (see Fig. 7-16). and thus catecholamines are likely to be involved in the regulation of TRH gene expression by cold. TRH neurons in the PVH are densely innervated by NPY terminals, and a portion of the NPY terminals arising from the C1, C2, C3, and A1 cell groups of the brain stem and projecting to the PVH are shown to be catecholaminergic. Somatostatin, dopamine, and serotonin also play a variety of roles in the regulation of TRH.

Stress

Stress is another determinant of thyrotropin secretion. In humans physical stress inhibits thyrotropin release, as indicated by the finding that in the euthyroid sick syndrome low T₃ and T₄ do not cause compensatory increases in thyrotropin secretion as would occur in normal individuals. A number of observations demonstrate interactions between the thyroid and adrenal axes. Physiologically, the bulk of evidence suggests that glucocorticoids in humans act to blunt the thyroid axis through actions in the CNS. Some actions may be direct because the TRH gene (see Fig. 7-13) (Figure Not Available) contains the glucocorticoid response element consensus sequence and hypophyseotropic TRH neurons appear to contain glucocorticoid receptors.
TSH, but disruption of cortisol synthesis with metyrapone only modestly affects the TSH circadian rhythm.

Several lines of evidence, however, identify conditions in which elevated glucocorticoids are associated with stimulation of the thyroid axis. Human depression is often associated with hypercortisolism and hyperthyroxinemia, and TRH mRNA levels are elevated by glucocorticoids in a number of cell lines as well as in cultured fetal hypothalamic TRH neurons from the rat. Thus, although glucocorticoids probably stimulate TRH production in TRH neurons, their overall inhibitory effect on the thyroid axis results from indirect glucocorticoid negative feedback on structures such as the hippocampus. Disruption of hippocampal suppression of the hypothalamic-pituitary-adrenal (HPA) axis is proposed to be involved in the hypercortisolism commonly seen in affective illness, and disruption of hippocampal inputs to the hypothalamic have been shown to produce a rise in hypophysiotropic TRH in the rat.

**Starvation**

The thyroid axis is depressed during starvation, presumably to help conserve energy by depressing metabolism (see Fig. 7-18 E to G) (Figure Not Available). In humans, reduced T₃, T₄, and TSH are seen during starvation or fasting. There are also changes in the thyroid axis in anorexia nervosa, such as low blood levels of T₃ and low normal levels of T₄ (see Chapter 33). Inappropriately low levels of TSH are found, suggesting defective activation of TRH production by low thyroid hormone levels. During starvation in rodents, reduced TRH release into hypophysial portal blood and reduced pro-TRH mRNA levels are seen, despite lowered thyroid hormone levels. Reduced basal TSH levels are also usually present.

The hypothyroidism seen in fasting or in the leptin-deficient Lep⁻/⁻ Lep⁻/⁻ mouse can be reversed by administration of leptin, and the evidence suggests that the mechanism involves leptin's ability to up-regulate TRH gene expression in the PVH (see Fig. 7-18 E to G) (Figure Not Available). Leptin appears to act both directly through leptin receptors on hypophysiotropic TRH neurons and indirectly through its actions on other hypothalamic cell groups, such as arcuate nucleus POMC and NPY-agouti-related peptide (AgRP) neurons. TRH neurons in the PVH receive dense NPY-AgRP and POMC projections from the arcuate and express NPY and melanocortin-4 receptors. The regulation of TRH by metabolic state is likely to be under redundant control.

**Infection and Inflammation**

The molecular basis of infection- or inflammation-induced thyrotropin suppression is now established. Sterile abscesses or the injection of interleukin-1 (IL-1; endogenous pyrogen, a secretory peptide of activated lymphocytes) or of tumor necrosis factor (TNF-) inhibits thyrotropin secretion. TNF- inhibits thyrotropin secretion directly and induces functional changes in the rat characteristic of the "sick euthyroid" state. It is likely that the thyrotropin inhibition in animal models of the sick euthyroid syndrome is due to cytokine-induced changes in hypothalamic and pituitary function.
Area Postrema

The area postrema lies at the caudal end of the fourth ventricle adjacent to the nucleus of the solitary tract (see Fig. 7-10 and Fig. 8-66). In experimental animals such as the rat and mouse, it is a midline structure lying above the nucleus of the solitary tract. However, in humans the area postrema is a bilateral structure. As the area postrema overlies the nucleus of the solitary tract, it also receives direct visceral afferent input from the glossopharyngeal nerve (including the carotid sinus nerve) and the vagus nerve. In addition, the area postrema receives direct input from several hypothalamic nuclei. The efferent projections of the area postrema include projections to the nucleus of the solitary tract, ventral lateral medulla, and the parabrachial nucleus. Consistent with a role as a sensory organ, the area postrema is enriched with receptors for several peptide hormones including glucagon-like peptide-1 and cholecystokinin. It also contains chemosensory neurons that include osmoreceptors. Notably, the area postrema is thought to be critical in the detection of potential toxins and can induce vomiting in response to foreign substances. In fact, the area postrema is often referred to as the chemoreceptor trigger zone.

The best-described physiologic role of the area postrema is probably the coordinated control of blood pressure. For example, the area postrema contains binding sites for angiotensin II, AVP, and atrial natriuretic peptide. Moreover, lesions of the area postrema in rats blunt the rise in blood pressure induced by angiotensin II. Finally, administration of angiotensin II induces the expression of Fos in neurons of the area postrema. The area postrema has also been hypothesized to play a role in responding to inflammatory cytokines during the acute febrile response (see Fig. 7-50 and Fig. 7-51 and "Neuroendocrine-Immune Interactions").
Subcommissural Organ

The subcommissural organ (SCO) is located below the posterior commissure near the junction of the third ventricle and cerebral aqueduct below the pineal gland (see Fig. 7-10). The SCO is composed of specialized ependymal cells that secrete a highly glycosylated protein of unknown function. The secretion of this protein leads to aggregation and formation of the so-called Reissner fibers. The glycoproteins are extruded through the aqueduct, the fourth ventricle, and the spinal cord lumen to terminate in the caudal spinal canal. In humans, intracellular secretory granules are identifiable in the SCO but Reissner’s fibers are absent. The SCO secretion in humans is therefore presumed to be more soluble and to be absorbed directly from the CSF. Compared with other CVOs, relatively little is known about the physiologic role of the SCO. Hypothesized roles for the SCO include clearance of substances from the CSF.
PINEAL GLAND

Historically, the functional significance of the pineal gland has been difficult to discern. For example, Descartes called the pineal gland the "seat of the soul." The pineal gland is both an endocrine and a circumventricular organ; it is derived from cells of the roof of the third ventricle and lies above the posterior commissure near the level of the habenular complex and the sylvian aqueduct. However, neuroanatomic studies have established that light-encoded information is relayed to the pineal in an indirect and multisynaptic fashion. This series of synapses ultimately results in innervation of the gland by noradrenergic sympathetic nerve terminals that are critical regulators of melatonin production and release. Specifically, the retina provides direct innervation to the suprachiasmatic nucleus (SCN) of the hypothalamus through the retinohypothalamic tract. The SCN in turn provides input to the paraventricular nucleus of the hypothalamus (PVH), a key cell group in neuroendocrine and autonomic control. This input is provided through direct and indirect pathways by intrahypothalamic projections. The PVH in turn provides direct innervation to sympathetic preganglionic neurons in the intermediolateral cell column of the thoracic regions of the spinal cord. Sympathetic preganglionic neurons innervate postganglionic neurons in the superior cervical ganglion, which in turn provide the noradrenergic innervation to the pineal (see "Hypothalamic-Pituitary Unit"). This rather circuitous pathway is thought to represent the anatomic substrate for light to regulate the secretion of melatonin. It is important to note that in the absence of light input, the pineal gland rhythms persist but are not entrained to the external light-dark cycle.

The Pineal Is the Source of Melatonin

The predominant hormone secreted by the pineal gland is melatonin. However, the pineal also contains biogenic amines, peptides, and GABA. Pineal-derived melatonin is synthesized from tryptophan, through serotonin, with the rate-limiting

<table>
<thead>
<tr>
<th>TABLE 7-2 -- Neuroactive Materials in the Paraventricular Nucleus and the Arcuate Nucleus</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Paraventricular Nucleus</strong></td>
</tr>
<tr>
<td><strong>Magnocellular Division</strong></td>
</tr>
<tr>
<td>Angiotensin II</td>
</tr>
<tr>
<td>Cholecystokinin</td>
</tr>
<tr>
<td>Glucagon</td>
</tr>
<tr>
<td>Oxytocin</td>
</tr>
<tr>
<td>Peptide 7B2</td>
</tr>
<tr>
<td>Proenkephalin B (dynorphin, rimorphin, -neoendorphin)</td>
</tr>
<tr>
<td>Vasopressin</td>
</tr>
<tr>
<td>Nitric oxide (NO)</td>
</tr>
<tr>
<td><strong>Parvicellular Division</strong></td>
</tr>
<tr>
<td>-Aminobutyric acid (GABA)</td>
</tr>
<tr>
<td>Angiotensin II</td>
</tr>
<tr>
<td>Atrial natriuretic factor</td>
</tr>
<tr>
<td>Cholecystokinin</td>
</tr>
<tr>
<td>Corticotropin-releasing hormone</td>
</tr>
<tr>
<td>Dopamine</td>
</tr>
<tr>
<td>Follicle-stimulating hormone-releasing factor</td>
</tr>
<tr>
<td>Galanin</td>
</tr>
<tr>
<td>Glucagon</td>
</tr>
<tr>
<td>Neuropeptide Y</td>
</tr>
<tr>
<td>Neotensin</td>
</tr>
<tr>
<td>Peptide 7B2</td>
</tr>
<tr>
<td>Proenkephalin A (met-enkephalin, leu-enkephalin, BAM 22P, metorphamide, met-enkephalin-Arg^1-Phe^7, met-enkephalin-Arg^7-Gly^7-Leu^8)</td>
</tr>
<tr>
<td>Somatostatin</td>
</tr>
<tr>
<td>Thyrotropin-releasing hormone (TRH)</td>
</tr>
<tr>
<td>Vasopressin</td>
</tr>
<tr>
<td>Interleukin-1 (IL-1)</td>
</tr>
<tr>
<td>Vasoactive intestinal peptide (VIP)/peptide-histidine-isoleucine (PHI)</td>
</tr>
<tr>
<td>Nitric oxide</td>
</tr>
<tr>
<td><strong>Arcuate Nucleus</strong></td>
</tr>
<tr>
<td>Acetylcholine (Ach)</td>
</tr>
<tr>
<td>-Aminobutyric acid</td>
</tr>
<tr>
<td>Dopamine</td>
</tr>
<tr>
<td>Galanin</td>
</tr>
<tr>
<td>Growth hormone-releasing hormone (GHRH)</td>
</tr>
<tr>
<td>Luteinizing hormone-releasing hormone (LHRH)</td>
</tr>
</tbody>
</table>
Neuropeptide Y
Neurotensin
Pancreatic polypeptide
Proenkephalin A
Prolactin
Melanocortins (corticotropin, -melanocyte-stimulating hormone [-MSH], -melanocyte-stimulating hormone [-MSH])
Endogenous opioids (-endorphin, -lipotropin [-LPH])
Somatostatin
Substance P


The final step of melatonin synthesis is catalyzed by hydroxyindole-O-methyltransferase (HIOMT). These enzymes are expressed in a pineal specific manner; however, hydroxyindole-O-methyltransferase is also expressed in the retina and red blood cells. It is now established that melatonin plays a key role in regulating a myriad of circadian rhythms, and a fundamental principle of circadian biology is that the synthesis of melatonin is exquisitely controlled. This control is exerted at several levels. NAT mRNA levels, NAT activity, and melatonin synthesis and release are regulated in a circadian fashion and are entrained by the light-dark cycle, with darkness thought to be the most important signal. For example, melatonin and NAT levels are highest during the dark and decrease sharply with the onset of light. Melatonin is not thought to be stored to any degree and thus is released into blood or CSF directly.

The CNS control of melatonin secretion during the dark is mediated by the neuroanatomic pathway already outlined. Lack of light ultimately results in the release of norepinephrine from postganglionic sympathetic nerve terminals that act on -adrenergic receptors in pinealocytes, resulting in an increase in adenylate cyclase activity. The resultant increased levels of cyclic adenosine monophosphate (cAMP) activate signal transduction cascades, including increased protein kinase activity and phosphorylation of cAMP response element binding protein. Notably, cAMP response elements have been identified in the promoter of NAT. Thus, light (or lack of it) acting through the sympathetic nervous system induces an increase in cAMP, representing a fundamental regulator of NAT transcription and melatonin synthesis that ultimately results in a dramatic change of melatonin levels across the day.
Physiologic Roles of Melatonin

One of the best characterized roles of melatonin is the regulation of the reproductive axis, including gonadotropin secretion and the timing and onset of puberty (see "Gonadotropin-Releasing Hormone and Control of the Reproductive Axis"). The potent regulation of the reproductive axis by melatonin is established in rodents and domestic animals such as the sheep. It was observed experimentally with the demonstration that removal of the pineal leads to precocious puberty and ameliorates the effects of constant darkness to induce gonadal involution. In addition, male rats exposed to constant darkness or made blind experimentally display testicular atrophy and decreased levels of testosterone. These profound effects are normalized by removal of the pineal gland. The physiologic significance of melatonin is probably most important in species referred to as seasonal breeders. Indeed, the role of melatonin in regulating reproductive capacity in species such as the sheep and the horse is now established. This type of reproductive strategy probably evolved to synchronize the length of day with the gestational period of the species to ensure that the offspring are born at favorable times of the year and maximize the viability of the young. Despite the potent effects of day length on reproduction in these species, exact mechanisms of melatonin regulation of GnRH release are unsettled. However, melatonin inhibits LH release from the rat pars tuberalis. The role of the pineal in human reproduction is even more unsettled. Earlier onset of menarche in blind women has been reported. In addition, a decline in melatonin at puberty has been described. However, it was not found in other studies. Thus, the role of melatonin in human reproduction is not clear. Nonetheless, the therapeutic potential of melatonin in regulating and shifting biologic rhythms in humans has received great attention.
Melatonin Receptors

It is now established that melatonin mediates its effects by acting on a family of G protein-coupled receptors, which have been characterized by pharmacologic, neuroanatomic, and molecular approaches. The first member of the family, Mel$_{1a}$, is a high-affinity receptor that was isolated originally from *Xenopus* melanophores. The second, Mel$_{1b}$, has approximately 60% homology with the Mel$_{1a}$ receptor. A third receptor, the Mel$_{1c}$ melatonin receptor, has been cloned from zebra fish, *Xenopus*, and chickens but not as yet from mammals.

The mechanisms for the effects of melatonin on regulating and entraining circadian rhythms are becoming increasingly understood. For example, melatonin inhibits the activity of neurons in the SCN of the hypothalamus, the master circadian pacemaker in the mammalian brain. Melatonin can entrain several mammalian circadian rhythms, probably by the inhibition of neurons in the SCN. Neuroanatomic evidence suggests that many of the effects of melatonin on circadian rhythms involve actions on Mel$_{1a}$ receptors, as the distribution of Mel$_{1a}$ mRNA overlaps with radiolabeled melatonin binding sites in the relevant brain regions. These sites include the SCN, the retina, and the pars tuberalis of the adenohypophysis. The Mel$_{1b}$ melatonin receptor is also expressed in retina and brain; however, this is thought to be at much lower levels.

Genetic studies in mice have also helped to illuminate the relative roles of each melatonin receptor in mediating the effects of this hormone. Targeted deletion (knockout) of the Mel$_{1a}$ receptor abolished the ability of melatonin to inhibit the activity of SCN neurons. Several studies have suggested that the inhibition of SCN neurons by melatonin is of great physiologic significance. For example, Reppert and colleagues have suggested that elevations of melatonin at night could decrease the responsiveness of the SCN to activity-related stimuli that could result in phase shifts. As noted, light potently inhibits melatonin synthesis and release. Thus, melatonin may underlie the mechanism by which light induces phase shifts. However, it should be noted that lack of the Mel$_{1a}$ gene does not block the ability of melatonin to induce phase shifts. These unexpected and somewhat confusing results have resulted in the hypothesis that Mel$_{1b}$ is involved in melatonin-induced phase shifts, as this receptor may be expressed in the human brain.
Melatonin Therapy in Humans

The role of melatonin as a "wonder drug" has received great attention from the lay press. The proposed beneficial and therapeutic uses of melatonin include treatment of jet lag, slowing or reversing the progression of aging, and enhancing immune function. As noted earlier, the most studied and established role of melatonin is that of phase shifting and resetting circadian rhythms. In this context, melatonin has been used to treat jet lag and may be effective in treating circadian-based sleep disorders. In addition, melatonin administration has been shown to regulate sleep in humans. Specifically, melatonin has a hypnotic effect at relatively low doses. Melatonin therapy has also been suggested as a way to treat seasonal affective disorders.

It is important to note that melatonin is now available over the counter and without a prescription throughout the United States. However, there is a striking paucity of controlled clinical studies of the relative efficacy and safety of melatonin administration. This should be viewed as problematic because melatonin is an endocrine hormone, and most hormones are not widely available without a prescription. Clearly, controlled clinical studies are needed to assess fully the therapeutic potential and safety of melatonin in humans.
HORMONAL FEEDBACK CONTROL SYSTEMS

HYPOTHALAMIC-REGULATED HORMONES AND NEUROENDOCRINE AXES

With the demonstration by the first half of the 1900s that pituitary secretion was controlled by hypothalamic hormones released into the portal circulation, the search was on for the hypothalamic releasing factors. The search for hypothalamic neurohormones with anterior pituitary regulating properties focused on extracts of stalk, median eminence, neural lobe, and hypothalamus from sheep and pigs. To give some idea of the herculean nature of this effort, approximately 250,000 hypothalamic fragments were required to purify and characterize the first such factor, TRH. The identification and characterization of TRH in 1970 and of other releasing hormones ultimately led to the Nobel Prize in Medicine in 1977 for Andrew Schally and Roger Guillemin. Such hypothysotropic substances were initially called releasing factors but are now more commonly called releasing hormones.

All of the hypothalamic-pituitary regulating hormones are peptides with the exception of dopamine, which is a biogenic amine that is the principal prolactin-inhibiting factor (PIF) (Table 7-3). All are now available for human investigation and treatment, and therapeutic analogues have been synthesized for dopamine, GnRH, and somatostatin.

In addition to regulating hormone release, some hypothysisotropic factors control pituitary cell differentiation and proliferation and hormone synthesis. Somatostatin and dopamine are inhibitory, and some act on more than one pituitary hormone. For example, TRH is a potent releaser of prolactin (PRL) and of thyrotropin and under some circumstances releases corticotropin and growth hormone (GH). GnRH releases both LH and follicle-stimulating hormone (FSH). Somatostatin inhibits the secretion of GH, thyrotropin, and a wide variety of nonpituitary hormones. The principal inhibitor of PRL secretion, dopamine, also inhibits secretion of thyrotropin, gonadotropin, and, under certain conditions, GH. Dual control is exerted by the interaction of inhibitory and stimulatory hypothalamic hormones. For example, somatostatin interacts with growth hormonereleasing hormone (GHRH) and TRH to control secretion of GH and thyrotropin, respectively, and dopamine interacts with prolactin-releasing factors (PRFs) to regulate PRL secretion. Some hypothalamic hormones act synergistically; for example, CRH and vasopressin act together to regulate the release of pituitary adrenocorticotropic hormone (ACTH).

Secretion of the releasing hormones in turn is regulated by

<table>
<thead>
<tr>
<th>Table 7-3 -- Structural Formulas of Principal Human Hypothalamic Peptides Directly Related to Pituitary Secretion</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vasopressin</strong></td>
</tr>
<tr>
<td>Cys-Tyr-Phe-Gln-Asn-Cys-Pro-Arg-Gly-NH$_2$ (MW = 1084.38)</td>
</tr>
<tr>
<td><strong>Oxytocin</strong></td>
</tr>
<tr>
<td>Cys-Tyr-Ile-Gln-Asn-Cys-Pro-Leu-Gly-NH$_2$ (MW = 1007.35)</td>
</tr>
<tr>
<td><strong>Thyrotropin-releasing hormone</strong></td>
</tr>
<tr>
<td>pGlu-His-Pro-NH$_2$ (MW = 362.42)</td>
</tr>
<tr>
<td><strong>Gonadotropin-releasing hormone</strong></td>
</tr>
<tr>
<td>pGlu-His-Tyr-Gly-Leu-Arg-Pro-Gly-NH$_2$ (MW = 1182.39)</td>
</tr>
<tr>
<td><strong>Corticotropin-releasing hormone</strong></td>
</tr>
<tr>
<td><strong>Growth hormone releasing hormone (GHRH 140, 144-NH$_2$, Human)</strong></td>
</tr>
<tr>
<td><strong>Somatostatin</strong></td>
</tr>
<tr>
<td>Ala-Gly-Cys-Lys-Asn-Phe-Phe-Trp-Lys-Thr-Phe-Thr-Ser-Cys (MW = 1638.12)</td>
</tr>
<tr>
<td><strong>Somatostatin-28</strong></td>
</tr>
<tr>
<td>Ser-Ala-Asn-Ser-Ala-Pro-Ala-Met-Ala-Pro-Arg-Glu-Lys-Ala-Gly-Cys-Lys-Asn-Phe-Phe-Trp-Lys-Thr-Phe-Thr-Ser-Cys (MW = 3149.0)</td>
</tr>
<tr>
<td><strong>Somatostatin-28 (112)</strong></td>
</tr>
<tr>
<td>Ser-Ala-Asn-Ser-Ala-Pro-Ala-Met-Ala-Pro-Arg-Glu (MW = 1244.49)</td>
</tr>
<tr>
<td><strong>Vasoactive intestinal peptide (pig, rat)</strong></td>
</tr>
<tr>
<td><strong>Prolactin-releasing peptide (PrRP31, PrRP20)</strong></td>
</tr>
<tr>
<td>Ser-Arg-Thr-His-Ser-Met-Glu-Ile-Ile-Arg-Thr-Pro-Asp-Ala-Ile-Pro-Asn-Pro-Ala-Trp-Tyr-Ala-Arg-Gly-Ile-Ile-Arg-Pro-Val-Gly-Phe-NH$_2$ (MW = 3665.16, 2273.58)</td>
</tr>
<tr>
<td><strong>Ghrelin</strong></td>
</tr>
<tr>
<td>Gly-Ser-Phe-Leu-Ser-Pro-Glu-His-Gln-Val-Gln-Gln-Gly-Ser-Lys-Ser-Lys-Pro-Pro-Ala-Leu-Gln-Glu-Pro-Arg (MW = 3514.9) [Ser 3 is n-octanoylated]</td>
</tr>
</tbody>
</table>

MW, molecular weight.

neurotransmitters and neuropeptides released by a complex array of neurons synapsing with hypothysotropic neurons. Control of secretion is also exerted through feedback control by hormones such as glucocorticoids, gonadal steroids, thyroid hormone, anterior pituitary hormones (short-loop feedback control), and hypothysisotropic factors themselves (ultrashort-loop feedback control).

The distribution of the hypothysotropic hormones is not limited to the hypothalamus. Most are also found in nonhypothysotropic hypothalamic neurons, in extrahypothalamic regions of the brain, and in other organs where they may have functions (e.g., effects on behavior or homeostasis) unrelated to pituitary regulation. Most, although not all, of the peptides, hormones, and neurotransmitters involved in the regulation of hypothalamic-pituitary control belong to the G protein-coupled receptor family (Table 7-4).

Feedback Concepts in Neuroendocrinology

In order to understand the regulation of each hypothalamic-pituitary-target organ axis, it is important to understand some basic concepts of homeostatic systems. A simplified account of feedback control in relation to neuroendocrine regulation is presented in this section. Hormonal systems form part of a feedback loop in which the controlled variable (generally the blood hormone level or some biochemical surrogate of the hormone) determines the rate of secretion of the hormone. In negative feedback systems the controlled variable inhibits hormone output, and in positive feedback control systems the controlled variable increases...
hormone secretion. Both negative and positive endocrine feedback control systems can be part of a closed loop, in which regulation is entirely restricted to the interacting regulatory glands, or an open loop, in which the nervous system influences the feedback loop. All pituitary feedback systems have nervous system inputs that either alter the set-point of the feedback control system or introduce open-loop elements that can influence or override the closed-loop control elements.

In engineering formulations of feedback, three controlled variables can be identified: a sensing element that detects the concentration of the controlled variable, a reference input that defines the proper control levels, and an error signal that determines the output of the system. The reference input is the set-point of the system.

Hormonal feedback control systems resemble engineering systems in that the concentration of the hormone in the blood (or some function of the hormone) regulates the output of the controlling gland. Hormonal feedback differs from engineering systems in that the sensor element and the reference input element are not readily distinguishable. The set-point of the controlled variable is determined by a complex cascade beginning with the kinetics of binding to a receptor and the activities of successive intermediate messengers. Sophisticated models incorporating control elements, compartmental analysis, and hormone production and clearance rates have been developed for many systems.
Virtually all functions of living animals (regardless of their position on the evolutionary scale) are subject to periodic or cyclic changes, many of which are influenced mainly by the nervous system (see Table 7-5 for definitions). Most periodic changes are free-running; that is, they are intrinsic to the organism independent of the environment and are driven by a biologic “clock.”

Most free-running rhythms can be coordinated (entrained) by external signals (cues), such as light-dark changes, meal patterns, cycles of the lunar periods, or the ratio of the length of day to the length of night. External signals of this type (zeitgeber or time givers) do not bring about the rhythm but provide

<table>
<thead>
<tr>
<th>Group</th>
<th>Name</th>
<th>General Structure</th>
<th>Transmembrane Sequences</th>
<th>Mode of Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biogenic amines</td>
<td>1A-Adrenoreceptors</td>
<td>1A</td>
<td>7</td>
<td>G9,11</td>
</tr>
<tr>
<td></td>
<td>1B</td>
<td>7</td>
<td>G9,11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1D</td>
<td>7</td>
<td>G9,11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2A-Adrenoreceptors</td>
<td>2A</td>
<td>7</td>
<td>G9,11</td>
</tr>
<tr>
<td></td>
<td>2B</td>
<td>7</td>
<td>G9,11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2C</td>
<td>7</td>
<td>G9,11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-Adrenoreceptors</td>
<td>1</td>
<td>7</td>
<td>G9</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>7</td>
<td>G9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>7</td>
<td>G9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Serotonin (5-OH-tryptamine)</td>
<td>5-HT1A</td>
<td>7</td>
<td>G9,11</td>
</tr>
<tr>
<td></td>
<td>5-HT1B</td>
<td>7</td>
<td>G9,11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5-HT1D</td>
<td>7</td>
<td>G9,11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5-HT1E</td>
<td>7</td>
<td>G9,11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5-HT2A</td>
<td>7</td>
<td>G9,11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5-HT2B</td>
<td>7</td>
<td>G9,11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5-HT2C</td>
<td>7</td>
<td>G9,11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5-HT3</td>
<td>4</td>
<td>Cation channel</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5-HT4</td>
<td>7</td>
<td>G9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dopamine</td>
<td>D1</td>
<td>7</td>
<td>G9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D2</td>
<td>7</td>
<td>G9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D3</td>
<td>7</td>
<td>G9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D4</td>
<td>7</td>
<td>G9</td>
</tr>
<tr>
<td></td>
<td>Histamine</td>
<td>H1</td>
<td>7</td>
<td>G9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H2</td>
<td>7</td>
<td>G9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H3</td>
<td>7</td>
<td>G9</td>
</tr>
<tr>
<td></td>
<td>Acetylcholine</td>
<td>M1</td>
<td>7</td>
<td>G9,11</td>
</tr>
<tr>
<td></td>
<td>Muscarinic</td>
<td>M2</td>
<td>7</td>
<td>G9,11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M3</td>
<td>7</td>
<td>G9,11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M4</td>
<td>7</td>
<td>G9</td>
</tr>
<tr>
<td></td>
<td>N9</td>
<td>7</td>
<td>G9,11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nicotinic</td>
<td>Muscle</td>
<td>4 Multiunit</td>
<td>Na/K/Ca</td>
</tr>
<tr>
<td></td>
<td>Ganglionic</td>
<td>4 Multiunit</td>
<td>Na/K/Ca</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Central nervous system</td>
<td>4 Multiunit</td>
<td>Na/K/Ca</td>
<td></td>
</tr>
<tr>
<td>Excitatory amino acids (glutamate)</td>
<td>Ionotrophic</td>
<td>NMDA</td>
<td>Multiunit 7TM</td>
<td>Na/K/Ca</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AMPA</td>
<td>Multiunit 3TM</td>
<td>Na/K/Ca</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kainate</td>
<td>7</td>
<td>Na/K/Ca</td>
</tr>
<tr>
<td></td>
<td>Metabotropic</td>
<td>mGlu1</td>
<td>7</td>
<td>G9,11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mGlu2</td>
<td>7</td>
<td>G9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mGlu3</td>
<td>7</td>
<td>G9,11</td>
</tr>
<tr>
<td>Neuropeptides</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------------------------------</td>
<td>--</td>
<td>--</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neuropeptides</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neurohypophyseal hormones</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vasopressin</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V₁A</td>
<td>7</td>
<td>G₁₁₁</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V₁B</td>
<td>7</td>
<td>G₁₁₁</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V₂</td>
<td>7</td>
<td>G₉</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oxytocin</td>
<td>OT</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypophyseotropin hormones</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRH</td>
<td>TRH</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GHRH</td>
<td>GHRH</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LHRH</td>
<td>LHRH</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRH</td>
<td>CRH</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Somatostatin</td>
<td>SST₁</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SST₂</td>
<td>7</td>
<td>G₉</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SST₃</td>
<td>7</td>
<td>G₉</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SST₄</td>
<td>7</td>
<td>G₉</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SST₅</td>
<td>7</td>
<td>G₉</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Endogenous opioid peptides</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Melanocortins</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MC₁</td>
<td>MC₁</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MC₂ (corticotropin)</td>
<td>7</td>
<td>G₉</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MC₃</td>
<td>7</td>
<td>G₉</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MC₄</td>
<td>7</td>
<td>G₉</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MC₅</td>
<td>7</td>
<td>G₉</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gut-brain peptides</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tachykinins</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Substance P</td>
<td>NK₁</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Substance K</td>
<td>NK₂</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neurokinin B</td>
<td>NK₃</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neurotensin</td>
<td>NT</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VIP</td>
<td>VIP₁</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VIP₂</td>
<td>7</td>
<td>G₉</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PACAP</td>
<td>PACAP</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Galanin</td>
<td>G</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cholecystokinin</td>
<td>CCK₄</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cholecystokinin (gastrin receptor)</td>
<td>CCK₅</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neuropeptide Y</td>
<td>Y₁</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y₂</td>
<td>7</td>
<td>G₉</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vasoactive peptides</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y₅</td>
<td>7</td>
<td>G₉</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Angiotensin</td>
<td>AT₁</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT₂</td>
<td>7</td>
<td>cGMP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atrial natriuretic peptide</td>
<td>ANP₄</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANP₅</td>
<td>1</td>
<td>cGMP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Endothelin</td>
<td>ET₁</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ET₂</td>
<td>7</td>
<td>G₁₁₁</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| ANP, AMPA, -amino-3-hydroxy-5-methyl-4-isoxazoloproprionic acid; TRH, thyrotropin-releasing hormone; GHRH, growth hormone-releasing hormone; LHRH, luteinizing hormone-releasing hormone; CRH, corticotropin-releasing hormone; NT, neurotensin; VIP, vasoactive peptide; PACAP, pituitary adenylate cyclase activating peptide; cGMP, cyclic guanosine monophosphate. cGMP: Guanylate cyclase activity intrinsic to the receptor.
Some receptors have intrinsic tyrosine phosphorylase activity, others have intrinsic tyrosine hydroxylase activity. The former stimulate phosphorylation of tyrosine kinases; the latter stimulate breakdown of tyrosine kinase.

The designation of functional type is oversimplified. Many examples can be cited in which receptor activation can stimulate both adenylate cyclase and phosphoinositide turnover.


Receptors cited are human or rat if human not available.

---

**TABLE 7-5 — Terms Used to Describe Cyclic Endocrine Phenomena**

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Period</td>
<td>Length of the cycle</td>
</tr>
<tr>
<td>Circadian</td>
<td>Around a day</td>
</tr>
<tr>
<td>Diurnal</td>
<td>Exactly a day</td>
</tr>
<tr>
<td>Ultradian</td>
<td>Less than a day, i.e., minutes or hours</td>
</tr>
<tr>
<td>Infradian</td>
<td>Longer than a day, i.e., month or year</td>
</tr>
<tr>
<td>Mean</td>
<td>Arithmetic mean of all values within a cycle</td>
</tr>
<tr>
<td>Range</td>
<td>Difference between the highest and lowest values</td>
</tr>
<tr>
<td>Nadir</td>
<td>Minimal level (inferred from mathematical curve fitting calculations)</td>
</tr>
<tr>
<td>Acrophase</td>
<td>Time of maximal levels (inferred from curve fitting)</td>
</tr>
<tr>
<td>Zeitgeber</td>
<td>&quot;Time-giver&quot; (German), the external cue, usually the light-dark cycle that synchronizes endogenous rhythms</td>
</tr>
<tr>
<td>Entainment</td>
<td>The process by which an endogenous rhythm is regulated by a zeitgeber</td>
</tr>
<tr>
<td>Phase shift</td>
<td>Induced change in an endogenous rhythm</td>
</tr>
<tr>
<td>Intrinsic clock</td>
<td>Neural structures that possess intrinsic capacity for spontaneous rhythms; for circadian rhythms these are located in the suprachiasmatic nucleus</td>
</tr>
</tbody>
</table>

Thyrotropin-Releasing Hormone

Chemistry and Evolution

TRH, the smallest known peptide hormone, is the tripeptide pyroGlui-His-Pro-NH₂. The TRH peptide sequence is repeated six times within the human TRH pre-prohormone gene (Fig. 7-13) (Figure Not Available). The rat pro-TRH precursor contains five TRH peptide repeats flanked by dibasic residues (Lys-Arg or Arg-Ang), along with seven or more non-TRH peptides. Two prohormone convertases, PC1 and PC2, cleave the TRH tripeptides at the dibasic residues within the regulated secretory pathway. Carboxypeptidase E then removes the dibasic residues, leaving the sequence Gin-His-Pro-Gly. This peptide is then amidated at the C-terminus by peptidylglycine alpha-amidating monoxygenase, with Gly acting as the amide donor. The amino-terminal pyro-Glu residue results from cyclization of the Gin.

Although the TRH tripeptide is the only established hormone encoded within its large prohormone, the rat pro-TRH yields seven additional peptides that have unique tissue distributions. Several biologic activities of these peptides have been observed: pre-pro-TRH may be a hypophysiotropic factor because it is released from hypothalamic slices and potentiates the thyrotropin-releasing effects of TRH. Pro-TRH is also released from the median eminence and appears to inhibit ACTH release.

TRH is a phylogenetically ancient peptide, being found in primitive vertebrates, such as the lamprey, and even invertebrates such as the snail. TRH is widely distributed in the hypothalamic cDNA encoding human preprothyrotropin-releasing hormone. Mol Endocrinol 1990; 4:551-556.

TRH is a phylogenetically ancient peptide, being found in primitive vertebrates, such as the lamprey, and even invertebrates such as the snail. TRH is widely expressed in both CNS and periphery in amphibians, reptiles, and fishes but does not stimulate thyrotropin release in these poikilothermic vertebrates. Thus, TRH has multiple peripheral and central activities and was co-opted as a hypophysiotropic factor midway during the evolution of vertebrates, perhaps specifically as a factor needed for coordinate regulation of temperature homeostasis.

Effects on the Pituitary Gland and Mechanism of Action

After intravenous injection of TRH in humans, serum thyrotropin levels rise within a few minutes (Fig. 7-14) followed by a rise in serum triiodothyronine (T₃) levels. There is an increase in thyroxine (T₄) release as well, but a change in blood levels of T₄ is usually not demonstrable because the pool of circulating T₄ (most of which is bound to carrier proteins) is so large. The clinical applications of TRH testing are covered later in this chapter and in Chapter 10. TRH action on the pituitary is blocked by previous treatment with thyroid hormone, which is a crucial element in feedback control of pituitary thyrotropin secretion.

TRH is also a potent PRF (Fig. 7-15). The time course of response of blood PRL levels to TRH, the dose-response characteristics, and the suppression by thyroid hormone pretreatment (all of which parallel changes in thyrotropin secretion) suggest that TRH may be involved in the regulation of PRL secretion. Moreover, TRH is present in the hypophysial-portal blood of lactating rats. However, it is unlikely to be a physiologic regulator of PRL secretion because the PRL response to nursing in humans is unaccompanied by changes in plasma thyrotropin levels. Nevertheless, TRH may occasionally cause hyperprolactinemia (with or without galactorrhea) in patients with hypothyroidism.

In normal individuals TRH has no influence on the secretions of pituitary hormones other than thyrotropin and PRL, but it enhances the release of human growth hormone (hGH) in acromegaly and of corticotropin in some patients with Cushing’s disease. Furthermore, prolonged stimulation of the normal pituitary with GHRH can sensitize it to the hGH-releasing effects of TRH. TRH also causes the release of hGH in some patients with uremia, hepatic disease, anorexia nervosa, and psychotic depression and in children with hypothyroidism. TRH inhibits sleep-induced hGH release through its actions in the CNS (see later in the section on extrapituitary actions of TRH).

Stimulatory effects of TRH are initiated by binding of the peptide to specific receptors on the plasma membrane of the thyrotroph. Neither thyroid hormone nor somatostatin, both of which antagonize the effects of TRH, interfere with its binding. TRH was originally thought to activate membrane adenylate cyclase to stimulate formation of cAMP and cAMP in turn was thought to stimulate thyrotropin secretion. However, CAMP does not increase under all conditions of TRH-induced thyrotropin release, and it is now clear that TRH action is mediated mainly through hydrolysis of phosphatidylinositol, with phosphorylation of key protein kinases and an increase in intracellular free Ca²⁺ as the crucial step in postreceptor activation. TRH effects can be mimicked by exposure to a Ca²⁺ ionophore and are partially abolished by a Ca²⁺-free medium. TRH stimulates the formation of mRNAs coding for thyrotropin and PRL, confirming that this peptide is trophic as well as a releasing factor.

TRH is degraded to acid TRH and to the dipeptide histidylprolineamide, which cyclizes nonenzymatically to histidylproline diketopiperazine (cyclic His-Pro). Acid

Figure 7-13 Structure of human thyrotropin-releasing hormone (TRH) gene and peptide, showing six repeating codons for the TRH sequence. CPE, carboxypeptidase E; PAM, peptidylglycine alpha-amidating monoxygenase; PC1, prohormone convertase 1. (From Yamada M, Radvick S, Wendtson FE, et al. Cloning and structure of human genomic DNA and hypothalamic cDNA encoding human preprothyrotropin-releasing hormone. Mol Endocrinol 1990; 4:551-556.)


Figure 7-15 Prolactin (PRL) and thyrotropin (TSH) secretory responses to intravenous injection of 800 µg of thyrotropin-releasing hormone (TRH) in humans. This figure shows that TRH induces discharge of both PRL and thyrotrophin, that the effect in females is greater than that in males (presumably owing to estrogen sensitization of the pituitary), and that thyrototoxicosis inhibits the response of both PRL and thyrotropin to TRH. An inhibitory effect on the TRH response is noted at the upper limit of the normal range of thyroid hormone levels and is a sensitive test of minor degrees of thyroid hormone excess. Although TRH is a potent prolactin-releasing factor (PRF), there is evidence that there is another PRF physiologically connected to PRL regulation. (Replotted from data of Bowers C, Friesen HG, Hwang P, et al. Prolactin and thyrotropin release in man by synthetic pyroglutamylhistidyl-prolinamide. Biochem Biophys Res Commun 1971; 45:1033-1041.)
TRH has some behavioral effects in rats that are similar to those of TRH but no other proven actions. Cyclic His-Pro is reported to act as a PRF and to have other neural effects, including reversal of ethanol-induced sleep (TRH is also effective in this system), elevation of brain cyclic guanosine monophosphate levels, an increase in stereotypic behavior, modification of body temperature, and inhibition of eating behavior. Some of the effects of TRH may be mediated through cyclic His-Pro, but the fact that cyclic His-Pro is abundant in some areas and is not proportional to the amount of TRH suggests that the peptidoe may not be derived solely from TRH.

Extrapituitary Function

TRH is present in virtually all parts of the brain: cerebral cortex, circumventricular structures, neuropehysis, pineal gland, and spinal cord. TRH is also found in pancreatic islet cells and in the gastrointestinal tract. Although it is present in low concentration, the total amount in extrahypothalamic tissues exceeds the amount in the hypothalamus.

The extensive extrahypothalamic distribution of TRH, its localization in nerve endings, and the presence of TRH receptors in brain tissue suggest that TRH serves as a neurotransmitter or neuromodulator outside the hypothalamus. TRH is a general stimulant and induces hyperthermia on intracerebroventricular injection, suggesting a role in central thermoregulation.

Clinical Applications

The use of TRH for the diagnosis of hyperthyroidism is less common since the development of ultrasensitive assays for thyroid-stimulating hormone (TSH). Its use to discriminate between hypothalamic and pituitary causes of thyrotropin deficiency has also declined because of the test’s poor specificity, but the application of ultrasensitive assays in conjunction with the TRH test has not been fully evaluated. TRH testing is also not of value in the differential diagnosis of causes of hyperprolactinemia but is useful for the demonstration of residual abnormal somatotropin-secreting cells in acromegalic patients who release hGH in response to TRH before treatment.

Studies of the effect of TRH on depression have shown inconsistent results, possibly because of poor blood-brain barrier penetration. Intrathecal administration of TRH may improve responses in depressed patients, but its clinical utility is unknown. Although a role for TRH in depression is not established, many depressed patients have a blunted thyrotropin response to TRH and changes in TRH responsiveness correlate with the clinical course. The mechanism by which blunting occurs is unknown.

TRH has been proposed as a treatment for women with threatened premature labor to stimulate the production of lung surfactant in the preterm fetus. Despite encouraging results in early studies, several large-scale trials failed to show improvement in the survival of babies so treated.

TRH has been evaluated for the treatment of spinal muscle atrophy and amyotrophic lateral sclerosis; transient improvement in strength was reported in both disorders, but the combined experience at many centers using a variety of treatment protocols including long-term intrathecal administration failed to confirm efficacy. TRH administration also reduces the severity of experimentally induced spinal and ischemic shock and head trauma. TRH has been used to treat children with neurologic disorders including West's syndrome, Lennox-Gastaut syndrome, early infantile epileptic encephalopathy, and intractable epilepsy. TRH has been proposed to be an analeptic agent. Sleeping or drug-sedated animals were awakened by the administration of TRH, reportedly reversed sedative effects of ethanol in humans, and TRH is said to have awakened a patient with a profound sleep disorder caused by a hypothalamic and midbrain eosinophilic granuloma.

Regulation of Thyrotropin Release

The secretion of thyrotropin is regulated by two interacting elements: negative feedback by thyroid hormone and open-loop neural control by hypothalamic hypothypsyedetic factors.

In the context of a feedback system, the level of thyroid hormone in blood or of its unbound fraction is the controlled variable and the set-point is the normal resting level of plasma thyroid hormone. Secretion of thyrotropin is inversely regulated by the level of thyroid hormone so that deviations from the set-point of control lead to appropriate changes in the rate of thyrotropin secretion. Factors that determine the rate of thyrotropin secretion required to maintain a given level of thyroid hormone include the rate at which thyrotropin and thyroid hormone disappear from the blood (turnover rate) and the rate at which T4 is converted to its more active form, T3.

TRH reportedly reversed sedative effects of ethanol in human and T4 inhibition of the thyrotropin response to TRH, and barely detectable decreases in plasma thyroid hormone levels were sufficient to sensitize the pituitary to TRH. TRH stimulates thyrotropin secretion within a few minutes through its action on a membrane receptor, whereas thyroid hormone actions,
mediated by intranuclear receptors, require several hours to take effect (see Chapter 10).

The secretion of hypothalamic TRH is also regulated by thyroid hormone feedback. Systemic injections of $T_3$ or implantations of tiny $T_3$ pellets in the paraventricular nucleus of hypothyroid rats (Fig. 7-18) (Figure Not Available) reduced the concentration of TRH mRNA and TRH prohormone in TRH-secreting cells. Thyroid hormone also suppressed TRH secretion into hypophyseal-portal blood in sheep.

$T_3$ in the blood gains access to TRH-secreting neurons in the hypothalamus by way of the CSF. The hormone is taken up by epithelial cells of the choroid plexus of the lateral ventricle of the brain, bound within the cell to locally produced transferrin ($T_4$-binding protein), and then secreted across the blood-brain barrier. Within the brain, $T_3$ is converted to $T_2$ by type II deiodinase, and $T_2$ interacts with subtypes of the thyroid hormone receptor, $TR_1$, $TR_2$, and $TR_3$ in the paraventricular nucleus and other brain cells (see Chapter 10). The set-point of the pituitary-thyroid axis is determined by thyroid hormone levels within the brain. $T_3$ in the circulation is not transported into brain in this manner but presumably gains access to the paraventricular TRH neurons across the blood-brain barrier. The brain $T_3$ transport and deiodinase system account for the fact that higher levels of $T_3$ are required to suppress pituitary-thyroid function after administration of $T_3$ than after administration of $T_4$.

Transferrin is present in the brain of early reptiles and in addition is synthesized by the liver in warm-blooded animals. During embryogenesis in mammals, transferrin is first detected when the blood-brain barrier appears, ensuring thyroid hormone access to the developing nervous system.

**Neural Control**

The hypothalamus determines the set-point of feedback control around which the usual feedback regulatory responses are elicited. Lesions of the thyrotropic area lower basal thyroid hormone levels and make the pituitary more sensitive to inhibition by thyroid hormone, and high doses of TRH raise thyrotropin and thyroid hormone levels. Synthesis of TRH in the paraventricular nucleus is regulated by feedback actions of thyroid hormones. The hypothalamus can override normal feedback control through an open-loop mechanism involving neuronal inputs to the hypothysyrotropic TRH neurons (see Fig. 7-16). For example, cold exposure caused a sharp increase in thyrotropin release in animals and in human newborns. Circadian changes in thyrotropin secretion are another example of brain-directed changes in the set-point of feedback control, but if thyroid hormone levels are sufficiently elevated, as in hyperthyroidism, TRH cannot overcome the inhibition.

Hypothalamic regulation of thyrotropin secretion is also influenced by two inhibitory factors, somatostatin and dopamine. Antisomatostatin antibodies increase basal thyrotropin levels and potentiate the response to stimuli that normally induce thyrotropin release in the rat, such as cold exposure and TRH administration. Thyroid hormone in turn inhibits the release of somatostatin, implying coordinated, reciprocal regulation of TRH and somatostatin by thyroid hormone. GH stimulates hypothalamic somatostatin synthesis and can inhibit thyrotropin secretion. The role of somatostatin in the regulation of thyrotropin secretion in humans is uncertain.

Dopamine has modest effects on thyrotropin secretion, and blockade of dopamine receptors (in the human) stimulates thyrotropin secretion slightly. Changes in the metabolism of thyroid hormone also influence $T_3$ homeostasis within the brain. In states of thyroid hormone deficiency, brain $T_3$ levels are maintained by an increase in the deiodinase that converts $T_2$ to $T_3$.

The pineal gland has been reported to inhibit thyroid function in some but not all studies. The pineal gland contains TRH, and in the frog its content changes with the season and with light and dark cycles independently of hypothalamic thyrotropin.

**Circadian Rhythm**

Plasma thyrotropin in humans is characterized by a circadian periodicity, with a maximum between 9 AM and 5 PM and a minimum between 4 PM and 7 PM. Smaller ultradian thyrotropin peaks occur every 90 to 180 minutes, probably because of bursts of TRH release from the hypothalamus, and are physiologically important in controlling the synthesis and glycosylation of thyrotropin. Glycosylation is a determinant of thyrotropin potency.

**Temperature**

External cold exposure activates and high ambient temperature inhibits the pituitary-thyroid axis in animals, and analogous changes occur in humans under certain conditions. Exposure of infants to cold at birth causes an increase in blood thyrotropin levels, possibly because of alterations in the turnover and degradation of the thyroid hormones. Blood thyroid hormone levels are higher in the winter than in the summer in individuals in cold climates but not in other climates. However, it is difficult to show that changes in environmental or body temperature in adults influence thyrotropin secretion. For example, exposure to cold ambient temperature or central hypothalamic cooling does not modify thyrotropin levels in young men. Behavioral changes, activation of the sympathetic nervous system, and shivering appear to be more important in temperature regulation in adults than the thyroid response.

The autonomic nervous system and the thyroid axis work together to maintain temperature homeostasis in mammals, and TRH plays a role in both pathways. Hypothalamic TRH release is rapidly (30 to 45 minutes) increased in rats exposed to cold. Rapid inhibition of somatostatin release in the median eminence has also been documented, and both changes appear to play important roles in the rise in plasma TSH induced by cold exposure. TRH mRNA is elevated within an hour of cold exposure (see Fig. 7-18 C and D) (Figure Not Available). The regulation of hypothalamic TRH release and expression by cold is largely mediated by catecholamines. Noradrenergic and adrenergic fibers, originating in the brain stem, are found in close proximity to TRH nerve endings in the median eminence, and a rapid rise in TRH release was seen after noradrenaline treatment of hypothalamic fragments containing mainly median eminence. Brain stem adrenergic and noradrenergic fibers also make synaptic contacts with TRH neurons in the PVH (see Fig. 7-16), and thus catecholamines are likely to be involved in the regulation of TRH gene expression by cold. TRH neurons in the PVH are densely innervated by NPY terminals, and a portion of the NPY terminals arising from the C1, C2, and C3, and A1 cell groups of the brain stem and projecting to the PVH are shown to be catecholaminergic. Somatostatin, dopamine, and serotonin also play a variety of roles in the regulation of TRH.

**Stress**

Stress is another determinant of thyrotropin secretion. In humans physical stress inhibits thyrotropin release, as indicated by the finding that in the euthyroid sick syndrome low $T_3$ and $T_4$ do not cause compensatory increases in thyrotropin secretion as would occur in normal individuals.

A number of observations demonstrate interactions between the thyroid and adrenal axes. Physiologically, the bulk of evidence suggests that glucocorticoids in humans and rodents act to blunt the thyroid axis through actions in the CNS. Some actions may be direct because the TRH gene (see Fig. 7-13) (Figure Not Available) contains the glucocorticoid response element consensus sequence and hypothalamic TRH neurons appear to contain glucocorticoid receptors. The diurnal rhythm of cortisol is opposite that of TSH (see Fig. 7-12) (Figure Not Available) and acute administration of glucocorticoids can block the nocturnal rise in
TSH, but disruption of cortisol synthesis with metyrapone only modestly affects the TSH circadian rhythm.  

Several lines of evidence, however, identify conditions in which elevated glucocorticoids are associated with stimulation of the thyroid axis. Human depression is often associated with hypercortisolism and hyperthyroxinemia, and TRH mRNA levels are elevated by glucocorticoids in a number of cell lines as well as in cultured fetal hypothalamic TRH neurons from the rat. Thus, although glucocorticoids probably stimulate TRH production in TRH neurons, their overall inhibitory effect on the thyroid axis results from indirect glucocorticoid negative feedback on structures such as the hippocampus. Disruption of hippocampal suppression of the hypothalamic-pituitary-adrenal (HPA) axis is proposed to be involved in the hypercortisolemia commonly seen in affective illness, and disruption of hippocampal inputs to the hypothalamus have been shown to produce a rise in hypophysiotropic TRH in the rat.  

Starvation  

The thyroid axis is depressed during starvation, presumably to help conserve energy by depressing metabolism (see Fig. 7-18 E to G) (Figure Not Available). In humans, reduced T3, T4, and TSH are seen during starvation or fasting. There are also changes in the thyroid axis in anorexia nervosa, such as low blood levels of T3 and low normal levels of T4 (see Chapter 33). Inappropriately low levels of TSH are found, suggesting defective activation of TRH production by low thyroid hormone levels. During starvation in rodents, reduced TRH release into hypophyseal portal blood and reduced pro-TRH mRNA levels are seen, despite lowered thyroid hormone levels. Reduced basal TSH levels are also usually present.

The hypothyroidism seen in fasting or in the leptin-deficient Lepob/Lepob mouse can be reversed by administration of leptin, and the evidence suggests that the mechanism involves leptin's ability to up-regulate TRH gene expression in the PVH (see Fig. 7-18 E to G) (Figure Not Available). Leptin appears to act both directly through leptin receptors on hypophysiotropic TRH neurons and indirectly through its actions on other hypothalamic cell groups, such as arcuate nucleus POMC and NPY-agouti-related peptide (AgRP) neurons. TRH neurons in the PVH receive dense NPY-AgRP and POMC projections from the arcuate and express NPY and melanocortin-4 receptors, and -MSH administration partially prevents the fasting-induced drop in thyroid hormone levels. Indeed, the TRH promoter contains a signal transducer and activator of transcription (STAT) response element and a cAMP response element that have been demonstrated to mediate induction of TRH gene expression by leptin and -MSH, respectively, in a heterologous cell system (see Fig. 7-13) (Figure Not Available).  

The regulation of TRH by metabolic state is likely to be under redundant control, however, because, unlike rodents, leptin-deficient children are euthyroid, and both melanocortin-4 receptor (MC4R)deficient rodents and humans are euthyroid. Central TRH outside the paraventricular nucleus also plays a role in thermoregulation through the autonomic nervous system.  

Infection and Inflammation  

The molecular basis of infection- or inflammation-induced thyrotropin suppression is now established. Sterile abscesses or the injection of interleukin-1 (IL-1; endogenous pyrogen, a secretory peptide of activated lymphocytes) or of tumor necrosis factor (TNF-) inhibits thyrotropin secretion, TNF- inhibits thyrotropin secretion directly and induces functional changes in the rat characteristic of the "sick euthyroid" state. It is likely that the thyrotropin inhibition in animal models of the sick euthyroid syndrome is due to cytokine-induced changes in hypothalamic and pituitary function. IL-6, IL-1, and TNF- contribute to the suppression of TSH in the sick euthyroid syndrome.
Corticotropin-Releasing Hormone

Chemistry and Evolution

The HPA axis is the humoral component of an integrated neural and endocrine system that functions to respond to internal and external challenges to homeostasis (stressors). The system comprises the neuronal pathways linked to release of catecholamines from the adrenal medulla (fight-or-flight response) and the hypothalamic-pituitary control of ACTH release in the control of glucocorticoid production by the adrenal cortex. Pituitary ACTH release is stimulated primarily by CRH and to a lesser extent by AVP (see Chapter 9). The hypophysiotropic CRH neurons are located in the paraventricular division of the PVH and project to the median eminence (see Fig. 7-6 (Figure Not Available) Fig. 7-7 Fig. 7-8 Fig. 7-9).

In a broader context, the CRH system in the CNS is also quite important in the behavioral response to stress. This complex system includes not only nonhypophysiotropic CRH neurons but also three CRH-like peptides (urocortin I, urocortin II or stresscopin-like peptide, and urocortin III or stresscopin), at least two cognate receptors (CRH-R1 and CRH-R2), and a high-affinity CRH-binding protein, each with distinct and complex distributions in the CNS.

The Schally and Guillemin laboratories demonstrated in 1955 that extracts from the hypothalamus stimulated ACTH release from the pituitary. The primary active principle, CRH, was purified and characterized from the sheep in 1981 by Vale and colleagues. Human CRH is an amided 41-amino-acid peptide that is cleaved from the carboxyl terminus of a 196-amino-acid pro-hormone precursor by PC1 and PC2 and amidated (Fig. 7-7B). In general, the peptide is highly conserved; the human peptide is identical in sequence to the mouse and rat peptides but differs at seven residues from the ovine sequence. CRH and urocortin I, II, and III in mammals, fish urotensin, anuran sauvagine, and the insect diuretic peptides are members of an ancient family of peptides that evolved from an ancestral precursor early in the evolution of metazoans, approximately 500 million years ago. Comparison of peptide sequences in the vertebrate suggests grouping of the peptides into two families, CRH-urotensin-urocortin-sauvagine and urocortin II-urocortin III (Fig. 7-29). Urocortin and sauvagine appear to represent tetrapod orthologues of fish urotensin. Sauvagine, isolated originally from Phyllomedusa sauvagei, is an osmoregulatory peptide produced in the skin of certain frogs; urotensin is an osmoregulatory peptide produced in the cauliflower secretory system of the fish. Whereas isolation of CRH required 250,000 ovine hypothalami, the cloning of urocortin II and III was accomplished by computer search of the human genome database.

The CRH peptides signal by binding to CRH-R1 and CRH-R2 receptors that are members of the gut-brain family of G protein-coupled receptors and couple to Gs and activation of adenyl cyclase. Two splice variants of the latter that differ in the extracellular amino-terminal domain, CRH-R2 and CRH-R2, have been found in both rodents and humans, and a third N-terminal splice variant, CRH-R2, has been reported in the human. CRH, urotensin, and sauvagine are all potent agonists of CRH-R1, urocortin is a potent agonist of both receptors, and urocortins II and III are specific agonists of CRH-R2. CRH-mediated activation of the HPA axis appears to be exclusively mediated through CRH-R1 expressed in the corticotroph. The PVH is the site of the majority of CRH neurons projecting to the median eminence, although some CRH neurons projecting to the median eminence are found in most hypothalamic nuclei (Fig. 7-21A). Some CRH fibers in the PVH also project to the brain stem, and CRH neurons are also found elsewhere, primarily in limbic structures involved in processing sensory information and in regulating the autonomic nervous system. Sites include the prefrontal, insular, and cingulate cortices; amygdala; substantia nigra; periaqueductal gray; locus coeruleus; nucleus of the solitary tract; and parabrachial nucleus. In the periphery, CRH is found in human placenta, where it is up-regulated 6-fold to 40-fold during the third trimester; lymphocytes; autonomic nerves; and gastrointestinal tract. Urocortin is found at highest levels in the Edinger-Westphal nucleus, lateral superior oliv, and supraoptic nuclei of the rostral brain with, additional sites including the substantia nigra, ventral tegmental area, and dorsal raphe (Fig. 7-21B). In the human, urocortin is widely distributed with highest levels in the frontal cortex, temporal cortex, and hypothalamus and has also been reported in the Edinger-Westphal and olivary nuclei. In the periphery, urocortin is seen in placenta, mucosal inflammatory cells in the gastrointestinal tract, lymphocytes, and cardio myocytes. The tissue distribution of urocortins II and III is not well characterized as of this writing.

In addition to its expression in pituitary corticotrophs, CRH-R1 is found in the neocortex and cerebellar cortex, subcortical limbic structures, and amygdala, with little to no expression in the hypothalamus (Fig. 7-21C). CRH-R1 is also found in a variety of peripheral sites in humans, including ovary, endometrium, and skin. CRH-R2 is found mainly in the brain in rodents, with high levels of expression seen in the ventromedial hypothalamic nucleus and lateral septum (see Fig. 7-21C). CRH-R2 is seen centrally in cerebral arterioles and peripherally in gastrointestinal tract, heart, and muscle. In contrast, in humans CRH-R2 is seen in brain and periphery, and the and subtypes are primarily central. Little CRH-R2 message is seen in pituitary. Although CRH-R1 appears to be exclusively involved in regulation of pituitary ACTH synthesis and release, both receptors have been found to be expressed in the rodent adrenal cortex. Data suggest that this intra-adrenal CRH-ACTH system may be involved in fine-tuning of adrenocortical corticosterone release.

The CRH system is also regulated in both brain and periphery by a 37-kd high-affinity CRH-binding protein. This factor was initially postulated from the observation that CRH levels rise dramatically during the second and third trimesters of pregnancy without activating the pituitary-adrenal axis. Among hypophysiotropic factors, CRH is the only one for which a specific binding protein (in addition to the receptor) exists in tissue or blood. The placenta is the principal source of pregnancy-related CRH-binding protein. Human and rat CRH-binding proteins are homologous (85% amino acid identity), but in the rat the protein is expressed only in brain. The binding protein is species specific; bovine CRH, which is almost identical in sequence to rat-human CRH, has a lower affinity of binding to the human binding protein.

The functional significance of the CRH-binding protein is not fully understood. CRH-binding protein does not bind to the CRH receptor but does inhibit CRH action. For this reason CRH-binding protein probably acts to modulate CRH actions at the cellular level. Corticophin cells in the anterior pituitary have membrane CRH receptors and intracellular CRH-binding protein; conceivably, the binding protein acts to sequester or terminate the action of membrane-bound CRH. CRH-binding protein is present in many regions of the CNS, including cells that synthesize CRH and cells that receive innervation from CRH-containing neurons. The anatomic distribution of the protein, the variability of its location in relation to the presence of CRH, and its relative sparseness in the CRH tuberohypophysal neuronal system suggest a control system that is as yet poorly understood.

Figure 7-19 Structure of human corticotropin-releasing hormone (CRH) gene and protein. The sequence coding for CRH occurs at the terminus of the prohormone. Cleavage sites and the terminal Gly position are shown. PAM, peptidylglycine alpha-amidating monooxygenase; PC1/PC2, prohormone convertases 1 and 2; CRE, glucocorticoid regulating element; AMP, cyclic AMP-responsive element; UTR, untranslated. (Redrawn from data of Shibahara S, Morimoto Y, Furutani Y, et al. Isolation and sequence analysis of the human corticotropin-releasing factor precursor gene. EMBO J 1983;2:75779.)

Figure 7-20 Sequence comparison of members of the corticotropin-releasing hormone (CRH) peptide family. SPP, stresscopin related peptide; SCP, stresscopin.
Structure-activity relationship studies have demonstrated that C-terminal amidation and an -helical secondary structure 129 are both important for biologic activity of CRH. The first CRH antagonist described was termed -helical CRH12. A second, more potent antagonist, termed astressin, had the structure cyclo(3033)(-Phe11, Nle12, Gu13, Lys15, JhCRH27). Both peptides are somewhat nonspecific, antagonizing both CRH-R1 and CRH-R2. Because of the anxiogenic activity of CRH and urocortin, a number of pharmaceutical companies have developed small molecule CRH antagonists; several of the

Figure 7-23 (Figure Not Available) Distribution of corticotropin-releasing hormone (CRH), urocortin, and the CRH receptor 1 (CRH-R1) and CRH-R2 messenger ribonucleic acid sequences in the rat brain. A, noradrenergic cell group 1; A, noradrenergic cell group 5; ac, anterior commissure; BST, bed nucleus of the stria terminalis; CA1, central nucleus amygdala; CG, central gray; DR, dorsal raphe; DVC, dorsal vagal complex; H, hippocampus; LC, locus coeruleus; LDT, laterodorsal tegmental nucleus; LH, lateral hypothalamic area; ME, median eminence; MID, midline thamic nuclei; mPOA, medial preoptic area; MR, medial raphe; MVL, medial vestibular nucleus; PB, parabrachial nucleus; POR, paraventricular nucleus; PP, posterior pituitary; PVN, paraventricular nucleus; SEPT, septal region; Si, sub substantia innominata; st, stria. (From Swanson LW, Sawchenko PE, Rivier J, et al. Organization of ovine corticotropin-releasing factor immunoreactive cells and fibers in the rat brain: an immunohistochemical study. Neuroendocrinology 1983; 36:165186; Bittencourt et al. J Comp Neurol 1999; 415:285, Fig. 17; Steckler and Holste. Biol Psychol 1999; 46:1490, Fig. 1.)

molecules are currently in clinical trials for anxiety and depression (discussed in more detail later). Thus far, this structurally diverse group of small molecule compounds, such as antalarmin, CP-154,526, and NBI27914, are potent antagonists of CRH-R1, with little activity at CRH-R2. The efficacy of these compounds across the entire behavioral, neuroendocrine, and autonomic repertoire of response to stress has been demonstrated in a number of laboratory animal studies. For example, oral administration of antalarmin in a social stress model in the primate (introduction of strange males) reduced behavioral measures of anxiety such as lack of exploratory behavior, decreased plasma ACTH and cortisol, and reduced plasma epinephrine and norepinephrine. A peptide antagonist with 100-fold selectivity for the CRH 2 receptor, ( -Phe11, His13, Jh14, Jh15) sauvagine 1140 or antisauvagine-30, has also been described. 130 Effects on the Pituitary and Mechanism of Action

Administration of CRH to humans causes prompt release of corticotropin into the blood, followed by secretion of cortisol (Fig. 7-22) and other adrenal steroids including aldosterone. 131,132 Most studies have used ovine CRH, which is more potent and longer acting than human CRH, but human and porcine CRHs appear to have equal diagnostic value. 133 The effect of CRH is specific to corticotropin release and is inhibited by glucocorticoids.

As mentioned before, CRH acts on the pituitary corticotroph primarily by binding to CRH-R1 and activating adenyl cyclase. The concentration of cAMP in the tissue is increased in parallel with the biologic effects and is reduced by glucocorticoids. The rate of transcription of the mRNA that encodes the corticotropin prohormone POMC is also enhanced by CRH, indicating that CRH is a trophic factor as well as a releasing hormone.

Figure 7-22 Changes in plasma levels of corticotropin and serum levels of cortisol after intravenous injection of corticotropin-releasing hormone in a group of six normal men. The initial prompt response in corticotropin is followed by a delayed secondary change in cortisol. To convert corticotropin values to picograms per liter, multiply by 0.2202. To convert cortisol values to milligrams per liter, multiply by 27.59. ACTH, adrenocorticotropic hormone. (From Grossman A, Kruysseman ACN, Perry L, et al. New hypothalamic hormone, corticotropin-releasing factor, specifically stimulates the release of adrenocorticotropic hormone and cortisol in man. Lancet 1982; 1:921922.)

Extrapituitary Functions

CRH and the urocortin peptides have a wide range of biologic activities in addition to the hypophyseotropic role of CRH in regulating ACTH synthesis and release. Centrally, these peptides have behavioral activities in anxiety, mood, arousal, locomotion, reward, and feeding 134,135,136 and increase sympathetic activation. Many of the nonhypophyseotropic behavioral and autonomic functions of these peptides can be viewed as complementary to activation of the HPA axis in the maintenance of homeostasis under exposure to stress. In the periphery, activities have been reported in immunity, cardiac function, gastrointestinal function, and reproduction.

The CRH and urocortin peptides have a repertoire of behavioral and autonomic actions after central administration that suggests a role for these pathways in mediating the behavioral-autonomic components of the stress response. Hyperactivity of the HPA axis is a common neuroendocrine finding in affective disorders (Fig. 7-23) (for reviews see references 133,137,138). Furthermore, normalization of HPA regulation is highly predictive of successful treatment. Defective dexamethasone suppression of CRH release, implying defective corticosterone receptor signaling, 139 is seen not only in depressed patients but also in healthy subjects with a family history of depression. 140 Depressed patients also show elevated levels of CRH in the CSF. 141

Central administration of CRH or urocortin activates neuronal cell groups involved in cardiovascular control and increases blood pressure, heart rate, and cardiac output. 142 However, urocortin is expressed in cardiac myocytes, 143 and intravenous administration of CRH or urocortin decreases blood pressure and increases heart rate in most species, including humans. 144 This hypertensive effect is probably mediated peripherally because ganglion blockade did not disrupt the hypotensive effects of intravenous urocortin. 145 Furthermore, high levels of CRH-R2 have been seen in the cardiac atria and ventricles. 146,147,148,149 and knockout of the CRH-R2 gene in the mouse removed the hypotensive effects of intravenous urocortin administration. 145,147,148

CRH and AVP also play an important role in the regulation of inflammatory responses. As described later (“Neuroendocrine-Immune Interaction”), cytokines play an important role in extinguishing inflammatory responses through activation of CRH and AVP neurons in the paraventricular nucleus and subsequent elevation of antiinflammatory glucocorticoids. Interestingly, CRH is generally seen to be proinflammatory in the periphery, where it is found in sympathetic efferents, sensory afferent nerves, leukocytes, and in macrophages in some species. 150

CRH is also made as a paracrine factor by the endometrium, where it may play a role in decidualization and implantation and act as a uterine vasodilator. 151

The relative contributions of each of the CRH-urocortin peptides and receptors to the different biologic functions reported has been the topic of considerable analysis, given the receptor-specific antagonists already described as well as the CRH. 129 CRH-R1, 132 CRH-R2, 152,153 and knockout mice available for study. 152,153 Examining three potent stressorstrastress, ether, and fastingthese studies demonstrated that other ACTH secretagogues, such as vasopressin, oxytocin, and catecholamines, could not replace CRH in its role in mounting the stress response. In contrast, augmentation of glucocorticoid secretion by a stressor after prolonged stress was not effective in the CRH knockout mouse, implicating CRH-independent mechanisms. 154

Although CRH is a potent anxiogenic peptide, 155 the CRH knockout mouse exhibits normal anxiety behaviors in, for example, conditioned fear paradigms. 156 The nonpeptide CRH-R1 specific antagonist CP-154,526 was anxiolytic in a shock-induced freezing paradigm in both wild-type and CRH knockout mice, 156 suggesting that the anxiogenic activity is a CRH-like peptide acting at the CRH-R1 receptor.
CRH and urocortin peptides also have potent anorexigenic activity, implicating the CRH system in stress-induced inhibition of feeding. Stress-induced inhibition of feeding remained intact, however, in the CRH knockout mouse.\textsuperscript{364} Likewise, suppression of the proestrous LH surge by restraint was intact in the CRH knockout mouse.\textsuperscript{365} Both CRH-R1 and CRH-R2 knockout strains had normal weight and feeding behavior but were distinctly different from wild-type mice in the anorexigenic response to centrally administered urocortin or CRH. The CRH-R1 deficient mice lacked the acute anorexigenic response (0 to 1.5 hours) to urocortin seen in wild-type mice.\textsuperscript{366} Both wild-type and CRH-R1 / mice exhibited comparable reduction in feeding 3 to 11 hours after administration. In contrast, the late phase of urocortin responsiveness appeared to depend on the presence of CRH-R2.\textsuperscript{367} Thus, signaling through CRH-R1 and CRH-R2 appears to play a complex role in the acute effects of stress on feeding behavior.

Clinical Applications

No useful therapeutic applications of CRH or CRH-like peptides have been reported, although the peptide has been demonstrated to have a number of activities in human and primate studies. For example, intravenous administration of CRH was found to stimulate energy expenditure and has been proposed for use in weight loss. The development of small molecule, orally available, CRH-R1 antagonists has, however, led to phase I clinical trials for anxiety and depression. An early study of 20 patients demonstrated significant reductions in scores of anxiety and depression, using ratings determined by either patient or clinician.\textsuperscript{368}

Feedback Control

The administration of glucocorticoids inhibits corticotropin secretion; removal of the adenals (or administration of drugs that impair secretion of glucocorticoids) leads to increased corticotropin release. The set-point of pituitary feedback is determined by the hypothalamus acting through hypothalamic releasing hormones CRH and vasopressin. The synthesis of their respective mRNAs are inhibited.\textsuperscript{369} Glucocorticoids act on both the pituitary corticotrophs and the hypothalamic neurons that secrete CRH and vasopressin. These regulatory actions are analogous to the control of the pituitary-thyroid axis. However, whereas thyrotropin becomes completely unresponsive to TRH when thyroid hormone levels are sufficiently high, severe neurogenic stress and large amounts of CRH can break through the feedback inhibition by glucocorticoids. A still higher level of feedback control is exerted by glucocorticoid-responsive neurons in the hippocampus that project to the hypothalamus; these neurons affect the activity of CRH hypophysiotropic neurons and determine the set-point of pituitary responsiveness to glucocorticoids.\textsuperscript{370}

Glucocorticoids are lipid soluble and enter the brain through the blood-brain barrier.\textsuperscript{371} In brain and pituitary they can bind to two receptors, type I (the mineralocorticoid receptor, so named because it binds aldosterone and glucocorticoids with high affinity) and type II (the glucocorticoid receptor, which has low affinity for mineralocorticoids).\textsuperscript{372} Glucocorticoid action involves binding of the steroid-receptor complex to regulator sequences in the genome.\textsuperscript{373} Type I receptors are saturated by basal levels of glucocorticoids, whereas type II receptors are not saturated under basal conditions but approach saturation during peak phases of the circadian rhythm and during stress. These differences and differences in regional distribution within the brain suggest that type I receptors determine basal activity of the hypothalamic-pituitary axis and that type II receptors mediate stress responses.

In the pituitary, glucocorticoids inhibit secretion of corticotropin and the synthesis of POMC mRNA in the hypothalamus, the secretion of CRH and vasopressin and the synthesis of their respective mRNAs is inhibited.\textsuperscript{374} Glucocorticoids may also act directly on neuronal cell membranes to change corticotropin secretion rapidly.\textsuperscript{375}

Glucocorticoids block stress-induced corticotropin release. The latency of the inhibitory effect is so short (less than 30 minutes)\textsuperscript{376} that it is possible that gene regulation is not the sole basis of the response. Long-term suppression (more than 1 hour) clearly acts through genomic mechanisms. Glucocorticoid receptors are also found outside the hypothalamus in the septum and amygdala,\textsuperscript{377} structures that are involved in the emotional changes in hypercortisolism and hypocortisolism. Hippocampal neurons are damaged by prolonged elevation of glucocorticoids during prolonged stress.

Neural Control

Significant physiologic or psychological stressors evoke an adaptive response that commonly includes activation of both the HPA axis and the sympathoadrenal axis. The end products of these pathways then help to mobilize resources to cope with the physiologic demands in emergency situations, acutely through the fight-or-flight response and over the long term through systemic effects of glucocorticoids on functions such as gluconeogenesis and energy mobilization (see Chapter 33). The HPA axis also has unique stress-specific homeostatic roles, the best example being the role of glucocorticoids in down-regulating immune responses after infection and other events that stimulate cytokine production by the immune system (see “Neuroendocrine-Immune Interactions”).

The paraventricular nucleus is the primary hypothalamic nucleus responsible for providing the integrated whole-animal response to stress.\textsuperscript{378} This nucleus contains three major types of effector neurons that are spatially distinct from one another within it: (1) magnocellular oxytocin and vasopressin neurons that project to the posterior pituitary and participate in the regulation of blood pressure, fluid homeostasis, lactation, and parturition; (2) neurons projecting to the brain stem and spinal cord that regulate a variety of autonomic responses including sympathoadrenal activation; and (3) parvocellular CRH neurons that project to the median eminence and regulate ACTH synthesis and release. Many CRH neurons coexpress AVP, which acts as an auxiliary ACTH secretagogue, synergistic with CRH to maintain regular circadian variation in parvocellular versus magnocellular neurons but is also regulated somewhat differently from CRH by stresses on parvocellular cells expressing both peptides.\textsuperscript{379} Different stressors result in different patterns of activation of the three major visceromotor cell groups within the paraventricular nucleus, as measured by the general neuronal activation marker c-fos (Fig. 7-24).\textsuperscript{380} For example, salt loading down-regulates CRH mRNA in parvocellular CRH cells, up-regulates CRH in a small number of magnocellular CRH cells, but only activates magnocellular cells. Hemorrhage activates every division of the paraventricular nucleus, whereas cytokine administration primarily activates parvocellular CRH cells with some minor activation of magnocellular and autonomic divisions.

The synthesis and release of AVP, which regulates renal water absorption and vascular smooth muscle, are controlled mainly by the volume and toxicity of the blood. This information is relayed to the magnocellular AVP cell through the nucleus of the solitary tract and A1 noradrenergic cell group of the ventrolateral medulla and projections from a triad of CVGs lining the third ventricle, the SFO, MePO, and OVLT. Oxytocin is primarily involved in reproductive functions, such as parturition, lactation, and milk ejection, although it is cosecreted with AVP in response to osmotic and volume challenges, and oxytocin cells receive direct projections from the nucleus of the solitary tract as well as from the SFO, median preoptic nucleus (MePO), and OVLT. In contrast to the neurosecretory neurons functionally defined by the three peptides, CRH, oxytocin, and AVP, PVH neurons projecting to brain stem and spinal cord include neurons expressing each of these peptides.

In the rodent, a wide variety of stressors have been determined to activate parvocellular CRH neurons, including cytokine injection, salt loading, hemorrhage, adrenalectomy, restraint, foot shock, hypoglycemia, fasting, and other exposure. Thus, in contrast to the simplicity of inputs to magnocellular cells (Fig. 7-25A), it is not surprising that parvocellular CRH neurons receive a diverse and complex assortment of inputs (Fig. 7-26; see Fig. 7-25B). These may be divided into three major categories, brain stem, limbic forebrain, and hypothalamus. Because the PVH is not known to receive any direct projections from the cerebral cortex or thalamus, stressors involving emotional or cognitive processing must induce activity to the PVH.

Visceral sensory input to the PVH involves primarily two pathways. The nucleus of the solitary tract, the primary recipient of sensory information from the thoracic and abdominal viscera, sends dense catecholaminergic projections to the PVH, both directly and through relays in the ventrolateral medulla.\textsuperscript{384} These brain stem projections account for about half of the NPY fibers present in the PVH, exclusively through projections to the thalamus, stressors involving emotional or cognitive processing must involve activity to the PVH. By contrast, what are termed neurogenic, emotional, or psychological stressors involve, in addition, nociceptive or somatosensory pathways as well as cognitive and affective brain centers. Using elevation of c-fos as an indicator of neuronal activation, detailed studies have compared PVH-projecting neurons activated by IL-1
treatment (systemic stressor) versus foot shock (neurogenic stressor). Only catecholaminergic solitary tract nucleus and ventrolateral medulla neurons were activated by moderate doses of IL-1. In contrast, foot shock activated neurons of the solitary tract nucleus and ventrolateral medulla but also cell groups in the limbic forebrain and hypothalamus. Notably, pharmacologic or mechanical disruption of the ascending catecholaminergic fibers blocked IL-1-mediated activation but not foot shock-mediated activation of the HPA axis. Data suggest that pathways activated by other neurogenic and systemic stressors may overlap significantly with those activated by foot shock and IL-1 treatment, respectively.

Except for the catecholaminergic neurons of the nucleus of the solitary tract and ventrolateral medulla, parts of the bed nucleus of the stria terminals, and the dorsomedial nucleus of the hypothalamus, many inputs to the paraventricular nucleus, such as those deriving from the prefrontal cortex and lateral septum, are thought to act indirectly through local hypothalamic glutamatergic and GABAergic neurons with direct synapses to the CRH neurons. The bed nucleus of the stria terminals is the only limbic region with prominent direct projections to the PVH. With substantial projections from the amygdala, hippocampus, and septal nuclei, it may thus serve as a key integrative center for transmission of limbic information to the PVH.

Other Factors Influencing Secretion of Corticotropin

Circadian Rhythms

Levels of corticotropin and cortisol (in humans) peak in the early morning, fall during the day to reach a nadir at about midnight, and begin to rise between 1 AM and 4 AM (see Fig. 7-12) (Figure Not Available). Within the circadian cycle approximately 15 to 18 pulses of corticotropin can be discerned, their height varying with the time of day. The set-point of feedback control by glucocorticoids also varies in a circadian pattern. Pituitary-adrenal rhythms are entrained to the light-dark cycle and can be changed over several days by exposure to an altered light schedule. It has long been assumed that the rhythm of corticotropin secretion is driven by CRH rhythms, and CRH knockout mice were found to exhibit no circadian rhythm in corticosterone production. Remarkably, however, a diurnal rhythm in corticosterone was restored by a constant infusion of CRH to the CRH knockout mouse, suggesting that CRH is necessary to permit pituitary or adrenal responsiveness to another diurnal rhythm generator.

Corticotropin Releaseinhibiting Factor

Disconnection of the pituitary from the hypothalamus in several species leads to increased basal levels of corticotropin, and certain responses to physical stress (in contrast to psychological stress) are retained in such animals. These observations have led several investigators to postulate the existence of a corticotropin inhibitory factor analogous to dopamine in the control of PRL secretion and to somatostatin in the control of GH secretion. Candidate hypothalamic peptides to inhibit corticotropin release at the level of the pituitary include atrial natriuretic peptide, activins and inhibins, and sequence 178 to 199 of the TRH prohormone. There is not yet a consensus on the existence of a physiologically relevant corticotropin releaseinhibiting factor or on its identity.
HYPOPHYSOTROPIC HORMONES AND NEUROENDOCRINE AXES (Continued)

Growth HormoneReleasing Hormone

Evidence for neural control of GH secretion came from studies of its regulation in animals with lesions of the hypothalamus and from the demonstration that hypothalamic extracts stimulate the release of GH from the pituitary. When it was shown that GH is released episodically, follows a circadian rhythm, responds rapidly to stress, and is blocked by pituitary stalk section, the concept of neural control of GH secretion became a certainty. However, it was only with the discovery of the paraneoplastic syndrome of ectopic GHRH secretion by pancreatic adenomas in humans that sufficient starting material became available for peptide sequencing and subsequent cloning of a complementary deoxynucleotidic acid (cDNA).

Two principal molecular forms of GHRH occur in human hypothalamus: GHRH(144)-NH₂ and GHRH(140)-OH (Fig. 7-27). As with other neuropeptides, the various forms of GHRH arise from post-translational modification of a larger prohormone. The NH₂-terminal tyrosine of GHRH (or histidine in rodent GHRHs) is essential for bioactivity, but a COOH-terminal NH₂ group is not. Fragments as short as (129)-NH₂ are active, but GHRH(127)-NH₂ is inactive. A circulating type IV dipeptidylpeptidase potently inactivates GHRH to its principal and more stable metabolite, GHRH(344)-NH₂, which accounts for most of the immunoreactive peptide detected in plasma. As in the case of LHRH, there are species differences among GHRHs; the peptides from seven species range in sequence homology with the human peptide from 93% in the pig to 67% in the rat. The COOH-terminal end of GHRH exhibits the most sequence diversity among species, consistent with the exon arrangement of the gene and dispensability of these residues for GHRH receptor binding.

Despite its importance for the elucidation of GHRH structure, ectopic secretion of the peptide is a rare cause of acromegaly. Fewer than 1% of acromegalic patients have elevated plasma levels of GHRH (see Chapter 8). Approximately 20% of pancreatic adenomas and 5% of carcinoid tumors contain immunoreactive GHRH, but most are clinically silent.

In addition to expression in the hypothalamus, the GHRH gene is expressed eutopically in human ovary, uterus, and placenta, although its function in these tissues is not known. Studies in rat placenta indicate that an alternative transcriptional start site 10 kilobases upstream from the hypothalamic promoter is utilized together with an alternatively spliced exon 1a.

The GHRH receptor is a member of a subfamily of G protein-coupled receptors that includes receptors for VIP, pituitary adenyl cyclase-activating peptide, secretin, glucagon, glucagon-like peptide 1, calcitonin, parathyroid hormone or parathyroid hormone-related peptide, and gastric inhibitory polypeptide. GHRH elevates intracellular cAMP by its receptor coupling to a stimulatory G protein (Gs), which activates adenyl cyclase, increases intracellular free Ca²⁺, releases preformed GH, and stimulates GH mRNA transcription and new GH synthesis (see Chapter 8). GHRH also increases pituitary phosphatidylinositol turnover. Nonsense mutations in the human GHRH receptor gene are the cause of rare familial forms of GH deficiency and indicate that no other gene product can fully compensate for the specific receptor in pituitary.

Effects on the Pituitary and Mechanism of Action

Intravenous administration of GHRH to individuals with normal pituitaries caused a prompt, dose-related increase in serum GH that peaked between 15 and 45 minutes, followed by a return to basal levels by 90 to 120 minutes (Fig. 7-28). A maximally stimulating dose of GHRH is approximately 1 µg/kg, but the response differs considerably between individuals and within the same individual tested on different occasions, presumably because of cosecretagogue and somatostatin tone that exists at the time of GHRH injection. Repeated bolus administration or sustained infusions of GHRH over several hours cause a modest decrease in the subsequent GH secretory response to acute GHRH administration. However, unlike the marked desensitization of the LHRH receptor and decline in circulating gonadotropins that occur in response to continuous LHRH exposure, pulsatile GH secretion and insulin-like growth factor I (IGF-I) production are maintained by constant GHRH in the human. This response suggests the involvement of additional factors that mediate the intrinsic
GH secretion is regulated by hypothalamic GHRH and somatostatin interacting with circulating hormones and additional modulatory peptides at the level of both the pituitary hormones other than GH may also limit the applicability of GHS therapy. Finally, apart from actions on GH secretion, both GHRH and GHSs are being exploited in a cloning strategy that led to the identification of a G protein-coupled receptor GHS-R that is highly selective for the GH secretagogue class of ligands. The GHS-R is unrelated to the GHRH receptor and is highly expressed in the anterior pituitary gland and multiple brain areas, including the medial basal hypothalamus, the hippocampus, and the mesencephalic nuclei that are centers of dopamine and serotonin production. Peptidyl and nonpeptidyl GHSs are active when administered by intranasal and oral routes, are more potent on a weight basis than GHRH itself, are more effective in vivo than in vitro, synergize with coadministered GHRH and are almost ineffective in the absence of GHRH, and do not suppress somatostatin secretion. Prolonged infusions of GHRP amplify pulsatile GH secretion in normal men. GHRP administration, like that of GHRH, facilitates slow wave sleep. Patients with hypothalamic disease leading to GH deficiency have low or no response to hexarelin; similarly, pediatric patients with complete absence of the pituitary stalk have no GH secretory response to hexarelin.

GHRH has several known extrapituitary functions. The most important may be its activity as a sleep regulator. The administration of nocturnal GHRH boluses to normal men significantly increased the density of slow wave sleep, as also shown in other species. Furthermore, there is a striking correlation between the age-related declines in slow wave sleep and daily integrated GH secretion in healthy men. These and other data suggest that central GH secretion is under circadian entrainment and nocturnal elevations in GH secretion amplitude or frequency directly mediate sleep stage and sleep-induced increases in GH secretion.

GHRH has been reported to stimulate food intake in rats and sheep, and the effect is dependent on route of administration, time of administration, and macronutrient composition of the diet. The neuropeptide's physiologic relevance to feeding in humans is unknown, although a study indicated that GHRH stimulated food intake in patients with anorexia nervosa but reduced it in patients with bulimia or in normal female control subjects.

**Extrapituitary Functions**

GHRH has few known extrapituitary functions. In studies of the opioid control of GH secretion, several peptide analogues of met-enkephalin were found to be potent GH secretagogues. These include the GH-releasing peptide GHRP-6 (Fig. 7-29) (Figure Not Available), hexarelin (His-D2MeTrp-Ala-Trp-DPhe-Lys-NH2), and other more potent analogues including cyclic peptides and modified pentapeptides. Subsequently, a series of nonpeptidyl GHRP mimetics were synthesized with greater oral bioavailability, including the spiropiperidine MK-0677 and the shorter acting benzylpiperidine L-163,540 (see Fig. 7-29). Common to all these compounds, and the basis of their differentiation from GHRH in pharmacologic activity screens, is their activation of phospholipase C and inositol 1,4,5-trisphosphate. This property was exploited in a cloning strategy that led to the identification of a G protein-coupled receptor GHS-R that is highly selective for the GH secretagogue class of ligands. The GHS-R is unrelated to the GHRH receptor and is highly expressed in the anterior pituitary gland and multiple brain areas, including the medial basal hypothalamus, the hippocampus, and the mesencephalic nuclei that are centers of dopamine and serotonin production. Peptidyl and nonpeptidyl GHSs are active when administered by intranasal and oral routes, are more potent on a weight basis than GHRH itself, are more effective in vivo than in vitro, synergize with coadministered GHRH and are almost ineffective in the absence of GHRH, and do not suppress somatostatin secretion. Prolonged infusions of GHRP amplify pulsatile GH secretion in normal men. GHRP administration, like that of GHRH, facilitates slow wave sleep. Patients with hypothalamic disease leading to GH deficiency have low or no response to hexarelin; similarly, pediatric patients with complete absence of the pituitary stalk have no GH secretory response to hexarelin.

The potent biologic effects of GHRPs and the identification of the GHS-R suggested the existence of a natural ligand for the receptor that is involved in the physiologic regulation of GH secretion. A probable candidate for this ligand is the acylated peptide ghrelin, produced and secreted into the circulation from the stomach (see Fig. 7-29). The effects of ghrelin on GH secretion in humans are identical to or more potent than those of the non-natural GHRPs (see Fig. 7-26). In addition, ghrelin acutely increases circulating PRL, ACTH, cortisol, and aldosterone levels. There is debate concerning the extent and localization of ghrelin expression in the brain that must be resolved before the implications of gastric-derived ghrelin in the regulation of pituitary hormone secretion are fully understood. A proposed role for ghrelin in appetite and regulation of food intake is discussed later in this chapter.

**Clinical Applications**

GHRH stimulates growth in children with intact pituitaries, but the optimal dosage, route, and frequency of administration, as well as possible usefulness by the nasal route, have not been determined. The availability of recombinant hGH (which requires post-translational modification among the known neuropeptides.}
Negative feedback control of GH release is mediated by GH itself and by IGF-I, which is synthesized in the liver under control of GH. Direct GH effects on the hypothalamus are produced by short-loop feedback, whereas those involving IGF-I and other circulating factors influenced by GH, including free fatty acids and glucose, are long-loop systems analogous to the pituitary-thyroid and pituitary-adrenal axes. Control of GH secretion thus includes two closed-loop systems (GH and IGF-I) and one open-loop regulatory system (neural).

Although most of the evidence for a direct role of GH in its own negative feedback has been derived from animals, an elegant study in normal men demonstrated that GH pretreatment blocks the subsequent GH secretory response to GHRH by a mechanism that is dependent on somatostatin. The mechanism responsible for GH feedback through the hypothalamus has been largely elucidated in rodent models. GH receptors are selectively expressed on somatostatin neurons in the hypothalamic periventricular nucleus and on NPY neurons in the arcuate nucleus. C-fos gene expression is acutely elevated in both populations of GH receptor-positive neurons by GH administration, indicating an activation of hypothalamic circuitry that includes these neurons. Similarly, GHRH neurons in the arcuate nucleus are acutely activated by MK-0677 because of their selective expression of the GHS-R. Zheng and colleagues showed in the latter group of neurons that c-fos induction after MK-0677 administration was blocked by pretreatment of mice with GH. The effect must be indirect because there are no GH receptors on GHRH neurons. However, there are type 2 somatostatin receptors expressed on GHRH neurons, and the somatostatin analogue octreotide also significantly blocked c-fos activation in the arcuate nucleus by MK-0677. The inhibitory effects of either GH or octreotide pretreatment were abolished in knockout mice lacking the specific somatostatin receptor. Together with data from many other experiments, these results strongly support a model of GH negative feedback regulation that involves the primary activation of periventricular somatostatin neurons by GH. These tuberoinfundibular neurons then inhibit GH secretion directly by release of somatostatin in the median eminence, but they also indirectly inhibit GH secretion by way of collateral axonal projections to the arcuate nucleus that synapse on and inhibit GHRH neurons. It is probable from evidence in rodents that NPY and galanin also play a part in the short-loop feedback of GH secretion, but a definitive mechanism in humans is not yet established.

IGF-I has a major inhibitory action on GH secretion at the level of the pituitary gland. IGF-I receptors are expressed on human somatotroph adenoma cells and inhibit both spontaneous and GHRH-stimulated GH release. In addition, gene expression of both GH and the pituitary-specific transcription factor Pit-1 is inhibited by IGF-I. Conflicting data among species suggest that circulating IGF-I may also regulate GH secretion by actions within the brain. The feedback effects of IGF-I account for the fact that in conditions in which circulating levels of IGF-I are low, such as anorexia nervosa, protein-calorie starvation, and Laron dwarfism (the result of a defect in the GH receptor), serum GH levels are elevated.

The predominant hypothalamic influence on GH release is stimulatory, and section of the pituitary stalk or lesions of the basal hypothalamic cause reduction of basal and induced GH release. When the somatostatinergic component is inactivated (e.g., by antisomatostatin antibody injection in rats), basal GH levels and GH responses to the usual provocative stimuli are enhanced.

GH-containing nerve fibers that terminate adjacent to portal vessels in the external zone of the median eminence arise principally from within, above, and lateral to the infundibular nucleus in the hypothalamus, corresponding to rodent arcuate and ventromedial nuclei. Perikarya of the tuberoinfundibular somatostatin neurons are located almost completely in the median eminence and parvocellular component of the anterior paraventricular nucleus. Neuroanatomic and functional evidence suggests a bidirectional synaptic interaction between the two peptidergic systems. Multiple extrahypothalamic brain regions provide effenter connections to the hypothalamus and regulate GHRH and somatostatin neuronal activity. Somatosensory and affective information is integrated and filtered through the amygdaloid complex. The basolateral amygdala provides an excitatory input to the hypothalamus, and the central extended amygdala, which includes the central and medial nuclei of the amygdala together with the bed nucleus of the stria terminals, provides a GABAergic inhibitory input. Many intrinsic neurons of the hypothalamus also release GABA, often with a peptide cotransmitter. Excitatory cholinergic fibers arise to a small extent from forebrain projection nuclei but mostly from hypothalamic cholinergic interneurons, which...
The inhibitory actions of NPY and CRH on GH secretion are firmly established in the rodent and are secondary to increased somatostatin tone. Contradictory evidence exists in the human for both.

In many instances, the inhibition can be demonstrated only as a suppression of GH release induced by a pharmacologic stimulus. Serotonin's effect on GH release in humans was difficult to decipher because of the large number of receptor subtypes. However, clinical studies with the receptor-selective agonist clonidine reliably stimulates GH release, and for this reason a clonidine test was a standard diagnostic tool in pediatric endocrinology. The stimulatory effect is blocked by the specific \( \alpha_2 \)-agonist clonidine, which appears to involve a dual mechanism of action, inhibition of somatostatin neurons and activation of GHRH neurons. In addition, partial attenuation of the effects of clonidine by mixed 5-hydroxytryptamine type 1 and type 2 antagonists suggests that some of the relevant receptors are located presynaptically on serotoninergic nerve terminals and increase serotonin release. Both norepinephrine and epinephrine play physiologic roles in the adrenergic stimulation of GH secretion. The \( \alpha_2 \)-agonists have no effect on GH secretion in humans, but \( \alpha_2 \)-agonists such as the bronchodilator salbutamol inhibit GH secretion by stimulating the release of somatostatin from nerve terminals in the median eminence. These effects are blocked by propranolol, a nonspecific \( \beta \)-receptor antagonist. Dopamine generally has a net effect to stimulate GH secretion, but the mechanism is not clear because of multiple dopamine receptor subtypes and the apparent activation of both GHRH and somatostatin neurons.

Serotonin's effect on GH release in humans was difficult to decipher because of the large number of receptor subtypes. However, clinical studies with the receptor-selective agonist clonidine reliably stimulates GH release, and for this reason a clonidine test was a standard diagnostic tool in pediatric endocrinology. The stimulatory effect is blocked by the specific \( \alpha_2 \)-agonist clonidine, which appears to involve a dual mechanism of action, inhibition of somatostatin neurons and activation of GHRH neurons. In addition, partial attenuation of the effects of clonidine by mixed 5-hydroxytryptamine type 1 and type 2 antagonists suggests that some of the relevant receptors are located presynaptically on serotoninergic nerve terminals and increase serotonin release. Both norepinephrine and epinephrine play physiologic roles in the adrenergic stimulation of GH secretion. The \( \alpha_2 \)-agonists have no effect on GH secretion in humans, but \( \alpha_2 \)-agonists such as the bronchodilator salbutamol inhibit GH secretion by stimulating the release of somatostatin from nerve terminals in the median eminence. These effects are blocked by propranolol, a nonspecific \( \beta \)-receptor antagonist. Dopamine generally has a net effect to stimulate GH secretion, but the mechanism is not clear because of multiple dopamine receptor subtypes and the apparent activation of both GHRH and somatostatin neurons.

Serotonin's effect on GH release in humans was difficult to decipher because of the large number of receptor subtypes. However, clinical studies with the receptor-selective agonist clonidine reliably stimulates GH release, and for this reason a clonidine test was a standard diagnostic tool in pediatric endocrinology. The stimulatory effect is blocked by the specific \( \alpha_2 \)-agonist clonidine, which appears to involve a dual mechanism of action, inhibition of somatostatin neurons and activation of GHRH neurons. In addition, partial attenuation of the effects of clonidine by mixed 5-hydroxytryptamine type 1 and type 2 antagonists suggests that some of the relevant receptors are located presynaptically on serotoninergic nerve terminals and increase serotonin release. Both norepinephrine and epinephrine play physiologic roles in the adrenergic stimulation of GH secretion. The \( \alpha_2 \)-agonists have no effect on GH secretion in humans, but \( \alpha_2 \)-agonists such as the bronchodilator salbutamol inhibit GH secretion by stimulating the release of somatostatin from nerve terminals in the median eminence. These effects are blocked by propranolol, a nonspecific \( \beta \)-receptor antagonist. Dopamine generally has a net effect to stimulate GH secretion, but the mechanism is not clear because of multiple dopamine receptor subtypes and the apparent activation of both GHRH and somatostatin neurons.

Serotonin's effect on GH release in humans was difficult to decipher because of the large number of receptor subtypes. However, clinical studies with the receptor-selective agonist clonidine reliably stimulates GH release, and for this reason a clonidine test was a standard diagnostic tool in pediatric endocrinology. The stimulatory effect is blocked by the specific \( \alpha_2 \)-agonist clonidine, which appears to involve a dual mechanism of action, inhibition of somatostatin neurons and activation of GHRH neurons. In addition, partial attenuation of the effects of clonidine by mixed 5-hydroxytryptamine type 1 and type 2 antagonists suggests that some of the relevant receptors are located presynaptically on serotoninergic nerve terminals and increase serotonin release. Both norepinephrine and epinephrine play physiologic roles in the adrenergic stimulation of GH secretion. The \( \alpha_2 \)-agonists have no effect on GH secretion in humans, but \( \alpha_2 \)-agonists such as the bronchodilator salbutamol inhibit GH secretion by stimulating the release of somatostatin from nerve terminals in the median eminence. These effects are blocked by propranolol, a nonspecific \( \beta \)-receptor antagonist. Dopamine generally has a net effect to stimulate GH secretion, but the mechanism is not clear because of multiple dopamine receptor subtypes and the apparent activation of both GHRH and somatostatin neurons.
Somatostatin clearly implicate the 5-hydroxytryptamine 1D receptor subtype in the stimulation of basal GH levels. The drug also potentiates the effect of a maximal dose of GHRH, suggesting that the recurring theme of GH disinhibition by inhibition of hypothalamic somatostatin neurons in its mechanism of action. Histaminergic pathways acting through H1 receptors play only a minor, conditional stimulatory role in GH secretion in humans.

Acetylcholine appears to be an important physiologic regulator of GH secretion. Blockade of acetylcholinergic muscarinic receptors reduces or abolishes GH secretory responses to GHRH, glucagon and arginine, morphine, and exercise. In contrast, drugs that potentiate cholinergic transmission increase basal GH levels and enhance the GH response to GHRH in normal individuals or in subjects with obesity or Cushing's disease. In vitro acetylcholine inhibits somatostatin release from hypothalamic fragments, and acetylcholine can act directly on the pituitary to inhibit GH release. There may even be a paracrine cholinergic control system within the pituitary. However, the sum of evidence suggests that the primary mechanism of action of M1 agonists is inhibition of somatostatin neuronal activity or the release of peptide from somatostatinergic terminals. Short-term cholinergic blockade with the M1 muscarinic receptor antagonist pirenzipine reduced the GH excess of patients with poorly controlled diabetes mellitus. However, in the long term, cholinergic blockade did not prevent complications associated with the hyperinsulinemic state.

Many neuropeptides in addition to GHRH and somatostatin are involved in the modulation of GH secretion in humans. Among these, the evidence is most compelling for a stimulatory role of galanin acting in the human hypothalamus by a GHRH-dependent mechanism. Many GHRH neurons are immunopositive for galanin as well as neuropeptide and tyrosine hydroxylase. Galanin's actions may be explained, in part, by presynaptic facilitation of catecholamine release from nerve terminals and subsequent direct adrenergic stimulation of GHRH release. Opioid peptides also stimulate GH release, probably by activation of GHRH neurons, but under normal circumstances endogenous opioid tone in the hypothalamus is presumed to be low because opioid antagonists have little acute effect on GH secretion.

A larger number of neuropeptides are known or suspected to inhibit GH secretion in humans, at least under certain circumstances. The list includes NPY, CRH, calcitonin, oxytocin, neurotensin, VIP, and TRH. Inhibitory actions of NPY are well established in the rat. The effect on GH secretion is secondary to stimulation of somatostatin neurons and is of particular interest because of the presumed role in GH auto-feedback (discussed earlier) and the integration of GH secretion with regulation of energy intake and expenditure (discussed in a later section). Finally, TRH has the well-established paradoxical effect of increasing GH secretion in patients with acromegaly, type 1 diabetes mellitus, hypothyroidism, or hepatic and renal failure.

Factors Influencing Secretion of Growth Hormone

Human Growth Hormone Rhythms

The unraveling of rhythmic GH secretion has relied on a combination of technical innovations in sampling and GH assay, and sophisticated mathematical modeling including deconvolution analysis and the calculation of approximate entropy as a measure of orderliness or regularity in minute-to-minute secretory patterns. At least three distinct categories of GH rhythms, which differ markedly in their time scales, can be considered here. The daily GH secretion rate varies over two orders of magnitude from a maximum of nearly 2.0 mg/day in late puberty to a minimum of 20 µg/day in older or obese adults. The neonatal period is characterized by markedly amplified GH secretory bursts followed by a prepubertal decade of stable, moderate GH secretion of 200 to 600 µg/day. There is a marked increase in daily GH secretion during puberty that is accompanied by a commensurate rise in plasma IGF-I levels that constitute a state of physiologic hypersomatotropism. This pubertal increase in GH secretion is due to increased GH mass per secretory burst and not to increased pulse frequency. Although the changes are clearly related to the increases in gonadal steroid hormones and can be mimicked by administration of estrogen or testosterone to hypogonadal children, the underlying neuroendocrine mechanisms are not fully understood. Gonadal sex steroids play both an organizational and activational role in the adult, regulating expression of the genes for many of the peptides and receptors central to GH secretion, and this difference is postulated to be a key factor in producing the sexual dimorphism.

The neuroendocrine basis for sex differences in the ultradian rhythm of GH secretion is not fully understood. Gonadal sex steroids play both an organizational role during development of the hypothalamus and an activational role in the adult, regulating expression of the genes for many of the peptides and receptors central to GH regulation. In the human, unlike the rat, the hypothalamic actions of testosterone appear to be predominantly due to its aromatization to 17-estradiol and interaction with estrogen receptors. Hypothalamic somatostatin appears to play a more prominent role in men than in women in the regulation of pulsatile GH secretion, and this difference is postulated to be a key factor in producing the sexual dimorphism.

External and Metabolic Signals

The various peripheral signals that modulate GH secretion in humans are summarized in Table 7-6 (also see Fig. 7-30 and Fig. 7-32). Of particular importance are factors related to energy intake and metabolism because they provide a common signal between the peripheral tissues and hypothalamic centers regulating nondiocrine homeostatic signaling pathways as well as to the classical hypothalaseurotropic neurons. It is also in this complex arena that species-specific regulatory responses are particularly prominent, making extrapolations between rodent experimental models and human GH regulation less reliable. The neuroendocrine basis for sex differences in the ultradian rhythm of GH secretion is not fully understood. Gonadal sex steroids play both an organizational role during development of the hypothalamus and an activational role in the adult, regulating expression of the genes for many of the peptides and receptors central to GH regulation. In the human, unlike the rat, the hypothalamic actions of testosterone appear to be predominantly due to its aromatization to 17-estradiol and interaction with estrogen receptors. Hypothalamic somatostatin appears to play a more prominent role in men than in women in the regulation of pulsatile GH secretion, and this difference is postulated to be a key factor in producing the sexual dimorphism. Important triggers of GH release include the normal decrease in blood glucose level after intake of a carbohydrate-rich meal, absolute hypoglycemia, exercise, physical and emotional stress, and high intake of protein (mediated by amino acids). Some of the pathologic causes of elevated GH represent extremes of these physiologic signals and include protein-calorie starvation, anorexia nervosa, liver failure, and type 1 diabetes mellitus. A critical concept is that many of these GH triggers work through the same final common mechanism of somatostatin withdrawal and consequent disinhibition of GH secretion. In contrast, postprandial hyperglycemia, glucose infusion, elevated plasma free fatty acids, type 2 diabetes mellitus (with obesity and insulin resistance), and obesity are all associated with inhibition of GH secretion. The role of leptin in mediating either increases or decreases in GH release is complicated by its multiple sites of action and coexistent secretory environment. Similarly, other members of the cytokine family including IL-1, IL-2, IL-6, and endotoxin have been inconsistently shown to stimulate GH in humans.

The actions of steroid hormones on GH secretion in humans are complex because of their multiple loci of action within the proximal hypothalamic-pituitary components in addition to secondary effects on other neural and endocrine systems. Glucocorticoids in particular produce opposite responses that are dependent on the chronicity of administration. Moreover, glucocorticoid effects follow an inverted U-shaped dose-response curve. Both low and high glucocorticoid levels reduce GH secretion, the former because of decreased GH gene expression and somatotroph responsiveness to GHRH and the latter because of increased hypothalamic somatostatin tone and decreased GHRH. Similarly, physiologic levels of thyroid hormones are necessary to maintain GH secretion and promote GH gene expression. Excessive thyroid hormone is also inhibitory to the GH axis, and the mechanism is speculated to be a combination of increased hypothalamic somatostatin tone, GH deficiency, and suppressed pituitary GH production.
Somatostatin, the peptide responsible for this inhibition of GH secretion and the inhibition of insulin secretion by a pancreatic islet cell, was eventually isolated from hypothalamus and sequenced by Brazeau and colleagues in 1973. The term somatostatin was originally applied to a cyclic peptide containing 14 amino acids (somatostatin-14 [SST-14]; see Fig. 7-33). Subsequently, a second form, N-extended somatostatin-28 (SST-28), was identified as a secretory product. Both forms of somatostatin are derived by independent cleavage of a common prohormone by prohormone convertases. In addition, the isolation of SST-28(112) in some tissues suggests that SST-14 can be secondarily processed from SST-28. SST-14 is the predominant form in the brain (including the hypothalamus), whereas SST-28 is the major form in the gastrointestinal tract, especially the duodenum and jejunum.

The name somatostatin is descriptively inadequate because the molecule also inhibits thyrotropin secretion from the pituitary and has nonpituitary roles including activity as a neurotransmitter or neuromodulator in the central and peripheral nervous systems and as a regulatory peptide in gut and pancreas. As a pituitary regulator, somatostatin is a true neurohormone, that is, a neuronal secretory product that enters the blood (hypophyseseal-portal circulation) to affect cell function at remote sites. In the gut, somatostatin is present in both the myenteric plexus, where it acts as a neurotransmitter, and epithelial cells, where it influences the function of adjacent cells as a paracrine secretion. Somatostatin can influence its own secretion from delta cells (an autocrine function) in addition to acting as a paracrine factor in pancreatic islets. Gut exocrine secretion can be modulated by intraluminal action, so it is also a lumone. Because of its wide distribution, broad spectrum of regulatory effects, and evolutionary history, this peptide can be regarded as an archetypal pan-system modulator.

The genes that encode somatostatin in humans (see Fig. 7-33) and a number of other species exhibit striking sequence homology, even in primitive fish such as the anglerfish. Furthermore, the amino acid sequence of SST-14 is identical in all vertebrates. Formerly, it was accepted that all tetrapods have a single gene encoding both SST-14 and SST-28 whereas teleost fish have two nonallelic pre-prosomatostatin genes (PPSI and PPSII), each of which encodes only one form of the mature somatostatin peptides. This situation implied that a common ancestral gene underwent a duplication event after the split of teleosts from the descendants of tetrapods. However, both lampreys and amphibians, which predate and postdate the teleost evolutionary divergence, respectively, have now been shown to have at least two PPS genes. A more distantly related gene has been identified in mammals that encodes cortistatin, a somatostatin-like peptide that is highly expressed in cortex and hippocampus. Cortistatin-14 differs from SST-14 by three amino acid residues but has high affinity for all known subtypes of somatostatin receptors (see later). The human gene sequence predicts a tripeptide-extended cortistatin-17 and a further N-terminally extended cortistatin-29. A revised evolutionary concept of the somatostatin gene family is that a primordial gene underwent duplication at or before the advent of chordates and the two resulting genes underwent mutation at different rates to produce the distinct pre-prosomatostatin and pre-procortistatin genes in mammals. A second gene duplication probably occurred in teleosts to generate PPSI and PPSII from the ancestral somatostatin gene.

Apart from its expression in neurons of the periventricular and arcuate hypothalamic nuclei and involvement in GH secretion discussed earlier, somatostatin is highly expressed in the cortex, lateral septum, extended amygdala, reticular nucleus of the thalamus, hippocampus, and many brain stem nuclei. Cortistatin is present in the brain at a small fraction of the levels of somatostatin and in a more limited distribution primarily confined to cortex and hippocampus. The molecular mechanisms underlying the development and hormonal regulation of somatostatin gene transcription have been most extensively studied in pancreatic islet cells. Less is known concerning the regulation of somatostatin gene expression in neurons except that activation is strongly controlled by binding of the phosphorylated transcription factor cAMP response element binding protein to its cognate cAMP response element contained in the promoter sequence. Enhancer elements in the somatostatin gene promoter that bind complexes of homeodomain-containing transcription factors (PAX6, PBX, PREP1) and up-regulate gene expression in pancreatic islets may actually represent core somatostatin elements in neurons (see Fig. 7-33), promoter elements TSE, and UE-A). Conversely, another related cis element in the somatostatin gene (see Fig. 7-33, promoter element TSE) apparently binds a homeodomain transcription factor PDX1 (also called STF1/ID2/ID1/PF1) that is common to developing brain, pancreas, and foregut and regulates gene expression in both the CNS and gut.

The function of somatostatin in GH and thyrotropin regulation was considered earlier in this chapter. Its actions in the extrahypothalamic brain and diagnostic and therapeutic roles are considered in the remainder of this section and in Chapter 8. An additional function of somatostatin in pancreatic islet cell regulation is described in Chapter 29, and the manifestations of somatostatin excess as in somatostatinoma are described in Chapter 35.

Somatostatin Receptors

Five somatostatin receptor subtypes (SSTR1 to SSTR5) have been identified by gene cloning techniques (Table 7-7), and one of these (SSTR2) is expressed in two alternatively spliced forms. These subtypes are encoded by separate genes located on different chromosomes, are expressed in unique or partially overlapping distributions in multiple target organs, and differ in their coupling to second messenger signaling molecules and therefore in their range and mechanism of intracellular actions. The subtypes also differ in their binding affinity to specific somatostatin analogues. Certain of these differences have important implications for the use of somatostatin analogues in therapy and in diagnostic imaging.

All SSTR subtypes are coupled to pertussis toxinsensitive G proteins and bind SST-14 and SST-28 with high affinity in the low nanomolar range, although SST-28 is a uniquely high affinity for SSTR5. SSTR1 and SSTR2 are the most abundant subtypes in brain and probably function as presynaptic autoreceptors in the hypothalamus and limbic forebrain, respectively, in addition to their postsynaptic actions. SSTR4 is most prominent in hippocampus. All the subtypes are expressed in the pituitary, but SSTR2 and SSTR5 are the most abundant on somatotrophs. These two subtypes are also the most physiologically important in pancreatic islets, with SSTR5 responsible for inhibition of insulin secretion from beta cells and SSTR2 responsible for inhibition of glucagon from alpha cells.

Binding of somatostatin to its receptor leads to activation of one or more plasma membranebound inhibitory G proteins, which in turn inhibit adenyl cyclase activity and lower intracellular cAMP. Other G proteinmediated actions common to all SSTRs are activation of a vanadate-sensitive phosphotyrosine phosphatase and modulation of mitogen-activated protein kinase (MAPK). Different subsets of SSTRs are also coupled to inwardly rectifying K+ channels, voltage-dependent Ca2+.
channels, an Na+/H+ exchanger, -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid (AMPA)-kainate glutamate receptors,

TABLE 7-7 -- Characteristics of the Human Somatostatin Receptors

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>SSTR1</th>
<th>SSTR2</th>
<th>SSTR3</th>
<th>SSTR4</th>
<th>SSTR5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chromosome</td>
<td>14q13</td>
<td>17q24</td>
<td>22q13.1</td>
<td>20p11.2</td>
<td>16p13.3</td>
</tr>
<tr>
<td>Tissue distribution</td>
<td>Brain</td>
<td>Brain</td>
<td>Brain</td>
<td>Brain</td>
<td>Brain</td>
</tr>
<tr>
<td></td>
<td>Pituitary</td>
<td>Pituitary</td>
<td>Pituitary</td>
<td>Pituitary</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Islet</td>
<td>Islet</td>
<td>Islet</td>
<td>Islet</td>
<td>Islet</td>
</tr>
<tr>
<td></td>
<td>Stomach</td>
<td>Stomach</td>
<td>Stomach</td>
<td>Stomach</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Kidney</td>
<td>Kidney</td>
<td>Lung</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Liver</td>
<td></td>
<td>Placenta</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

phospholipase C, and phospholipase A2. The lowering of intracellular cAMP and Ca\(^{2+}\) is the most important mechanism for the inhibition of hormone secretion, and actions on phosphotyrosine phosphatase and MAPK are postulated to play a role in somatostatin's antiproliferative effect on tumor cells.

Effects on Target Tissues and Mechanism of Action

In the pituitary, somatostatin inhibits secretion of GH and thyrotropin and, under certain conditions, of PRL and ACTH as well. It exerts inhibitory effects on virtually all endocrine and exocrine secretions of the pancreas, gut, and gallbladder (Table 7-8). Somatostatin inhibits secretion by the salivary glands and, under some conditions, the secretion of parathyroid hormone and calcitonin. Somatostatin blocks hormone release in many endocrine-secreting tumors, including insulinomas, glucagonomas, VIPomas, carcinoid tumors, and some gastrinomas.

The physiologic actions of somatostatin in extrahypothalamic brain remain the subject of investigation. In the striatum, somatostatin increases the release of dopamine from nerve terminals by a glutamate-dependent mechanism. It is widely expressed in GABAergic interneurons of limbic cortex and hippocampus, where it modulates the excitability of pyramidal neurons. Temporal lobe epilepsy is associated with a marked reduction in somatostatin-expressing neurons in the hippocampus consistent with a putative inhibitory action on seizures. A wealth of correlative data has linked reduced forebrain and CSF concentrations of somatostatin with Alzheimer's disease, major depression, and other neuropsychiatric disorders, raising speculation about the role of somatostatin in modulating neural circuits underlying cognitive and affective behaviors.

Clinical Applications of Somatostatin Analogues

An extensive pharmaceutical discovery program has produced somatostatin analogues with receptor subtype selectivity and improved pharmacokinetics and oral bioavailability compared with the native peptide. Initial efforts focused on the rational design of constrained cyclic peptides that incorporated \(\alpha\)-amino acid residues and included the Trp\(^1\)-Lys\(^8\) dipeptide of somatostatin, which was shown by structure-function studies to be necessary for high-affinity binding to its receptor (see Fig. 7-33). Many such analogues have been studied in clinical trials including octreotide, lanreotide, vapreotide, and the hexapeptide MK-678. These compounds are agonists with

TABLE 7-8 -- Biologic Actions of Somatostatin Outside the Central Nervous System

<table>
<thead>
<tr>
<th>Inhibits Hormone Secretion by</th>
<th>Pituitary gland</th>
<th>GH, thyrotropin, ACTH, prolactin</th>
<th>Gastrointestinal tract</th>
<th>Gastrin</th>
<th>Secretin</th>
<th>Gastrointestinal polypeptide</th>
<th>Motilin</th>
<th>Glicentin (enteroglucagon)</th>
<th>VIP</th>
<th>Pancreas</th>
<th>Insulin</th>
<th>Glucagon</th>
<th>Somatostatin</th>
<th>Genitourinary tract</th>
<th>Renin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inhibits Other Gastrointestinal Actions</td>
<td>Gastric acid secretion</td>
<td>Gastric and jejunal fluid secretion</td>
<td>Gastric emptying</td>
<td>Pancreatic bicarbonate secretion</td>
<td>Pancreatic enzyme secretion</td>
<td>(Stimulates intestinal absorption of water and electrolytes)</td>
<td>Gastrointestinal blood flow</td>
<td>AVP-stimulated water transport</td>
<td>Bile flow</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extra Gastrointestinal Actions</td>
<td>Inhibits the function of activated immune cells</td>
<td>Inhibition of tumor growth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

GH, growth hormone; ACTH, adrenocorticotropic hormone; AVP, arginine vasopressin; VIP, vasoactive intestinal peptide.
Similarly high-affinity binding to SSTR2 and SSTR5, moderate binding to SSTR3, and no (or low) binding to SSTR1 and SSTR4. A combinatorial chemistry approach has led to a new generation of nonpeptidyl somatostatin agonists that bind selectively and with subnanomolar affinity to each of the five SSTR subtypes. In contrast to the marked success in development of potent and selective somatostatin agonists, there is a relative paucity of useful antagonists.

The actions of octreotide (SMS 201-995 or Sandostatin) illustrate the general potential of somatostatin analogues in therapy. It controls excess secretion of GH in acromegaly in most patients and shrinks tumor size in about one third. Octreotide is also indicated for the treatment of thyrotropin-secreting adenomas that recur after surgery. It is used to treat other functioning metastatic neuroendocrine tumors, including carcinoid, VIPoma, glucagonoma, and insulinoma, but is seldom of use for the treatment of gastrinoma. It is also useful in the management of many forms of diarrhea (acting on salt and water excretion mechanisms in the gut) and in reducing external secretions in pancreatic fistulae (thus permitting healing). A decrease in blood flow to the gastrointestinal tract is the basis for its use in bleeding esophageal varices, but it is not effective in the treatment of bleeding from a peptic ulcer.

The only major undesirable side effect of octreotide is reduction of bile production and of gallbladder contractility, leading to "sludging" of bile and an increased incidence of gallstones. Other common adverse effects including nausea, abdominal cramps, diarrhea secondary to malabsorption of fat, and flatulence usually subside spontaneously within 2 weeks of continued treatment. Impaired glucose tolerance is not associated with long-term octreotide therapy, despite an inhibitory effect on insulin secretion, because of compensating reductions in carbohydrate absorption and GH and glucagon secretion that are also caused by the drug.

Somatostatin analogues labeled with a radioactive tracer have been used as external imaging agents for a wide range of disorders. A 111In-labeled analogue of octreotide (OctreoScan) has been approved for clinical use in the United States and several other countries. The majority of neuroendocrine tumors and many pituitary tumors that express somatostatin receptors are visualized by external imaging techniques after administration of this agent; a variety of nonendocrine tumors and inflammatory lesions are also visualized, all of which have in common the expression of somatostatin receptors. Such tumors include nonsmall cell cancer of the lung (100%), meningioma (100%), breast cancer (74%), and astrocytomas (67%). Because activated T cells of the immune system display somatostatin receptors, inflammatory lesions that take up the tracer include sarcoidosis, Wegener's granulomatosis, tuberculosis, and many cases of Hodgkin's disease and non-Hodgkin's lymphoma. Although the tracer lacks specificity in differential diagnosis, its ability to identify the presence of abnormality and the extent of the lesion provides important information for management, including tumor staging. The use of a small hand-held radiation detector in the operating room makes it possible to ensure the completeness of removal of medullary thyroid carcinoma metastases.

New developments in the synthesis of tracers chelated to octreotide for positron emission tomography have allowed the sensitive detection of somatostatin receptor expression. The majority of neuroendocrine tumors and inflammatory lesions are also visualized, all of which have in common the expression of somatostatin receptors. Such tumors include nonsmall cell cancer of the lung (100%), meningioma (100%), breast cancer (74%), and astrocytomas (67%). Because activated T cells of the immune system display somatostatin receptors, inflammatory lesions that take up the tracer include sarcoidosis, Wegener's granulomatosis, tuberculosis, and many cases of Hodgkin's disease and non-Hodgkin's lymphoma. Although the tracer lacks specificity in differential diagnosis, its ability to identify the presence of abnormality and the extent of the lesion provides important information for management, including tumor staging. The use of a small hand-held radiation detector in the operating room makes it possible to ensure the completeness of removal of medullary thyroid carcinoma metastases.

The use of somatostatin to inhibit the growth of normal and some neoplastic cell lines and to reduce the growth of experimentally induced tumors in animal models has stimulated interest in somatostatin analogues for the treatment of cancer. Somatostatin's tumoristatic effects may be a combination of direct actions on tumor cells related to inhibition of growth factor receptor expression, inhibition of MAPK, and stimulation of phosphorylase phosphatase. SSTR1, SSTR2, SSTR4, and SSTR5 can all promote cell cycle arrest associated with induction of the tumor suppressor retinoblastoma and p21, and SSTR3 can trigger apoptosis accompanied by induction of the tumor suppressor p53 and the proapoptotic protein Bax. In addition, somatostatin has indirect effects on tumor growth by its inhibition of circulating, paracrine, and autocrine tumor growth-promoting factors and it can modulate the activity of immune cells and influence tumor blood supply. Despite this promise, the therapeutic utility of octreotide as an antineoplastic agent remains controversial.

Two new treatment approaches in preclinical models may yet effectively utilize somatostatin receptors in the arrest of cancer cells. The first is receptor-targeted radionuclide therapy using octreotide chelated to a variety of gamma- or beta-emitting radioisotopes. Theoretical calculations and empirical data suggest that radiolabeled somatostatin analogues can deliver a tumoricidal radiotherapeutic dose to some tumors after receptor-mediated endocytosis. A variation on this theme is the chelation of a cytotoxic chemotherapeutic agent to a somatostatin analogue. A second approach involves somatic cell gene therapy to transfact SSTR-negative pancreatic cancer cells with an SSTR gene. Therapeutic results can be obtained with the creation of autocrine or paracrine inhibitory growth effects or the addition of targeted radionuclide treatments.
Prolactin-Regulating Factors

Dopamine

It is well known that PRL secretion, unlike the secretion of other pituitary hormones, is primarily under tonic inhibitory control by the hypothalamus. Destruction of the stalk median eminence or transplantation of the pituitary gland to ectopic sites causes a marked constitutive increase in PRL secretion, in contrast to a decrease in the release of GH, TSH, ACTH, and the gonadotropins. Many lines of evidence indicate that dopamine is the principal, physiologic prolactin-inhibiting factor (PIF) released from the hypothalamus. Dopamine is present in hypophysial-portal vessel blood in sufficient concentration to inhibit PRL release, and dopamine inhibits PRL secretion from lactotrophs both in vivo and in vitro. Mutant mice with a targeted disruption of the D2 receptor gene uniformly developed lactotroph hyperplasia, hyperprolactinemia, and eventually lactotroph adenomas, further emphasizing the importance of dopamine in the physiologic regulation of lactotroph proliferation in addition to hormone secretion.

The intrinsic dopamine neurons of the medial-basal hypothalamus constitute a dopaminergic population with regulatory properties that are distinct from those in other areas of the brain. Notably, they lack D2 autoreceptors but express PRL receptors, which are essential for positive feedback control as discussed in detail later. In the rat, these neurons are subdivided by location into the A12 group within the arcuate nucleus and the A14 group in the anterior periventricular nucleus. The caudal A12 dopamine neurons are further classified as tuberoinfundibular (TIDA) because of their axonal projections to the external zone of the median eminence. Tuberohypophyseal (THDA) neuronal soma are located more rostrally in the arcuate nucleus and project to both the neural lobe and intermediate lobe through axon collaterals that are found in the internal zone of the median eminence. Finally, the A14 periventricular hypophyseal (PHDA) neurons send their axons only to the intermediate lobe of the pituitary gland.

Although the TIDA neurons are generally considered to be the major source of dopamine to the anterior lobe through the long portal vessels originating in the median eminence, dopamine can also reach the anterior lobe from the neural and intermediate lobes by the interconnecting short portal veins. Consistent with this pathway for dopamine access to the anterior lobe, surgical removal of the neurointermediate lobe in rats caused a significant increase in basal PRL levels. In addition to direct actions of dopamine on lactotrophs, central dopamine can indirectly affect PRL secretion by altering the activity of inhibitory interneurons that in turn synapse on the TIDA neurons. These effects are complicated by opposing intracellular signaling pathways linked to D1 and D2 receptors located on different populations of interneurons.

The binding of dopamine or selective agonists such as bromocriptine to the D2 receptor has multiple effects on lactotroph function. D2 receptors are coupled to pertussis toxin-sensitive G proteins and inhibit adenyl cyclase and decrease intracellular cAMP levels. Other effects include activation of an inwardly rectifying K⁺ channel, increase of voltage-activated K⁺ currents, decrease of voltage-activated Ca²⁺ currents, and inhibition of inositol phosphate production. Together, this spectrum of intracellular signaling events decreases free Ca²⁺ concentrations and inhibits exocytosis of PRL secretory granules. Dopamine also has a modest effect on thyrotrophs to inhibit the secretion of TSH.

There is continuing debate concerning the mechanism by which D2 receptor activation inhibits transcription of the PRL gene. Likely pathways involve the inhibition of MAPK or protein kinase C, with a resultant reduction in the phosphorylation of Ets family transcription factors. Ets factors are important for the stimulatory responses of TRH, insulin, and epidermal growth factor on PRL expression and they interact cooperatively with the pituitary-specific POU protein Pit1, which is essential for cAMP-mediated PRL gene expression. The second messenger pathways used by the D2 receptor to inhibit lactotroph cell division are also unsettled. A study using primary pituitary cultures from rats demonstrated forskolin treatment, which activates protein kinase A and elevates intracellular cAMP, or insulin treatment, which activates a potent receptor tyrosine kinase, were both effective mitogenic stimuli for lactotrophs. Bromocriptine competitively antagonized the proliferative response caused by elevated cAMP. Furthermore, inhibition of MAPK signaling by PD98059 markedly suppressed the mitogenic action of both insulin and forskolin, suggesting an interaction of MAPK and protein kinase A signaling.

Another study used immortalized mammosomatotroph tumor cells that were transfected with a D2 receptor expression vector and concluded that stimulation of a phosphoinositide phosphatase activity was an important component of dopamine's antiproliferative action. Therefore, it is clear that dopamine actions on lactotrophs involve multiple different intracellular signaling pathways linked to activation of the D2 receptor, but different combinations of these pathways are relevant for the inhibitory effects on PRL secretion, PRL gene transcription, and lactotroph proliferation.

The other major action of dopamine in the pituitary is the inhibition of hormone secretion from the POMC-expressing cells of the intermediate lobe, although, as noted earlier, the adult human differs from most other mammals in the rudimentary nature of this lobe. THDA and PHDA axon terminals provide a dense plexus of synaptic-like contacts on melanotrophs. Dopamine release from these terminals is inversely correlated with serum MSH levels and also regulates POMC gene expression and melanotroph proliferation.

Other hypothalamic factors probably play a role secondary to that of dopamine as additional PIFs. The primary reason to conjecture the existence of these PIFs is the frequent inconsistency between portal dopamine levels and circulating PRL in different rat models. GABA is the strongest candidate and most likely acts through GABAₐ receptors in the anterior pituitary. Melanotrophs, like lactotrophs, are inhibited by both dopamine and GABA but with the principal involvement of GABAergic and metabotropic GABAₐ receptors. Because basal dopamine tone is high, the measurable inhibitory effects of GABA on PRL release are generally small under normal circumstances. Other putative PIFs include somatostatin and calcitonin.

Protein Releasing Factors

Although tonic suppression of PRL release by dopamine is the dominant effect of the hypothalamus on PRL secretion, a number of stimuli promote PRL release, not only...
merely by disinhibition of PIF effects but by causing release of one or more neurohormonal PRFs (see Fig. 7-35). The most important of the putative PRFs are TRH, oxytocin, and VIP, but vasopressin, angiotensin II, NPY, galanin, substance P, bombesin-like peptides, and neuropeptide C can also trigger PRL release under different physiologic circumstances. TRH was discussed in a previous section of this chapter (see Fig. 7-15). In humans there is an imperfect correlation between pulsatile PRL and TSH release, suggesting that TRH cannot be the sole physiologic PRF under basal conditions.

Like TRH, oxytocin, vasopressin, and VIP fulfill all the basic criteria for a PRF. They are produced in paraventricular hypothalamic neurons that project to the median eminence. Concentrations of the hormones in portal blood are much higher than in the peripheral circulation and are sufficient to stimulate PRL secretion in vitro. Moreover, there are functional receptors for each of the neurohormones in the anterior pituitary gland and either pharmacologic antagonism or passive immunization against each hormone can decrease PRL secretion, at least under certain circumstances. Vasopressin is released during stress and hypovolemic shock, as is PRL, suggesting a specific role for vasopressin as a PRF in these contexts. Similarly, another candidate PRF, peptide histidine isoleucine, may be specifically involved in the secretion of PRL in response to stress. Peptide histidine isoleucine and the human homologue PRP are structurally related to VIP and synthesized from the same prohormone precursor in their respective species. Both peptides are coexpressed with CRH in parvocellular paraventricular neurons and presumably released by the same stimuli that cause release of CRH into the hypophyseal-portal vessels.

There is evidence suggesting that dopamine itself may also act as a PRF, in contrast to its predominant function as a PIF. At concentrations three orders of magnitude lower than that associated with maximal inhibition of PRL secretion, dopamine was shown to be capable of stimulating secretion from primary cultures of rat pituitary cells. These cultures were extended to in vivo model by Arex and colleagues, who demonstrated that low-dose dopamine infusion in cannulated rats caused a further increase in circulating PRL above the already elevated baseline produced by pharmacologic blockade of endogenous dopamine biosynthesis. The physiologic relevance of these findings to humans has yet to be established.

Finally, reports of "new" PRFs continue to be published. Much excitement was generated by the isolation of a peptide from bovine hypothalamus named prolactin-releasing peptide (PRP). PRP binds with high affinity to an orphan G protein-coupled receptor (hGR3/GPR10) expressed specifically in human pituitary anterior lobe cells. PRP stimulates PRL release from rat pituitary cells with a potency similar to that of TRH. However, PRP is expressed predominantly in a subpopulation of noradrenergic neurons in the medulla and a serious question of whether PRP reaches the anterior pituitary and actually causes PRL secretion. Subsequent studies found no direct evidence for release of PRP in the arcuate nuclei/median eminence, further suggesting that the peptide is not a hypophyseotropic neurohormone. However, PRP probably does function as a neuromodulator within the hypothalamus.

CNS at sites receiving its receptor and may be involved in the neural circuitry mediating satiety.

Intrapituitary Regulation of Prolactin Secretion

Probably more than that of any other pituitary hormone, the secretion of PRL is regulated by autocrine-paracrine factors within the anterior lobe and by neurointermediate lobe factors that gain access to venous sinusoids of the anterior lobe by way of the short portal vessels. The wealth of local regulatory mechanisms within the anterior pituitary has been reviewed extensively and is also discussed in Chapter 8. Galanin, VIP, endothelin-like peptides, angiotensin II, epidermal growth factor, basic fibroblast growth factor, LHRH, and the cytokine IL-6 are among the most potent local stimulators of PRL secretion. Locally produced inhibitors include PRL itself, acetylcholine, transforming growth factor beta, and calcitonin. Although none of these stimulatory or inhibitory factors plays a dominant role in the regulation of lactotroph function and much of the research in this area has not been directly confirmed in human pituitary, it seems apparent that the local milieu of autocrine and paracrine factors plays an essential regulatory role in determining the responsiveness of lactotrophs to hypophysiotropic factors in different physiologic states.

As noted earlier, a proportion of the inhibitory dopamine tone to the anterior lobe lactotrophs is derived from the neurointermediate lobe. It was therefore unanticipated that surgical removal of this structure in rats would block suckling-induced PRL release over the moderate baseline increase attributed to partial dopamine disinhibition. Further studies showed that exposure of the anterior pituitary to intermediate lobectomy extracts (devoid of VIP, vasopressin, and other known PRFs) stimulated PRL secretion. At least two kinds of PRF activity have been isolated from intermediate lobe tumors of the mouse, but the specific molecules involved have yet to be identified. Other researchers have suggested a more passive role for the neurointermediate lobe in the regulation of PRL secretion. Melanotropin-derived N-acetylated MSH appears to act as a lactotroph responsiveness factor by recruiting nonsecretory cells to an active state and sensitizing secreting lactotrophs to the actions of other direct PRFs. However, the relevance of the neurointermediate lobe for PRL regulation in primates (including humans) is not clear because of its atrophied structure in these species.

Neuroendoctrine Regulation of Prolactin Secretion

Secretion of PRL, like that of other anterior pituitary hormones, is regulated by hormonal feedback and neural influences from the hypothalamus. Feedback is exerted by PRL itself at the level of the hypothalamus. PRL secretion is regulated by many physiologic states including the estrous and menstrual cycles, pregnancy, and lactation. Furthermore, PRL is stimulated by several extrereceptive stimuli including light, ultrasonic vocalization of pups, olfactory cues, and various modalities of stress. Expression and secretion of PRL are also influenced strongly by estrogens at the level of both the lactotrophs and TIDA neurons (see Fig. 7-35) and by paracrine regulators within the pituitary such as galanin and VIP.

Feedback Control

Negative feedback control of PRL secretion is mediated by a unique short-loop mechanism within the hypothalamus. PRL activates PRL receptors, which are expressed on all three subpopulations of A12 and A14 dopamine neurons, leading to increased tyrosine hydroxylase expression and dopamine synthesis and release. Ames dwarf mice that secrete virtually no PRL, GH, or TSH have decreased numbers of arcuate dopamine neurons and this hypoplasia can be reversed by neonatal administration of PRL, suggesting a trophic action on the neurons. However, another mouse model of isolated PRL deficiency generated by gene targeting appears to have normal numbers of hypofunctioning dopamine neurons secondary to the loss of PRL feedback.

Neural Control

Lactotrophs have spontaneously high secretory activity, and therefore the predominant effect of the hypothalamus on PRL secretion is tonic suppression, which is mediated by regulatory hormones synthesized by tuberohypophyseal neurons. Secretory bursts of PRL are caused by the acute withdrawal of dopamine inhibition, stimulation by PRFs, or combinations of both events. At any given moment, locally produced autocrine and paracrine regulators further modulate the responsiveness of individual lactotrophs to neurohormonal PIFs and PRFs.

Multiple neurotransmitter systems impinge on the hypothalamic dopamine and PRF neurons to regulate their neurosecretion (see Fig. 7-35). Nicotinic cholinergic and glutamatergic afferents activate TIDA neurons, whereas histamine, acting predominantly through H2 receptors, inhibits these neurons. An inhibitory peptidergic input to TIDA neurons of major physiologic significance is that associated with the endogenous opioid peptides enkephalin and dynorphin and their cognate - and -receptor subtypes. Opioid inhibition of dopamine release has been associated with increased PRL secretion under virtually all physiologic conditions, including the basal state, different phases of the estrous cycle, lactation, and stress.

Ascending serotoninergic inputs from the dorsal raphe nucleus are the major activator of PRF neurons in the paraventricular nucleus. There is still debate concerning the identity of the specific 5-hydroxytryptamine receptors involved in this activation.

The PRL regulatory system and its monoaminergic control have been scrutinized in detail because of the frequent occurrence of syndromes of PRL hypersecretion (see Chapter 8). Both the pituitary and the hypothalamus have dopamine receptors, and unfortunately the response to dopamine receptor stimulation and blockade does not distinguish between central and peripheral actions of the drug. Many commonly used neuroleptic drugs influence PRL secretion. Reserpine (a catecholamine depletor) and phenoxybenzamine such as chlorpromazine and haloperidol enhance PRL release by disinhibition of dopamine action on the pituitary, and the PRL response is an excellent predictor of the antipsychotic effects of phenothiazines because of its correlation with D2 receptor binding and activation.
antipsychotic neuroleptic agents act on brain dopamine receptors in the mesolimbic system and in the pituitary-regulating tuberoinfundibular system. Consequently, treatment of such patients with dopamine agonists such as bromocriptine can reverse the psychiatric benefits of such drugs. A report of three patients with psychosis and concomitant prolactinomas recommended the combination of clozapine and quinagolide as the treatment of choice to manage both diseases simultaneously.

Factors Influencing Secretion

Circadian Rhythm

PRL is detectable in plasma at all times during the day but is secreted in discrete pulses superimposed on basal secretion and exhibits a diurnal rhythm with peak values in the early morning. There is a true circadian rhythm in humans because it is maintained in a constant environment independent of the sleep rhythm. The combined body of data examining TIDA neuronal activity, dopamine concentrations in the median eminence, and manipulations of the SCN suggests that endogenous diurnal alterations in dopamine tone that are entrained by light constitute the major neuroendocrine mechanism underlying the circadian rhythm of PRL secretion.

External Stimuli

The suckling stimulus is the most important physiologic regulator of PRL secretion. Within 1 to 3 minutes of nipple stimulation, PRL levels rise and remain elevated for 10 to 20 minutes. This reflex is distinct from the milk let-down, which involves oxytocin release from the neurohypophysis and contraction of mammary alveolar myoepithelial cells. These reflexes provide a mechanism by which the infant regulates both the production and the delivery of milk. The nocturnal rise in PRL secretion in nursing women and in non-nursing women may have evolved as a mechanism of milk maintenance during prolonged nonsuckling periods at night.

Pathways involved in the suckling reflex arise in nerves innervating the nipple, enter the spinal cord by way of spinal afferent neurons, ascend the spinal cord through spinothalamic tracts to the midbrain, and enter the hypothalamus by way of the median forebrain bundle. In most of the pathway, neurons regulating the oxytocin-dependent milk let-down response accompany those involved in PRL regulation and then separate at the level of the paraventricular nuclei. The suckling reflex brings about an inhibition of PIF activity and a release of PRFs, although the identity of an undisputed suckling-induced PRF is unsettled.

Although the significance for PRL regulation in humans is not certain, environmental stimuli from seasonal changes in light duration and auditory and olfactory cues are clearly of great importance to many mammalian species. Seasonal breeders, such as the sheep, exhibit a reduction in PRL secretion in response to shortened days. The specific ultrasound vocalization of rodent pups is among the most potent stimuli for PRL secretion in lactating and virgin female rats. Olfactory stimuli from pheromones also have potent actions in rodents. A prime example is the Bruce effect or spontaneous abortion induced by exposure of a pregnant female rat to an unfamiliar male. It is mediated by a well-studied neural circuitry involving the vomeronasal nerves, corticomedial amygdala, medial preoptic area of the hypothalamus, and finally activation of TIDA neurons and a reduction in circulating PRL that is essential for maintenance of luteal function in the first half of pregnancy.

Stress in many forms dramatically affects PRL secretion, although the teleologic significance is uncertain. It may be related to actions of PRL on cells of the immune system or some other aspect of homeostasis. Different stressors are associated with either a reduction or an increase in PRL secretion, depending on the local regulatory environment at the time of the stress. However, whereas well-documented changes in PRL are associated with relatively severe forms of stress in laboratory animal models, a study of academic stress in college students failed to show any significant correlation among the time periods before, during, or after final examinations and diurnal PRL levels.
Gonadotropin-Releasing Hormone and Control of the Reproductive Axis

Chemistry and Evolution

The hypothalamic neuropeptide that controls the function of the reproductive axis is GnRH. GnRH is a 10-amino-acid peptide that is synthesized as part of a larger precursor molecule and is then enzymatically cleaved to remove a signal peptide from the N-terminus and GnRH-associated peptide (GAP) from the C-terminus (Fig. 7-36). All forms of the decapeptide have a pyroGlu at the N-terminus and Gly-amide at the C-terminus, indicating the functional importance of the terminal regions throughout evolutionary biology.

Within mammals, two genes encoding GnRH have been identified. The first encodes a 92-amino-acid precursor protein. This form of GnRH is now referred to as GnRH-I and is the form found in hypothalamic neurons that serves as a releasing factor to regulate pituitary gonadotropin function. The second GnRH gene, GnRH-II, encodes a decapetide that differs from the first by three amino acids. This form of GnRH is found in the midbrain region and serves as a neurotransmitter rather than as a pituitary releasing factor. Both GnRH-I and GnRH-II are found in phylogenetically diverse species, from fish to mammals, suggesting that these multiple forms of GnRH diverged from one another early in vertebrate evolution. A third form of GnRH, GnRH-III, has been identified in neurons of the telencephalon in teleost fish. This form of GnRH may have been lost in higher vertebrates or simply not yet discovered in other species.

As discussed subsequently in more detail, GnRH-I and GnRH-III are found in cells that originate in the olfactory placode in early embryonic development. In contrast, GnRH-I-containing cells are derived from the midbrain ventricle. GnRH is also found in cells outside the brain. The roles of GnRH peptides produced outside the brain are not well understood but are an area of current investigation.

All GnRH genes have the same basic structure, with the pre-prohormone mRNA encoded in four exons. Exon 1 contains the 5' untranslated region of the gene; exon 2 contains the signal peptide, GnRH, and the N-terminus of GAP; exon 3 contains the central portion of GAP; and exon 4 contains the C-terminus of GAP and the 3' untranslated region (see Fig. 7-36). Among species, the nucleotide sequences encoding the GnRH decapetide are highly homologous.

Two transcriptional start sites have been identified in GnRH genes at +1 and -579, with the +1 promoter being active in hypothalamic neurons and the other promoter active in placenta. The first 173 base pairs of the promoter are highly conserved among species. In the rat, this promoter region has been shown to contain two Oct-1 binding sites; three regions that bind the Pou domain family of transcription factors, SCIP, Oct-6, and Tst-1; and three regions that can bind the progesterone receptor. In addition, a variety of promoters and second messengers have been shown to regulate GnRH gene expression, and the majority of the cis-acting elements thus far characterized for hormonal control of GnRH transcription have been localized to the proximal promoter region. The 5' untranslated region of the GnRH gene also contains a 300-base-pair enhancer region that is 1.8 kilobases upstream of the transcription start site. It contains binding sites for Pou homeodomain transcription factors and GATA factors.

In this chapter, we focus on the hypothalamic GnRH that is derived from GnRH-I mRNA and plays an important role in the regulation of the hypothalamic-pituitary-gonadal axis. A mutant strain of mice with a deletion of the GnRH-I gene have hypogonadism, and the homozygous animals are infertile.

Anatomic Distribution

GnRH neurons are small, diffusely located cells that are not concentrated in a nucleus (Fig. 7-37A). They are generally bipolar and fusiform in shape, with long thin axons that can exhibit spines. The location of hypothalamic GnRH neurons is species-dependent. In the rat, hypothalamic GnRH neurons are concentrated in rostral areas including the medial preoptic area, the diagonal band of Broca, the septal areas, and the anterior hypothalamus. In primates, the majority of hypothalamic GnRH neurons are located more dorsally in the medial basal hypothalamus, the infundibulum, and periventricular to the third ventricle. Throughout the hypothalamus, neuroendocrine GnRH neurons, which extend their axon terminals to the median eminence, are interspersed with non-neuroendocrine GnRH neurons, which extend their axons to other regions of the brain including other hypothalamic regions and various regions of the cortex. GnRH secreted from non-neuroendocrine neurons has been implicated in the control of sexual behavior in rodents but not in higher primates.

Embryonic Development

GnRH neuronal development is an unusual neuronal population in that they originate outside the CNS, from the epithelial tissue of the nasal placode. During embryonic development GnRH neurons migrate across the surface of the brain and into the hypothalamus, with the final hypothalamic location differing somewhat among species. Migration is dependent on a scaffolding of neurons and glial cells along which the GnRH neurons move, with neural cell adhesion molecules playing a critical role in guiding the migration process.

Failure of GnRH neurons to migrate properly leads to a clinical condition, Kallman's syndrome, in which GnRH neuroendocrine neurons do not reach their final destination and thus do not stimulate pituitary gonadotropin secretion. Patients with Kallman's syndrome do not enter puberty spontaneously. X-linked Kallman's syndrome results from a deficiency of the KAL-1 gene, which encodes a putative protein of 680 amino acids and contains four fibronectin type III repeats and a four-disulfide core motif. However, this form of Kallman's syndrome accounts for only a small percentage (about 8%) of cases, and the cause of other forms remains unknown. Administration of exogenous GnRH effectively treats this form of hypothalamic hypogonadism. Patients with Kallman's syndrome often have other congenital midline defects, including anosmia, which results from hypoplasia of the olfactory bulb and tract.

Action at the Pituitary

Receptors

GnRH binds to a membrane receptor on pituitary gonadotrophs and stimulates both LH and FSH synthesis and secretion. The GnRH receptor is a seven-transmembrane-domain G protein-coupled receptor, but it lacks a typical intracellular C-terminal cytoplasmic domain. Under physiologic conditions, the GnRH receptor number varies and is usually directly correlated with the gonadotropin secretory capacity of pituitary gonadotrophs. For example, across the rat estrous cycle, a rise in GnRH receptors is seen just before the surge of gonadotropins that occurs on the afternoon of proestrus. GnRH receptor message levels are regulated by a variety of hormones and second messengers including steroid hormones (estradiol can both suppress and stimulate, and progesterone suppresses), gonadotropins (which suppress), and calcium and protein kinase C (which stimulate).

G_Na1 is the primary guanosine triphosphate-binding protein mediating GnRH responses; however, there is evidence that GnRH receptors can couple to other guanosine triphosphate-binding proteins including G_s and G_i. With activation, the GnRH receptor couples to a phosphoinoside-specific phospholipase C, which leads to increases in calcium transport into gonadotrophs and calcium release from internal stores through a diacylglycerol protein kinase C pathway. Increased calcium entry is a critical step in GnRH-stimulated release of gonadotropin secretion.

However, the MAPK cascade is also stimulated by GnRH.
When there is a decline in GnRH stimulation to the pituitary, as occurs in a variety of physiologic conditions including states of lactation, undemnurbation, or seasonal periods of reproductive quiescence, the number of GnRH receptors on pituitary gonadotrophs declines dramatically.\textsuperscript{2+} Subsequent exposure of the pituitary to pulses of GnRH restores receptor number by a Ca\textsuperscript{2+}-dependent mechanism that requires protein synthesis.\textsuperscript{3+} The effect of GnRH to induce its own receptor is termed up-regulation or self-priming. Only certain physiologic frequencies of pulsatile GnRH can augment GnRH receptor production, and these frequencies appear to differ among species.\textsuperscript{4+} Up-regulation of GnRH receptors after a period of low GnRH stimulation to the pituitary can take hours to days of exposure to pulsatile GnRH, depending on the duration and extent of the prior decrease in GnRH. The self-priming effect of GnRH to up-regulate its own receptors also plays a crucial role in the production of the gonadotropin surge that occurs at midcycle in females of spontaneously ovulating species and triggers ovulation. Just before the gonadotropin surge, two factors, the increased frequency of pulsatile GnRH release and a sensitization of the pituitary gonadotrophs by rising levels of estradiol, make the pituitary exquisitely sensitive to GnRH and allow an output of LH that is an order of magnitude greater than the release seen during the rest of the female reproductive cycle. This surge of LH triggers the ovulatory process at the ovary.

In contrast to up-regulation of GnRH receptors by pulsatile regimens of GnRH, continuous exposure to GnRH leads to down-regulation of GnRH receptors and an accompanying decrease in LH and FSH synthesis and secretion, termed desensitization.\textsuperscript{5+} Down-regulation does not require calcium mobilization or gonadotropin secretion.\textsuperscript{6+} It involves a rapid uncoupling of receptor from G proteins and sequestration of the receptors from the plasma membrane, followed by internalization and proteolytic degradation of the receptors.\textsuperscript{7+}

The concept of down-regulation has a number of clinical applications. For example, the most common current therapy for precocious puberty of hypothalamic origin (i.e., precocious GnRH secretion) is to treat the child with a long-acting

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure7-37a.png}
\caption{Regulation of the hypothalamic-pituitary-gonadal axis. A, Gonadotropin-releasing hormone (GnRH) neurons in a coronal section of the rat hypothalamus at 4× magnification. The inset is at 20× magnification. (Micrograph provided by Patricia Williamson and Kevin Grove, Oregon National Primate Center.)}
\end{figure}

GnRH agonist, which down-regulates pituitary GnRH receptors and effectively turns off the reproductive axis.\textsuperscript{8+} Children with precocious puberty can be maintained with long-acting GnRH agonists for years to suppress the premature activation of the reproductive axis, and at the normal age of puberty agonist treatment can be withdrawn, allowing a reactivation of pituitary gonadotrophs and a downstream increase in gonadal steroid hormone production. Long-acting GnRH agonists are also used in the treatment of forms of breast cancer that are estrogen-dependent as well as other gonadal steroid-dependent cancers.\textsuperscript{9+} Long-acting antagonists of GnRH have been developed that can also be used for these therapies.\textsuperscript{10} Antagonists have the advantage of not having a flare effect, that is, an acute stimulation of gonadotropin secretion that is seen during the initial treatment of individuals with superagonists.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure7-37b.png}
\caption{A, Schematic diagram of the hypothalamic-pituitary-gonadal axis showing neural systems that regulate GnRH secretion and feedback of gonadal steroid hormones at the level of the hypothalamus and pituitary: CRH, corticotropin-releasing hormone; FSH, follicle-stimulating hormone; GABA, -aminobutyric acid; LH, luteinizing hormone; NPY, neuropeptide Y. Frequent blood samples from the peripheral blood stream is used to define the pulsatile nature of LH secretion (i.e., frequency and amplitude of LH pulses), and pulsatile LH is used as an indirect measure of the activity of the GnRH secretory system. Indirect measurement of GnRH secretion is thus used in many animal studies examining the factors that govern the regulation of the pulsatile activity of the reproductive neuroendocrine axis. Unlike}
\end{figure}

Pulsatile Gonadotropin-Releasing Hormone Stimulation

Because a single pulse of GnRH stimulates the release of both LH and FSH and chronic exposure of the pituitary to pulsatile GnRH supports the synthesis of both LH and FSH, many people believe that there is only one releasing factor regulating the synthesis and secretion of LH and FSH.\textsuperscript{11+} However, in a number of physiologic conditions there are divergent patterns of LH and FSH secretion, and thus a second FSH-releasing peptide has been proposed, but such a peptide has not been isolated to date.\textsuperscript{12+} Other mechanisms, discussed in more detail later, are likely to account for the differential regulation of LH and FSH release.

The ensemble of GnRH neurons in the hypothalamus that send axons to the portal blood system in the median eminence fire in a coordinated, repetitive, episodic manner, producing distinct pulses of GnRH in the portal blood stream.\textsuperscript{13} The pulsatile nature of GnRH stimulation to the pituitary leads to the release of distinct pulses of LH into the peripheral blood stream.\textsuperscript{14} In experimental animals, in which it is possible to collect blood samples simultaneously from the portal and peripheral blood stream, GnRH and LH pulses have been found to correspond in about a one-to-one ratio at most physiologic rates of secretion (Fig. 7-38) (Figure Not Available).\textsuperscript{15} Because the portal blood stream is generally inaccessible in humans, the collection of

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure7-37c.png}
\caption{Simultaneous detection of pulses of gonadotropin-releasing hormone (GnRH) measured in blood collected from the hypothalamic-hypophyseal portal vessels and luteinizing hormone (LH) measured in blood collected from the peripheral vasculature of an oophorectomized ewe. (Redrawn from Clarke L, Cummins JT. Endocrinology 1992; 111:1737-1739.)}
\end{figure}

LH secretion, FSH secretion is not always pulsatile, and even when it is pulsatile, there is only partial concordance between LH and FSH pulses.\textsuperscript{16+}

It is possible to place multiple unit recording electrodes in the medial basal hypothalamus of monkeys and other species and find spikes of electrical activity that are concordant with the pulsatile discharge of LH secretion.\textsuperscript{17} It is unknown, however, whether these bursts of electrical activity reflect the activity of GnRH neurons themselves or the activity of neurons that impinge on GnRH neurons and govern their firing. With the development of mice in which the gene for green fluorescent protein has been under the regulation of the GnRH promoter, it has been possible to identify GnRH neurons in hypothalamic tissue slices using fluorescence microscopy and record from them intracellularly.\textsuperscript{18} These studies

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure7-37d.png}
\caption{The influence of gonadotropin-releasing hormone (GnRH) pulse frequency on luteinizing hormone (LH) and follicle-stimulating hormone (FSH) secretion in a female mule monkey with an acute nucleus lesion abating endogenous GnRH support of the pituitary. Decreasing GnRH pulse frequency from 1 pulse/hour to 1 pulse/3 hours leads to a decrease in plasma LH concentrations but an increase in plasma FSH concentrations. (Redrawn from Wilt L, Haueter A, Marshall G, et al. Endocrinology 1987; 109:376385.)}
\end{figure}

have shown that many, but not all, GnRH neurons show a bursting pattern of electrical activity. A central, unsolved question in the field of reproductive neuroendocrinology is what causes GnRH neurons to pulse in a coordinated manner. Studies using a line of clonal GnRH neurons have shown that these neurons grown in culture can release GnRH in a pulsatile pattern, suggesting that the pulse-generating capacity of GnRH neurons may be intrinsic.\textsuperscript{19} The term GnRH pulse generator is often used to acknowledge the fact that GnRH secretion occurs in pulses and to refer to the central mechanisms responsible for pulsatile GnRH
A critical factor governing LH and FSH secretion and release is the rate of pulsatile GnRH stimulation of the gonadotropes. Experimental studies in which the hypothalamus was lesioned and GnRH was replaced by pulsatile administration of exogenous GnRH showed that different frequencies of GnRH can lead to differential ratios of LH to FSH secretion from the pituitary. Figure 7-39 shows that in a monkey with a hypothalamic lesion, replacement of one pulse of GnRH per hour led to a relatively low ratio of FSH to LH secretion. Subsequent institution of a slower pulse frequency of one pulse of GnRH every 3 hours led to a decrease in LH secretion but an increase in FSH secretion such that the ratio of FSH to LH secretion was greatly elevated. It is likely that this effect of pulse frequency on the ratio of FSH to LH secretion accounts, at least in part, for the clinical finding that at times when the GnRH pulse generator is just turning on, such as at the onset of puberty and during recovery from chronic undernutrition, the ratio of FSH to LH is higher than when it is measured in adults experiencing regular reproductive function. As discussed subsequently, steroid hormones act at both the hypothalamus and pituitary to influence strongly the rate of pulsatile GnRH release and amount of LH and FSH secreted from the pituitary.

GnRH pulse frequency not only influences the rate of pulsatile gonadotropin release and the ratio of FSH to LH secretion but also plays an important role in modulating the structural makeup of the gonadotropins. LH and FSH are structurally similar glycoprotein hormones. Each of these hormones is made up of an α subunit and a β subunit, and each has a unique subunit that conveys tissue specificity to the intact hormone. Before secretion of gonadotropins, terms such as LH and FSH are attached to their respective gonadotropin molecule. The sugars include sialic acid, galactose, α-acetylgalactosamine, and mannose, but the most important is sialic acid. The extent of glycosylation of LH and FSH is important for the physiologic function of these hormones. Forms of gonadotropin with more sialic acid have a longer half-life because they are protected from degradation by the liver. Forms of gonadotropin with less sialic acid can have more potent effects at their biologic receptors. Both the rate of GnRH stimulation and ovarian hormone feedback at the level of the pituitary regulate the degree of LH and FSH glycosylation. For example, slow frequencies of GnRH, seen during follicular development, are associated with greater degrees of FSH glycosylation, which would provide sustained FSH support to growing follicles. In contrast, faster frequencies of GnRH, seen just before the midcycle gonadotropin surge, are associated with lesser degrees of FSH glycosylation, providing a more potent but shorter lasting form of FSH at the time of ovulation.

Regulatory Systems

Many neurotransmitter systems from the brain stem, limbic system, and other areas of the hypothalamic convey information to GnRH neurons (see Fig. 7-37; Fig. 7-38). These systems include suprachiasmatic, serotonergic, GABA-ergic, and a number of other peptide neurotransmitters. Glutamate and norepinephrine play important roles in providing stimulatory drive to the reproductive axis, whereas GABA and endogenous opioid peptides provide a substantial portion of the inhibitory drive to GnRH neurons. Influences of specific neurotransmitter systems are discussed where appropriate in later sections on the physiologic regulation of GnRH neurons.

GnRH neurons are surrounded by glial processes, and only a small percentage of their surface area is available to receive dendritic contacts fromafferent neurons. Changes in the steroid hormone milieu influence the degree of glial sheathing and may play important roles in regulating afferent input to GnRH neurons by this mechanism. Some glial cells also secrete substances that can modulate the activity of GnRH neurons. For example, alpupherty there is an increase in the hypothalamic expression of transforming growth factor, and transforming growth factor can stimulate GnRH release by acting on astroglial cells to stimulate their release of PGE₂, which is stimulatory to GnRH neurons.

Feedback Regulation

Steroid hormone receptors are abundant in the hypothalamus and in many neural systems that impinge on GnRH neurons, including noradrenergic, serotonergic, -endorphincontaining, and NPY neurons. Early studies identifying regions of the brain that bound labeled estrogens showed that in rodents the preoptic area and ventromedial hypothalamus had the highest concentrations of estrogen receptors in the brain. Further localization studies, identifying estrogen receptors by immunocytochemistry or in situ hybridization, confirmed the strong presence of estrogen receptors in the hypothalamus and in brain areas with strong connections to the hypothalamus, including the amygdala, septal nuclei, bed nucleus of the stria terminalis, medial part of the nucleus of the solitary tract, and lateral portion of the parabrachial nucleus. In 1986 a new member of the steroid hormone receptor superfamily with high sequence homology to the classical estrogen receptor (now referred to as estrogen receptor ) was isolated from rat prostate and named estrogen receptor . This novel estrogen receptor was shown to bind estradiol and to activate transcription by binding to estrogen response elements.

In situ hybridization studies examining the localization of estrogen receptor mRNA have shown that these receptors are present throughout the rostral-caudal extent of the brain, with a high level of expression in the preoptic area, bed nucleus of the stria terminalis, paraventricular and supraoptic nuclei, amygdala, and laminae II to VI of the cerebral cortex. Specific receptors for progesterone are induced by estrogen in hypothalamic regions of the brain, including the preoptic area, the ventromedial and ventrolateral nuclei, and the infundibular-arcuate nucleus, although there is also evidence for constitutive expression of progesterone receptors in some hypothalamic regions. Androgen receptor mRNA is found in the distribution of androgen and estrogen receptors throughout the brain. The highest density of androgen receptors was found in hypothalamic nuclei known to participate in the control of reproduction and sexual behaviors, including the arcuate nucleus, paraventricular nucleus, medial preoptic nucleus, ventromedial nucleus, and brain regions with strong connections to the hypothalamus including the amygdala, nuclei of the septal region, bed nucleus of the stria terminalis, nucleus of the solitary tract, and lateral division of the parabrachial nucleus.

The anterior pituitary also contains receptors for all of the gonadal steroid hormones. Steroid hormones can dramatically alter the pattern of pulsatile release of GnRH and of the gonadotropins through actions at both the hypothalamus and the pituitary (Fig. 7-40 and Fig. 7-41). At the hypothalamus, estradiol, progesterone, and testosterone can all act to slow the frequency of GnRH release into the portal blood stream, an action referred to as negative feedback. Because GnRH neurons have generally been shown to lack steroid hormone receptors, it is likely that the effects of steroid hormones on the firing rate of GnRH neurons are mediated by steroid hormone actions on other neural systems that provide afferent input to GnRH neurons. For example, progesterone-mediated negative feedback on GnRH secretion in primates appears to be regulated by -endorphincontaining neurons in the hypothalamus, acting primarily through μ-opioid receptors. If a μ-opioid antagonist, such as naloxone, is administered along with progesterone, the negative feedback action of progesterone on GnRH secretion can be blocked.

Negative feedback of steroid hormones can also occur directly at the level of the pituitary. For example, estradiol has been shown to be capable of binding to the pituitary, decreasing LH and FSH synthesis and release, and decreasing the sensitivity of pituitary gonadotropes to the actions of GnRH such that less LH and FSH are released when a pulse of GnRH stimulates the pituitary. Evidence for such a direct pituitary action of estradiol came from studies with rhesus monkeys that had been rendered deficient in endogenous GnRH by a lesion in the arcuate nucleus and showed a decline in endogenous gonadotropin secretion. When these monkeys received exogenous GnRH gonadotropin secretion, subsequent estradiol infusions dramatically suppressed the responsiveness of the pituitary to GnRH and suppressed the gonadotropin secretion that was being driven by the pulsatile administration of GnRH. Steroid hormones can have direct negative feedback actions at the pituitary; however, the extent of hypothalamic versus pituitary negative feedback actions is species-specific. In primate species including humans, there is considerable feedback of estradiol at the pituitary, but most of the progesterone and testosterone negative feedback occurs at the level of the hypothalamus.

Most of the time, the hypothalamic-pituitary axis is under the negative feedback influence of gonadal steroid hormones. If the gonads are removed surgically or their normal secretory function of steroid hormones is suppressed pharmacologically, there is a dramatic increase (10-fold to 20-fold) in circulating levels of LH and FSH secretion. This type of "castration response" occurs normally at the menopause in women, when...
Regulation of the Ovarian Cycle

Ovarian follicular development and thus ovarian production of large quantities of estradiol and progesterone decrease and eventually cease.

In addition to negative feedback, estradiol can have a positive feedback action at the level of the hypothalamus and pituitary to lead to a massive release of LH and FSH from the pituitary.

This massive release of gonadotropins occurs once each menstrual cycle and is referred to as the LH-FSH surge. The positive feedback action of estradiol acts as a response to the rising tide of estradiol that is produced during the process of dominant follicle development in the late follicular phase of the menstrual cycle. In women, elevated estradiol levels are generally maintained at about 500 pg/mL for about 36 hours prior to stimulation of the gonadotropin surge.

Experiments have shown that both a critical concentration of plasma estradiol and a critical duration of elevated estradiol are necessary to achieve positive feedback and a resulting gonadotropin surge (see Fig. 7-41). Moreover, the duration of estradiol elevation that is required to trigger a surge depends on the concentration of estradiol. If supra-physiological doses of estradiol are administered, the surge can occur as early as 18 hours after their administration. Because the ovary is responsible for the production of estradiol and the time course and magnitude of estradiol release control the rate of positive feedback, the ovary has been referred to as the zeitgeber of the menstrual cycle. The dependence of the positive feedback system on the magnitude of estradiol production helps explain the fact that the portion of the menstrual cycle that varies most in length is the follicular phase. Production of higher levels of estradiol by a dominant follicle in one cycle would lead to a more rapid positive feedback action with earlier ovulation and thus a shorter follicular phase compared with a cycle in which the dominant follicle produced lower levels of estradiol.

As with negative feedback in response to estradiol, the positive feedback actions of estradiol occur both at the hypothalamus, to increase GnRH secretion, and at the pituitary, to enhance greatly pituitary responsiveness to GnRH. At the pituitary, estradiol increases pituitary sensitivity to GnRH by increasing the synthesis of new GnRH receptors and by enhancing the responsiveness to GnRH at a postreceptor site of action. At the level of the hypothalamus in rodent species, estradiol appears to act at a “surge center” to induce the ovulatory surge of GnRH. Lesions in areas adjacent to the medial preoptic area, near the anterior commissure and sepal complexes, block the ability of estradiol to induce a surge in these species without blocking negative feedback effects of estradiol. In primate species, there does not appear to be a separate surge center mediating the positive feedback actions of estradiol. The cellular mechanisms that mediate the switch from negative to positive feedback of estradiol are not fully understood, but there is support for the concept that estrogen induction of various transcription factors and receptors (notably progesterone receptors) may play an important role in mediating this switch. Alternatively, estrogen has been shown to have biphasic actions on hypothalamic GABAergic neurons that impinge on GnRH neurons and are strong regulators of their activity, with the switch in action dependent on the duration of estradiol exposure. The molecular mechanisms by which estradiol influences GnRH gene expression are also not well understood, but it is likely that these influences occur through actions of neural systems afferent to GnRH because GnRH neurons do not appear to have estrogen receptors. Much more is known about the molecular mechanisms by which estradiol acts at the pituitary to regulate gonadotropin gene expression. Expression of LH subunit is strongly regulated (10-fold to 14-fold) by estradiol, but expression of FSH and subunits is regulated to a lesser extent (4-fold to 8-fold and 2-fold to 3-fold, respectively). Although in vivo studies indicate strong negative feedback actions of estradiol on LH gene transcription, such actions have not been replicated in in vitro studies with isolated pituitaries, leading to the conclusion that estradiol negative feedback on gonadotropin synthesis occurs predominantly by extrapituitary mechanisms. In contrast, estradiol can stimulate LH mRNA transcription directly at the level of the pituitary, acting by binding to an estrogen response element in the 5’ promoter region of the LH gene.

Regulation by Inhibins and Activins

Negative feedback of pituitary FSH secretion is also exerted by a family of peptide hormones produced by the gonads, the inhibins. Inhibins are produced by follicular and luteal cells of the ovary and by Sertoli cells in the testes. Inhibins are members of the transforming growth factor superfamily and comprise two subunits, an and a subunit. There are two forms of the subunit, A and B. Inhibins selectively inhibit the release of LH and FSH from gonadotrophs. Activins selectively suppress FSH secretion without simultaneous suppression of the LH secretion; thus, they provide one of the mechanisms whereby the pituitary can release differential amounts of LH and FSH, even though there appears to be only a single GnRH receptor. Alternatively, estrogen has been shown to have biphasic actions on hypothalamic GABAergic neurons that impinge on GnRH neurons and are strong regulators of their activity, with the switch in action dependent on the duration of estradiol exposure.

Activins received their name from their ability to facilitate FSH release. Activins have been shown to stimulate both basal and GnRH-induced FSH release from the anterior pituitary as well as increase FSH mRNA levels by enhancing transcription. An important role of endogenous activins in stimulating FSH secretion is supported by the finding that transgenic mice deficient in activin receptor IIa have reduced serum FSH levels. Activins have other actions in pituitary gonadotrophs as well, including up-regulation of GnRH receptors and enhancement of GnRH-stimulated LH release. Activins and inhibins also have local actions within the ovary influencing granulosa cell growth and differentiation, the responsiveness of the ovary to gonadotropins, steroid hormone production, follicular development, and oocyte maturation.

Regulation of the Ovarian Cycle

Whereas in males spermatogenesis occurs continually throughout the adult years, females show a cyclic pattern of ovarian activity with intermittent maturation and release of ova from the ovaries. Cyclic activity in the ovary is controlled by an interplay between steroid hormones produced by the ovary and the hypothalamic-pituitary neuroendocrine components of the reproductive axis. The duration of each phase of the ovarian cycle is species-dependent, but the general mechanisms controlling the cycle are similar in all species that have spontaneous ovarian cycles. In the human menstrual cycle, day 1 of the cycle is designated as the first day of menstrual bleeding. At this time, small and medium-sized follicles are present in the ovaries and only small amounts of estradiol are produced by the follicular cells. As a result, there is a low level of negative feedback to the hypothalamic-pituitary axis, LH pulse frequency is relatively fast (one pulse about every 60 minutes), and FSH concentrations are slightly elevated compared with much of the rest of the cycle. FSH acts at the level of the ovarian follicles to stimulate development and causes an increase in follicular estradiol production, which in turn provides increased negative feedback to the hypothalamic-pituitary unit.

A result of the increased negative feedback is a slowing of pulsatile LH secretion over the course of the follicular phase to a rate of about one pulse every 90 minutes. However, as the growing follicle (or follicles, depending on the species) secretes more estradiol, a positive feedback action of estradiol is triggered that leads to an increase in GnRH release and a surge release of LH and FSH. The surge of gonadotropins acts at the fully developed follicle to stimulate the dissolution of the follicular wall and leads to ovulation of the matured ovum into the nearby fallopian tube, where fertilization takes place if sperm are present.
after about 14 days and progesterone and estradiol secretion diminishes. This reduces the negative feedback signals to the hypothalamus and pituitary and allows an increase in FSH and LH secretion. The fall in progesterone is also a withdrawal of steroid hormone support to the endometrial lining of the uterus, and as a result the endometrium is shed as menses and a new cycle begins.

In other species, the interplay between the neuroendocrine and ovarian hormones is similar but the timing of events is different and other factors, such as circadian and seasonal regulatory factors, play a role in regulating the cycle. The rat has a 4- or 5-day ovarian cycle with no menses (the endometrial lining is absorbed rather than shed). The rat also shows strong circadian rhythmicity in the timing of the LH-FSH surge, with the surge always occurring in the afternoon of the day of proestrus. Sheep are an example of a species that has a strongly seasonal pattern of ovarian cyclicity. During the breeding season they have 15-day cycles, with a very short follicular phase and an extended luteal phase; during the nonbreeding season signals relay information about day length through the visual system, pineal, and SCN cause a dramatic suppression of GnRH neuronal activity, and cyclic ovarian function is prevented by a decrease in trophic hormonal support from the pituitary.

Early Development and Puberty

Neuroendocrine stimulation of the reproductive axis is initiated during fetal development, and in primates in midgestation circulating levels of LH and FSH reach values similar to those in castrated adults. Later in gestational development, gonadotropin levels decline, restrained by rising levels of circulating gonadal steroids. The steroids that have this effect are probably placental in origin in that after parturition there is a rise in circulating gonadotropin levels that is apparent for variable periods of the first year of life, depending on the species. The decline in reproductive hormone secretion in the postnatal period appears to be due to a decrease in GnRH stimulation of the reproductive axis because it occurs even in the castrate state and gonadotropin and gonadal steroid secretion can be supported by administration of pulses of GnRH.

Pubertal reawakening of the reproductive axis occurs in late childhood and is marked initially by nighttime elevations in gonadotropin and gonadal steroid hormone levels. The mechanisms controlling the pubertal reawakening of the GnRH pulse generator have been an area of intense investigation for the past two decades. Although the mechanisms are not fully understood, significant progress has been made in identifying central changes in the hypothalamus that appear to play a role in this process. There appear to be both a decrease in transsynaptic inhibition to the GnRH neuronal system at puberty and an increase in stimulatory input to GnRH neurons at this time. One of the major inhibitory inputs to the GnRH system is provided by GABAergic neurons. Studies in rhesus monkeys have shown that hypothalamic levels of GABA decrease early in puberty and that blocking GABAergic input before puberty, by intrahypothalamic administration of antiseizure oligodeoxynucleotides against the enzymes responsible for GABA synthesis, results in premature activation of the GnRH neuronal system.

It has been suggested, on the basis of findings that a subset of glutamate receptors (i.e., kainate receptors) increase in the hypothalamus at puberty, that the pubertal decrease in GABA tone may be caused by an increase in glutamatergic transmission. Further evidence for a role for glutamate comes from studies showing that administration of glutamate to prepubertal rhesus monkeys can drive the reawakening of the reproductive axis. Increased stimulatory drive to the GnRH neuronal system also appears to come from increases in norepinephrine and NPY at the time of puberty. Furthermore, as discussed earlier, there is evidence that growth factors act through release of prostaglandin from glial cells at puberty to play a role in stimulating GnRH neurons.

Despite an increased understanding of the neural changes occurring at puberty, the question of what signals trigger the pubertal awakening of the reproductive axis is unanswered at this time. Availability of food and nutritional status have been shown to affect the timing of puberty, but these signals appear to be only modulators of the pubertal process because puberty can be only moderately advanced by increasing food availability. Determining whether there is a genetic timing mechanism that regulates the timing of puberty or whether other signals from the body or the brain are responsible for timing the reactivation of the reproductive axis awaits further research.

Reproductive Function and Stress

Many forms of physical stresses, such as energy restriction, exercise, temperature stress, infection, pain, and injury, as well as psychological stresses, such as being subordinate in a dominance hierarchy or being acutely psychologically stressed, can suppress the activity of the reproductive axis. If the stress exposure is brief, there may be acute suppression of circulating gonadotropins and gonadal steroid hormones and in females disruption of normal menstrual cyclicity, but fertility is unlikely to be impaired. In contrast, prolonged periods of significant stress exposure can lead to complete impairment of reproductive function, also characterized by low circulating levels of gonadotropins and gonadal steroids. Stress appears to decrease the activity of the reproductive axis by decreasing GnRH drive to the pituitary because in all cases in which it has been examined, administration of exogenous GnRH can reverse the effects of the stress-induced decline in reproductive hormone secretion. Although we do not know the neural circuits through which many forms of stress suppress GnRH neuronal activity, some forms of stress-induced suppression of reproductive function are better understood.

In the case of foot shock stress in rats and immune stress (i.e., injection of IL-1) in primates, the suppression of gonadotropin secretion that occurs has been shown to be reversible by administration of a CRH antagonist, implying that endogenous CRH secretion mediates the effects of these stresses on GnRH neurons. In other studies, naloxone, a -opioid receptor antagonist, has been shown to be capable of reversing stress-induced suppression of gonadotropin secretion in monkeys; however, naloxone is ineffective in reversing the suppression of gonadotropin secretion that occurs during insulin-induced hypoglycemia. In the case of metabolic stresses, multiple regulators appear to mediate changes in the neural drive to the reproductive axis.

Various metabolic fuels including glucose and fatty acids can regulate the function of the reproductive axis, and blocking cellular utilization of these fuels can lead to suppression of gonadotropin secretion and decreased gonadal activity. Leptin, a hormone produced by fat cells, can also modulate the activity of the reproductive axis. Transgenic mice deficient in leptin or leptin receptors are infertile, and fertility can be restored by administration of leptin. Moreover, leptin administration has been shown to reverse the suppressive effects of undernutrition on the reproductive axis in some situations. Leptin receptors are found in several populations that are known to have a strong influence on the reproductive axis, notably NPY neurons.

In summary, it appears that a number of neural circuits can mediate effects of stress on the GnRH neuronal system and that the neural systems involved are at least somewhat specific to the type of stress that is experienced.
Leptin and the Brain-Gut-Adipose Axis

Long-term energy, stored as fat in adipose tissue, is homeostatically maintained by a hypothalamic system termed the lipostat or adipostat. Inputs to this system are many, including acute hormonal, nutritional, and vagal signals of hunger and satiety; the signal of long-term energy stores derived from the adipocyte hormone leptin; and powerful olfactory, visual, emotional, and cognitive inputs from higher brain centers. Outputs include those directed toward energy intake, primarily determined by feeding behavior, and energy expenditure, which can be broken down into basal metabolism, voluntary and involuntary activity, and diet-induced thermogenesis. As described throughout this chapter, energy homeostasis is maintained through the triad of behavioral, autonomic, and endocrine pathways. Thus, energy homeostasis is maintained by a complicated hypothalamic-brain stem-target organ axis that may be referred to as the brain-gut-adipose axis (Fig. 7-43).

The regulation of leptin levels is complex. For example, the hypothalamic system that regulates leptin release and action is under the influence of a variety of factors such as acute hormonal, nutritional, and vagal signals of hunger and satiety. Leptin is secreted primarily from the adipocyte; however, minor levels of regulated leptin expression also occur in other sites such as skeletal muscle, liver, heart, placenta, and stomach.

Effects of Leptin on the Hypothalamus and Neuropeptide Axes

A reduction in leptin levels occurs because of loss of adipose mass such as in anorexia nervosa, weight loss induced by diet or exercise, or starvation and is crucial to metabolic adaptation to a state of negative energy balance. This metabolic adaptation includes a decrease in metabolic rate that allows extended survival periods; inhibition of the reproductive, GH, and thyroid axes; and, at least in rodents, inhibition of the activity of the sympathetic nervous system and activation of the HPA axis.

In addition, leptin is a critical signal in the initiation of puberty. Leptin is a signal from the adipocyte tissue directed to the CNS that conveys readiness to proceed into puberty and is essential for fertility in the adult. Presumably, the leptin signal is a mechanism for the organism to determine whether adequate energy stores are present to maintain a pregnancy through term. For example, leptin administration restored fertility to db/db mice and prevented the starvation-induced delay in ovulation in female mice and rats.

Mechanism of Action

After secretion, leptin circulates in plasma in both free and bound forms. It is assumed that the binding protein is a soluble form of the leptin receptor, but other alternatives are being evaluated. In humans, the half-life of leptin is approximately 75 minutes.

The precise mechanism of the transport of leptin into the CNS is unknown. Active uptake of leptin has been described in the capillary endothelium and microvascularature of brains from humans and mice, suggesting a role of short isoforms of the leptin receptor. In addition, the transport of leptin into the choroid plexus is saturable.

After its transport through the blood-brain barrier, leptin binds to specific receptors in the hypothalamus. Leptin receptor mRNA is densely concentrated in the arcuate nucleus, and lower levels are found in the ventromedial and dorsomedial hypothalamic nuclei.

The leptin receptor is a member of the cytokine receptor superfamily. The leptin receptor binds Janus kinases (JAKs), tyrosine kinases involved in intracellular
signal transducer and activator of transcription (STAT) family of proteins. In turn, these STAT proteins activate transcription of leptin target genes.

A great deal of effort has gone into characterizing the mechanism of leptin action in the hypothalamus. Leptin appears to inhibit feeding and stimulate metabolism by acting on a small number of nuclei in the hypothalamus and brain stem, including the ventromedial hypothalamus and the arcuate, dorsomedial, and paraventricular hypothalamic nuclei. In these neurons, leptin up-regulates the expression of an assortment of anorexigenic peptides, such as -MSH, derived from the POMC pro-hormone gene (Fig. 7-46), cocaine and amphetamine-regulated transcript (CART), and neuropeptide Y, and decreases the expression of orexigenic peptides, NPY, AgRP, and melanin-concentrating hormone. Furthermore, leptin has been shown to depolarize and activate the firing rate of the anorexigenic POMC neurons and hyperpolarize the adjacent orexigenic NPY-ArgP neurons.

NPY, long known as a potent stimulator of feeding, was proved to have a role in leptin action when deletion of the NPY gene relieved a significant component of the obesity phenotype of the ob/ob mouse. The role of the NPY system in energy homeostasis was originally discovered from studies on the agouti mouse, one of the five naturally occurring monogenic obesity strains in the mouse. Intracerebroventricular administration of -MSH agonists and antagonist analogues inhibited stimulated feeding behavior, respectively. Furthermore, deletion of the melanocortin-4 receptor (MC4R), the primary neuronal receptor for the melanocortin peptides, caused an obesity syndrome identical to that seen in the obese leathal yellow agouti animal. The structure and distribution of the POMC and NPY-ArgP circuits are highly conserved in humans (Fig. 7-47) (Fig. Not Available). Furthermore, a null mutation in the POMC gene in humans caused an obesity syndrome similar to that seen in the leathal yellow mouse along with ACTH insufficiency and red hair (see

"Neuroendocrine Disease"). Thus, it appears that the central melanocortin system subserves the same purposes in mouse and humans. Although obesity related to genetic defects in leptin or the leptin receptor is rare in humans, haploinsufficiency of the MC4R appears to be responsible for up to 3% to 5% of severe pediatric obesity. Thus, the arcuate nucleus, site of the POMC-CART and NPY-ArgP neurons described earlier, is an important site of leptin action in humans.

Clinical Applications

Leptin deficiency and leptin receptor defects in humans are rare. In fact, serum leptin levels in humans are generally proportional to adipose mass. Thus, the vast majority of obese humans may be considered to manifest a leptin-resistant state rather than a deficient state. This concept of leptin resistance also remains poorly understood. However, it is thought that one mechanism of leptin resistance may be impaired leptin transport into the brain. Thus, suboptimal leptin transport through the blood-brain barrier may be one mechanism that underlies the development of leptin resistance in humans. Furthermore, the concept of leptin resistance leads to some reservations concerning the ability of exogenously administered leptin to overcome this leptin resistance and cause effective weight reduction in obese humans.

Clinical studies have now demonstrated that leptin treatment is safe and well tolerated and clearly effective in individuals with congenital leptin deficiency. In this study, low doses of methionyl leptin (met-leptin) were given to subjects with congenital deficiency that resulted in leptin levels 10% of that predicted on the basis of body fat. Leptin in this study was well tolerated and resulted in dramatic declines in appetite, body weight, and food intake. However, in individuals with common obesity, leptin had only modest effects on appetite and body weight. For example, studies evaluated the safety and efficacy of recombinant human met-leptin administration as well as pegylated human leptin. The first study was a double-blind, placebo-controlled, escalating-dose cohort trial in 54 lean and 73 obese subjects. Higher doses of met-leptin (0.01 to 0.3 mg/kg daily) were also given for 4 to 24 weeks. Met-leptin treatment resulted in significant dose-dependent weight loss: -1.3 kg (placebo...
Figure 7-48 Average plasma ghrelin, insulin, and leptin concentrations during a 24-hour period in 10 human subjects consuming breakfast (B), lunch (L), and dinner (D) at the times indicated (0800, 1200, and 1730 hours, respectively) (Reprinted with permission from Cummings DE, et al. A preprandial rise in plasma ghrelin levels suggest a role in meal initiation in humans. Diabetes 2001; 50:17141719.).

Of note, 95% of the weight loss achieved in the two highest dose cohorts was due to loss of fat mass and not any significant changes in fat free mass. The findings have supported the idea that leptin resistance may be partially overcome by a high enough overall leptin concentration.

Another study evaluated the efficacy of another long-acting leptin compound (A-200) in 200 obese subjects in a 24-week randomized, placebo-controlled pilot study with mild dietary intervention (500 calories below daily requirement). Results indicated that A-200 was safe, well tolerated, and resulted in a statistically significant decline in body weight and fat mass. Most of the weight loss again was determined to be secondary to decreases in fat mass. In a randomized, double-blind trial, 30 patients received either 20 mg of polyethylene glycol (PEG)-leptin or placebo weekly for 12 weeks. At the end of the study, patients receiving placebo had increased appetite and hunger levels in the fasting state, compared with reduced appetite and hunger in the treatment group. However, the treatment group did not experience reductions in daily food intake or body mass or changes in body composition compared with the control group. These findings led researchers to conclude that PEG-leptin has central rather than peripheral biologic activity in obese men.

Studies in rodents have demonstrated that leptin is highly efficacious as an antidiabetic agent in lipodystrophy, in which leptin deficiency is directly responsible for a hyperinsulinemic diabetic syndrome. The National Institute of Diabetes and Digestive and Kidney Diseases is currently studying the long-term efficacy of leptin replacement in patients with lipodystrophy (www.clinicaltrials.gov, National Institutes of Health protocols 02-DK-0146 and 02-DK-0022). Results from these studies are not yet available.

Feedback Control

Little is known about the cellular pathway involving leptin secretion. However, the rapid effects of -adrenergic stimulation on leptin release from adipose tissue suggest that leptin secretion is regulated by cAMP. As well, leptin secretion is upregulated by the hormones insulin and cortisol working synergistically and down-regulated by catecholamines, norepinephrine, and epinephrine. A report also suggests that cholecystokinin may regulate leptin secretion directly. Finally, TNF may be an important paracrine regulator of leptin secretion.

The interesting phenomenon of leptin resistance in obesity was initially suggested on the basis of the elevation of plasma leptin levels in obese humans. It turns out that, as with other cytokine receptors, activation of the leptin receptor induces expression of a protein called suppressor of cytokine signaling-3 (SOCS-3), which may inhibit further leptin signal transduction. The contribution of SOCS-3 to acquisition of leptin resistance and obesity remains an active area of investigation. As well, leptin receptors are expressed in the endothelial cells of the blood-brain barrier and it is plausible that dysfunction of this process may also lead to a state of obesity and leptin resistance.
Neuroendocrine-Immune Interactions

Stimulation of the immune system by foreign pathogens leads to a stereotyped set of responses orchestrated by the CNS. These responses are the result of the complex interaction of the immune system and the CNS and are often referred to as the cerebral component of the acute phase reaction. This constellation of stereotyped responses is adaptive, is mediated in large part by the hypothalamus, and includes coordinated autonomic, endocrine, and behavioral components. These responses include fever, alterations in the activity of nearly every neuroendocrine axis, changes in the sleep-wake cycle, anorexia, and inactivity.

It is now clear that cytokines produced by white blood cells of the immune system mediate the CNS responses. Early evidence supporting this hypothesis was provided by the seminal observations that cytokines such as IL-1 can activate the HPA axis. In fact, these and other observations provided the framework for a new area of research in neuroscience and neuroendocrinology. This discipline is often referred to as neuroimmunology. Thus, the term neuroimmunomodulation has been used to describe the study of the interactions of immune system cues and nervous system function.

Although it is established that cytokines modulate hypothalamic activity, it is also important to note that the immune system is modulated by the nervous system. This modulation occurs largely by two routes, endocrine mechanisms and direct innervation. The innervation includes lymphoid organs such as the thymus and spleen, which receive direct inputs from the autonomic nervous system. As noted earlier in the section on CRH, the hallmark of cytokine action on the hypothalamus is the activation of the HPA axis. The resultant glucocorticoid secretion acts as a classical negative feedback to the immune system to damp the immune response. This general, glucocorticoids inhibit most limits of the immune response, including lymphocyte proliferation, production of immunoglobulins, cytokines, and cytotoxicity. These inhibitory reactions form the basis of the anti-inflammatory actions of glucocorticoids.

Glucocorticoid feedback on immune responses is regulatory and beneficial because loss of this function makes animals with adrenal insufficiency vulnerable to inflammation. Moreover, this feedback response can have pathophysiologic consequences, as chronic activation of the HPA axis can certainly be detrimental. Indeed, it is now established that chronic stress can lead to immunosuppression. The fact that products of inflammation such as IL-1 can activate the HPA axis suggests the operation of a negative feedback control loop to regulate the intensity of inflammation. The role of the hypothalamus in regulating pituitary-adrenal function is an excellent example of neuroimmunomodulation.

This section addresses several of the hypothesized mechanisms by which cytokines engage neural pathways to mediate neuroendocrine and autonomic effects. In addition, some of the autonomic and endocrine pathways that are engaged by immune system cues are briefly discussed. It is important to note that many nonlymphocytic cells including endocrine and adipose cells and neurons also synthesize cytokines that exert effects independent of immunomodulation. Examples of cytokines secreted by adipocytes include leptin, adipin, and TNF, which have profound effects on metabolism.

Cytokines Signal the Central Nervous System

Cytokines made outside the CNS can alter the activity and function of populations of hypothalamic neurons. Although the interactions of cytokines with the nervous system have been studied extensively, the mechanisms by which immune signals influence the CNS remain unsettled. LPS (or endotoxin) is a cell wall component of all gram-negative bacteria that is a potent immune system stimulant. LPS administration is widely used as an experimental model and induces the secretion of several pyrogenic cytokines including IL-1, TNF, and IL-6 that mimic the patterns of cytokine production seen in natural infections.

Other cytokines secreted by adipocytes include leptin, adipin, and TNF, which have profound effects on metabolism.

Interaction of Cytokines with the Circumventricular Organs

The CVOs, described in detail earlier, are specialized regions on the margins of the ventricular system that have fenestrated capillaries and therefore no blood-brain barrier. Many circulating hormones such as angiotensin II act on neurons in the CVOs, converting blood-borne signals into CNS responses. Several models of fever production have hypothesized that cytokines may enter the CNS through the CVOs, particularly at the level of the OVLT (Fig. 7-50; see Fig. 7-49). However, definitive evidence establishing this model as a predominant mechanism is still lacking.

Large lesions of the preoptic area of the hypothalamus including the OVLT block fever, but they inevitably damage nearby regions that are critical for thermoregulation. Small lesions of the OVLT do not block fever or corticosterone responses. However, an inherent limitation of this type of study is that the lesion itself breaches the blood-brain barrier, allowing entry of cytokines. Moreover, knife cuts just caudal to the OVLT, interrupting connections from the OVLT to the PVH, did not block activation of the HPA axis by IL-1.

Other studies have focused on the area postrema, a CVO located in the medulla oblongata lying along the surface of the nucleus of the solitary tract at the caudal end of the fourth ventricle (see Fig. 7-50 and Fig. 7-51). Lesions of the area postrema can block the IL-1-induced activation of the HPA axis and the induction of c-fos mRNA in the PVH. However, the hypothalamic-pituitary-adrenal (HPA) axis by immune system stimulation. The immune system probably uses several pathways and sites of entry to communicate with the brain. This model predicts that circumventricular organs (organs devoid of blood-brain barrier; CVOs) and the blood vessels (bv) are crucial target sites of cytokines of systemic origin produced during the acute-phase response, whereas activated regions of the brain stem and deep limbic system might play a determinate role in the integration of information received from the periphery. Among these integrative structures, the PVN is critical in coordinating autonomic and endocrine responses including the activity of the HPA axis. For example, corticotropin-releasing factor (CRF) neurons of the paraventricular PVN expressed c-fos messenger ribonucleic acid, and that transcription of the gene coding CRF is activated essentially in this hypothalamic nucleus indicates the importance and the specificity of this neuroendocrine component of the acute phase reaction, particularly at the OVLT (Fig. 7-50). In general, glucocorticoids inhibit most limits of the immune response, including lymphocyte proliferation, production of immunoglobulins, cytokines, and cytotoxicity. These inhibitory reactions form the basis of the anti-inflammatory actions of glucocorticoids.

Although it is established that cytokines modulate hypothalamic activity, it is also important to note that the immune system is modulated by the nervous system. This modulation occurs largely by two routes, endocrine mechanisms and direct innervation. The innervation includes lymphoid organs such as the thymus and spleen, which receive direct inputs from the autonomic nervous system. As noted earlier in the section on CRH, the hallmark of cytokine action on the hypothalamus is the activation of the HPA axis. The resultant glucocorticoid secretion acts as a classical negative feedback to the immune system to damp the immune response (Fig. 7-50). Instead of this feedback response can have pathophysiologic consequences, as chronic activation of the HPA axis can certainly be detrimental. Indeed, it is now established that chronic stress can lead to immunosuppression. The fact that products of inflammation such as IL-1 can activate the HPA axis suggests the operation of a negative feedback control loop to regulate the intensity of inflammation. The role of the hypothalamus in regulating pituitary-adrenal function is an excellent example of neuroimmunomodulation.
nucleus in endotoxin-treated animals. The mechanisms and the circuitry controlling the CRF release and the activity of the HPA axis might also be different from those involved in the biosynthetic circuitry of CRF during immune challenge. ACTH, adrenocorticotropic hormone; AP, area postrema; ARC, arcuate nucleus; BNST, bed nucleus of the stria terminalis; bv, blood vessels; cP, choroid plexus; CeA, central nucleus of the amygdala; COX-2, cyclooxygenase-2; DMH, dorsomedial nucleus of the hypothalamus; EP, proopiomelanocortin E receptor; IL-1, interleukin 1; IL-1RI, IL-1 type 1 receptor; IL-6, interleukin 6; IFN, interferon; LC, locus coeruleus; LDT, laterodorsal tegmental nucleus; LPS, lipopolysaccharide; LRMN, lateral reticular nucleus medialis; ME, median eminence; MPDA, medial preoptic area; NF, nuclear factor; NTS, nucleus of the solitary tract; OVL, organum vasculosum of the lamina terminalis; PGE, prostaglandin E2; PB, parabrachial nucleus; PP, posterior pituitary; PVN, paraventricular nucleus of the hypothalamus; PVH, paraventricular nucleus [pc] and magnocellular [pm] divisions; SFO, subfornical organ; SON, supraoptic nucleus; TNF, tumor necrosis factor; VM, ventromedial nucleus.

Figure 7-51 Immune stimulation activates key brain regions. A series of photomicrographs demonstrating the distribution of Fos-like immunoreactivity (Fos-IR) in the rat brain 2 hours after intravenous injections of lipopolysaccharide (LPS; 125 µg/kg). LPS administration is a commonly used model of immune stimulation, and LPS-IR is a widely used marker of neuronal activation. LPS activates (induces Fos-IR) in the ventral medial preoptic area (vmPVT) and medial (mp) parvicellular and posterior magnocellular (pm) divisions. Also note that LPS activates neurons in the circumventricular organs (OVLT, SFO, AP); 3, third ventricle.

Macrophages (Fig. 7-62)_. Regardless of the cell type, it seems clear that circulating LPS or cytokines induce COX 2 in cells in the perivascular space, which in turn may produce prostaglandins to stimulate nearby brain regions inside the blood-brain barrier.

PGE 2, the predominant endogenous isoform of PGE in the brain, is thought to be an essential mediator of cytokine modulation of hypothalamic function. This claim is supported by the finding that microinjections of PGE receptor agonists into the brain of rats and other species produce fever. The preoptic area of the hypothalamus surrounding the OVL T is thought to be critical in the response to PGE 2 (see Fig. 7-51A). For example, microinjections of as little as 1 ng of PGE 2 into the anteroventral periaqueductal area of rats reliably produced fever. Conversely, the COX-2 inhibitor ketorolac attenuated LPS-induced fever with injections placed in the same region. This PGE-sensitive zone is the same as the region containing the highest concentrations of PGE 2 binding sites. The cloning of the prostaglandin E (EP) receptors has allowed more definitive analysis of the receptors in the hypothalamus that mediate the effects of PGE 2 (see Fig. 7-50).

Four EP receptor subtypes have been identified, EP 1, EP 2, EP 3, and EP 4. All four subtypes are expressed in the preoptic area of the hypothalamus. Despite the established role of PGE in producing fever and activating the HPA axis, the EP receptor subtypes that are crucial in the febrile response are not yet established. Pharmacologic evidence suggests that EP 1 and EP 2 receptor agonist administration has an effect mimicking PGE 2-induced fever. Moreover, an EP 1 receptor antagonist blocked PGE 2 fever. In contrast, targeted deletion of the EP 2 gene resulted in mice that did not show an early phase of fever after intracerebroventricular injection of LPS or PGE 2. Interestingly, there is up-regulation of EP 2 receptor expression in several areas of the brain, including the CRH neurons of the paraventricular nucleus, after immune challenge. In addition, paraventricular neurons that express Fos after intracerebroventricular PGE 2 also express EP 2 receptors. Thus, production of PGE 2 is certainly an obligate step in the pathogenesis of the febrile response; the identity of the EP receptor subtypes required for distinct components of the response remains to be established.

Entry of Cytokines into the Brain

Circulating cytokines are proteins that cannot easily penetrate the blood-brain barrier. The kinetics of entry of cytokines into the brain have been examined, and evidence suggests that there is saturable transport of IL-1, IL-6, and TNF into the brain. However, it is not clear whether sufficient levels of cytokines are detectable in brain after acute intravenous administration to account for CNS responses to acute infection. Thus, the physiologic setting and significance of this mechanism remain to be established.

Moreover, it is noteworthy that levels of circulating IL-1 do not rise significantly during immune challenges. In contrast, large increases in circulating and brain IL-6 are found during fever. Although it is not completely understood, it appears that synthesis of IL-6 within the blood-brain barrier and not peripheral IL-6 crossing the barrier is critical in the production of fever. Moreover, several studies have demonstrated that cells located at the blood-brain barrier and cells within the meninges respond to LPS stimulation with induction of IL-1 and TNF, the nuclear factor B inhibitor IB, and the LPS receptor CD14 (Fig. 7-52). Cells with a similar morphology lining the blood vessels that penetrate the CNS and the meninges that cover it also have IL-1 receptors, suggesting that they may respond to cytokines as well. Hence, endothelial and perivascular cells at the blood-brain interface may have the ability to elaborate cytokines after an LPS or cytokine signal. The physiologic role of centrally produced cytokines in the response to peripheral immune cues has been reviewed in detail.

Interactions of Cytokines with Peripheral Nerves

Another proposed model by which cytokines may alter the activity of CNS neurons involves stimulation of peripheral sensory nerves, the prototypical example being the vagus nerve. Several pieces of experimental evidence support the idea that the vagus may provide a conduit for cytokines to activate CNS pathways. For example, IL-1 receptor antagonist binds to vagal paranganglia. In addition, neurons in the nodose ganglia (vagal sensory neurons) expressed type 1 IL-1 receptor mRNA. Peripheral administration of LPS induced Fos expression, a marker of neuronal activation, in the vagal sensory nodose ganglia, and this could be blocked by prior vagotomy. Similarly, IL-1 administration induced the expression of Fos in neurons in the nodose ganglia and increased the firing rate of other vagal nerve fibers. The IL-1-induced response was blocked by pretreatment with a COX inhibitor, demonstrating the need for prostaglandin production in this model as well.

Severing the vagus nerve below the diaphragm blocked fever, sickness behavior, and induction of IL-1 mRNA and Fos protein in the brain after intraperitoneal LPS or IL-1. These observations suggest that, although vagal sensory mechanisms may contribute to CNS responses to immune stimuli, particularly with local
infections in the abdominal or thoracic cavities, blood-borne immune challenge may activate the CNS by other routes. In the end, it is likely that redundant mechanisms exist by which the CNS is made aware of inflammatory signals in the periphery. The relative contribution of distinct mechanisms may depend upon the route of administration and dose of inflammatory mediators, and future studies should increase our understanding of these mechanisms.

Cell Groups Throughout the Brain Responsive to Cytokines

Many studies have used the expression of immediate early genes such as c-fos or its protein product, Fos, as a marker of neuronal activity. In this way, investigators have assessed the involvement of extended neuronal systems during the complex physiologic responses after immune challenge. Mapping the patterns of activation in the CNS after either IL-1 or LPS administration has yielded new insights into the functional neuroanatomy underlying the coordinated autonomic, endocrine, and behavioral responses during the febrile response. Immune activation using moderate to high doses of LPS and IL-1 activates central autonomic and endocrine structures at nearly every level of the neuraxis including several neuroendocrine regulator sites such as the central nucleus of the amygdala, paraventricular hypothalamic nucleus, arcuate nucleus of the hypothalamus, SFO, OVLT, and ventral medial preoptic area. Immune activation of cells was found in the dorsal parvicellular division of the paraventricular nucleus in the hypothalamus. These results suggest that neurons in the parvicellular PVH specifically innervate sympathetic preganglionic neurons in the spinal cord that regulate LPS-induced fever. Furthermore, as noted earlier, activation of CRH neurons of the PVH is a signature of the CNS response to immune stimulation. Thus, the paraventricular hypothalamic nucleus is a key site for mediating both neuroendocrine and autonomic responses to immune stimulation.


noradrenergic cell group in the ventral pons, the caudal part of the nucleus of the solitary tract, the ventromedial medulla including the medullary raphe nuclei, and the rostral ventrolateral medulla, including the C1 adrenergic cell group. Lesions interrupting the input from the C1 cells to the PVH prevent the HPA response to IL-1. These studies suggest that the activation of C1 cells by locally produced prostaglandins may play a critical role in activating the HPA axis in response to IL-1.

Sympathetic preganglionic neurons in the intermediolateral cell column (IML), extending from the first thoracic through the upper lumbar segments of the spinal cord, also show Fos expression in response to LPS. Preganglionic neurons in the upper thoracic (T1 to T4) levels mediate thermogenesis by brown adipose tissue, which is a key mechanism used by rats to control heat production and body temperature. Sympathetic preganglionic neurons in the T2 to T5 levels are important for control of the heart, which is important because there are changes in cardiac output in the febrile state. Another important concept is that sympathetic preganglionic neurons receive direct, monosynaptic input from a series of well-defined nuclei in the brain stem and the hypothalamus. These cells provide another way in which the hypothalamus can contribute to the coordinated autonomic response to inflammatory signals. The major input to the sympathetic preganglionic column arises from neurons in the hypothalamus. This innervation includes the paraventricular nucleus (dorsal, ventral, and lateral parvicellular subnuclei), the lateral hypothalamic area, and the arcuate nucleus and retrochiasmatic area. Direct projections to the intermediolateral cell column also arise in the brain stem from the A5 noradrenergic cell group in the ventral pons, the caudal part of the nucleus of the solitary tract, the ventromedial medulla including the medullary raphe nuclei, and the rostral ventrolateral medulla, including the C1 adrenergic cell group. LPS-activated cells that innervate the IML are found in the rostral ventrolateral medulla (C1 adrenergic cell group) and the A5 noradrenergic cell group in the brain stem. Moreover, a prominent population of cells was found in the dorsal parvicellular division of the paraventricular nucleus in the hypothalamus. These results suggest that neurons in the parvicellular PVH specifically innervate sympathetic preganglionic neurons in the spinal cord that regulate LPS-induced fever. Furthermore, as noted earlier, activation of CRH neurons in the PVH is a signature of the CNS response to immune stimulation. Thus, the paraventricular hypothalamic nucleus is a key site for mediating both neuroendocrine and autonomic responses to immune stimulation.
NEUROENDOCRINE DISEASE

Disease of the hypothalamus can cause pituitary dysfunction, neuropsychiatric and behavioral disorders, and disturbances of autonomic and metabolic regulation. In the diagnosis and treatment of suspected hypothalamic or pituitary disease, four issues must be kept in mind: the extent of the lesion, the physiologic impact, the specific cause, and the psychosocial setting. The etiology of hypothalamic neuroendocrine disorders categorized by age and syndrome is summarized in Table 7.9 and Table 7.10.

Manifestations of pituitary insufficiency secondary to hypothalamic or pituitary stalk damage are not identical to those of primary pituitary insufficiency. Hypothalamic injury causes decreased secretion of most pituitary hormones but can cause hypersecretion of hormones normally under inhibitory control by the hypothalamus, as in hypersecretion of PRL after damage to the pituitary stalk and precocious puberty caused by loss of the normal restraint over gonadotropin maturation. Impairment of inhibitory control of the neurohypophysis can lead to the syndrome of inappropriate vasopressin secretion (SIADH) (see Chapter 9). More subtle abnormalities in secretion can result from impairment of the control system. For example, loss of the normal circadian rhythm of corticotropin secretion may occur after loss of pituitary-adrenal secretory reserve. and responses to physiologic stimuli may be paradoxical. Because hypothalamic hormone levels cannot be measured directly and pituitary secretion is regulated by complex, multilayered controls, assay of pituitary hormones in blood does not necessarily give a meaningful picture of events at hypothalamic and higher levels. Rarely, tumors secrete excessive amounts of releasing peptides and cause hypersecretion of hormones from the pituitary.

Disorders of the hypothalamic-pituitary unit can result from lesions at several levels. Defects can arise from destruction of the pituitary (as by tumor, infarct, inflammation, or autoimmune disease) or from a hereditary deficiency of a particular hormone as in rare cases of isolated FSH, GH, or POMC deficiency. Selective loss of thyroid hormone receptors in the pituitary can give rise to increased thyrotropin secretion and thyrotoxicosis. Furthermore, disorders can arise through disruption of the stalkmedian eminence contact zone, the stalk itself, or the nerve terminals of the tuberohypophyseal system; such disruption occurs after surgical stalk section, with tumors involving the stalk, and in some inflammatory diseases. At a higher level, tonic inhibitory and excitatory inputs can be lost as manifested by absence of circadian rhythms or the development of precocious puberty. Physical stress, cytokine products of inflammatory cells, toxins, and reflex inputs from peripheral homeostatic monitors also impinge on the tuberoinfundibular system. At the highest level of control, emotional stress and psychological disorders can activate the pituitary-adrenal stress response and suppress gonadotropin secretion (e.g., psychogenic amenorrhea) or inhibit GH secretion (e.g., psychosocial dwarfism) (see Chapter 23). Intrinsic disease of the anterior pituitary is reviewed in Chapter 6, and disturbances in neurohypophyseal function are discussed in Chapter 9. This chapter considers diseases of the hypothalamic-pituitary unit.

Pituitary Isolation Syndrome

Destructive lesions of the pituitary stalk, as occur with head injury, surgical transection, tumor, or granuloma, produce a characteristic pattern of pituitary dysfunction. Central diabetes insipidus (DI) develops in a large percentage of patients, depending on the level at which the stalk has been sectioned. If the cut is close to the hypothalamus, DI is almost always produced, whereas if the section is low on the stalk, the incidence is lower. The extent to which nerve terminals in the upper stalk are preserved determines the clinical course. The

---

TABLE 7.9 -- Etiology of Hypothalamic Disease by Age

<table>
<thead>
<tr>
<th>Premature Infants and Neonates</th>
<th>Intrauterine hemorrhage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meningitis: bacterial</td>
<td></td>
</tr>
<tr>
<td>Tumors: glioma, hemangiomata</td>
<td></td>
</tr>
<tr>
<td>Trauma</td>
<td></td>
</tr>
<tr>
<td>Hydrocephalus, kernicterus</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1 mo – 2 yr (infant)</th>
<th>Tumors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glioma, especially optic glioma</td>
<td></td>
</tr>
<tr>
<td>Histiocytosis X</td>
<td></td>
</tr>
<tr>
<td>Hemangiomata</td>
<td></td>
</tr>
<tr>
<td>Hydrocephalus</td>
<td></td>
</tr>
<tr>
<td>Meningitis</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Familial disorders</th>
<th>Laurence-Moon-Biedl syndrome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prader-Labhart-Willi syndrome</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>210 yr (child)</th>
<th>Neoplasms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Craniopharyngioma</td>
<td></td>
</tr>
<tr>
<td>Glioma, dysgerminoma, hamartoma</td>
<td></td>
</tr>
<tr>
<td>Histiocytosis X, leukemia</td>
<td></td>
</tr>
<tr>
<td>Ganglioneuroma, ependymoma</td>
<td></td>
</tr>
<tr>
<td>Medulloblastoma</td>
<td></td>
</tr>
<tr>
<td>Meningitis</td>
<td></td>
</tr>
</tbody>
</table>

---

150
<table>
<thead>
<tr>
<th>Condition</th>
<th>1025 yr</th>
<th>2550 yr</th>
<th>50 yr and Older</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bacterial</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Tuberculous</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Encephalitis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Viral</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exanthematos demyelinating</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Familial</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes insipidus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radiation therapy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetic ketoacidosis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moyamoya disease, circle of Willis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>1025 yr</strong></td>
<td>Tumors</td>
<td>Tumors</td>
<td>Tumors</td>
</tr>
<tr>
<td><strong>Craniohypophyseal</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glioma, hamartoma, dysgerminoma</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Histiocytosis X, leukemia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dermoid, lipoma, neuroblastoma</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Trauma</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Vascular</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subarachnoid hemorrhage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aneurysm</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arteriovenous malformation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Inflammatory disease</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meningitis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Encephalitis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sarcoidosis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tuberculosis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structural brain defect</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chronic hydrocephalus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Increased intracranial pressure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>2550 yr</strong></td>
<td></td>
<td>Nutritional: Wernicke's disease</td>
<td>Tumors: Pituitary tumors, sarcoma, glioblastoma, ependymoma, meningioma, colloid cysts, lymphoma</td>
</tr>
<tr>
<td><strong>Vascular</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aneurysm, subarachnoid hemorrhage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arteriovenous malformation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Damage from pituitary radiation therapy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>50 yr and Older</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nutritional: Wernicke's disease</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Tumors</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pituitary tumors, sarcoma, glioblastoma, ependymoma, meningioma, colloid cysts, lymphoma</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Vascular</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infarct, subarachnoid hemorrhage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pituitary apoplexy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inflammation: encephalitis, sarcoidosis, meningitis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Damage from radiation therapy for ear-nose-throat carcinoma, pituitary tumors</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Although head injury, granulomas, and tumors are the most common causes of acquired DI, other cases develop in the absence of a clear-cut cause. Some cases may be due to autoimmune disease of the hypothalamus as suggested by the finding of autoantibodies to neurohypophysial cells in a third of cases of “idiopathic” DI in one series. However, autoantibodies were also frequently found in association with histiocytosis-X. Later reports suggest the importance of continued vigilance in cases of idiopathic DI because a definite cause is frequently uncovered in time, including a high proportion of occult germinomas whose detection by magnetic...
resonance imaging may be preceded by elevated levels of human chorionic gonadotropin (hCG) in CSF.

**TABLE 7-10 -- Etiology of Endocrine Syndromes of Hypothalamic Origin**

<table>
<thead>
<tr>
<th>Hypophyseotropic Hormone Deficiency</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgical pituitary stalk section</td>
<td></td>
</tr>
<tr>
<td>Basilar meningitis and granuloma, sarcoidosis, tuberculosis, sphenoid osteomyelitis, eosinophilic granuloma</td>
<td></td>
</tr>
<tr>
<td>Craniopharyngioma</td>
<td></td>
</tr>
<tr>
<td>Hypothalamic tumor</td>
<td></td>
</tr>
<tr>
<td>Infundibuloma</td>
<td></td>
</tr>
<tr>
<td>Teratoma (ectopic pinealoma)</td>
<td></td>
</tr>
<tr>
<td>Neuroglial tumor, particularly astrocytoma</td>
<td></td>
</tr>
<tr>
<td>Maternal deprivation syndrome, psychosocial dwarfism</td>
<td></td>
</tr>
<tr>
<td>Isolated growth hormone-releasing hormone (GHRH) deficiency</td>
<td></td>
</tr>
<tr>
<td>Hypothalamic hypothyroidism</td>
<td></td>
</tr>
<tr>
<td>Panhypophyseotropic failure</td>
<td></td>
</tr>
</tbody>
</table>

**Disorders of Regulation of Gonadotropin-Releasing Hormone Secretion**

**Female**
- Precocious puberty
- GnRH-secreting hamartoma
- hCG-secreting germinoma
- Delayed puberty
- Neurogenic amenorrhea
- Pseudoamenorrhea
- Amenorrhea
- Drug-induced amenorrhea

**Male**
- Precocious puberty
- Fröhlich's syndrome
- Olfactory-genital dysplasia (Kallmann's syndrome)

**Disorders of Regulation of Prolactin-Regulating Factors**

<table>
<thead>
<tr>
<th>Tumor</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug-induced</td>
<td></td>
</tr>
<tr>
<td>Reflex</td>
<td></td>
</tr>
<tr>
<td>Herpes zoster of chest wall</td>
<td></td>
</tr>
<tr>
<td>Post-thoracotomy</td>
<td></td>
</tr>
<tr>
<td>Nipple manipulation</td>
<td></td>
</tr>
<tr>
<td>Spinal cord tumor</td>
<td></td>
</tr>
<tr>
<td>&quot;Psychogenic&quot;</td>
<td></td>
</tr>
<tr>
<td>Hypothyroidism</td>
<td></td>
</tr>
<tr>
<td>Carbon dioxide narcosis</td>
<td></td>
</tr>
</tbody>
</table>

**Disorders of Regulation of Corticotropin-Releasing Hormone**

| Paroxysmal corticotropin discharge (Wolff's syndrome) |  |
| Loss of circadian variation                        |  |
| Depression                                         |  |
| CRH-secreting gangliocytoma                        |  |
| CRH, corticotropin-releasing hormone; GnRH, gonadotropin-releasing hormone; hCG, human chorionic gonadotropin. |  |

Congenital DI can be part of a hereditary disease. DI in the Brattleboro rat is due to an autosomal recessive genetic defect that impairs production of vasopressin but not of oxytocin. Inherited forms of DI in humans have been attributed to mutations in the vasopressin V2 receptor gene or less frequently in the aquaporin or
Menstrual cycles cease after stalk section although urinary gonadotropins may still be detectable, unlike the situation after hypophysectomy. Plasma glucocorticoid levels and urinary excretion of cortisol and 17-hydroxycorticoids decline after hypophysectomy and stalk section, but the change is slower after stalk section. A transient increase in cortisol secretion after stalk section is believed to be due to release of ACTH from preformed stores. The ACTH response to the lowering of blood cortisol is markedly reduced but ACTH release after stress may be normal, possibly because of CRH-independent mechanisms. Reduction in thyroid function after stalk section is similar to that seen with hypophysectomy. The fall in GH secretion is said to be the most sensitive indication of damage to the stalk; however, the insidious nature of this endocrinologic change in adults who have suffered traumatic brain injuries may cause it to be overlooked and therefore contribute to delayed rehabilitation.

Humans with stalk sections or with tumors of the stalk region have widely varying levels of hyperprolactinemia and may have galactorrhea. PRL responses to hypoglycemia and to TRH are blunted, in part because of loss of neural connections with the hypothalamus. PRL responses to dopamine agonists and antagonists in the pituitary isolation syndrome are similar to those in patients with prolactinomas. Interestingly, PRL secretion continues to show a diurnal variation in patients with either hypothalamic-pituitary disconnection or microprolactinoma. Both forms of hyperprolactinemia are characterized by a similarly increased frequency of PRL pulses and a marked rise in nonpulsatile or basal PRL secretion, although the disruption is greater in the tumoral hyperprolactinemia.

An incomplete pituitary isolation syndrome may occur with the empty sella syndrome, intrasellar cysts, or pituitary adenomas. Anterior pituitary failure after stalk section is in part due to loss of specific neural and vascular links to the hypothalamus and in part due to pituitary infarction.
Selective pituitary failure can be due to a deficiency of specific pituitary cell types or a deficiency of one or more hypothalamic hormones. Isolated GnRH deficiency is the most common hypophyseotropic hormone deficiency. In Kallmann’s syndrome (gonadotropin deficiency commonly associated with hyposmia), hereditary agenesis of the olfactory lobe may be demonstrable by magnetic resonance imaging. Abnormal development of the GnRH system is due to defective migration of the GnRH-containing neurons from the olfactory nasal epithelium in early embryologic life (see the earlier section on GnRH). Other malformations of the cranial midline structures, such as absence of the septum pelucidum in septo-optic dysplasia (De Morsier’s syndrome), can cause hypogonadotropic hypogonadism (HH) or, less commonly, precocious puberty. A surprisingly large percentage of children with septo-optic dysplasia who otherwise have multiple hypothalamic-pituitary abnormalities actually retain normal gonadotropin function and enter puberty spontaneously. The genetic basis of HH has now been established in approximately 10% of patients. Mutations in the KAL (Kallmann’s syndrome) gene and the AHC-DAX1 (adrenal hypoplasia congenital-HH) gene cause X-linked recessive disease. Autosomal recessive HH has been associated with mutations in the GnRH receptor, leptin, leptin receptor, FSH, LH, PROP-1 (combined pituitary deficiency), and HESX (septo-optic dysplasia) genes.

The GnRH response test is of little value in the differential diagnosis of hypothyroidism. Most patients with GnRH deficiency show little or no response to an initial test dose, but normal responses are seen after repeated injection. This slow response has been attributed to down-regulation of GnRH receptors in response to prolonged GnRH deficiency. Furthermore, with intrinsic pituitary disease the response to GnRH may be absent or normal. Consequently, it is not possible to distinguish between hypothalamic and pituitary disease with a single injection of GnRH. Prolonged infusions or repeated administration of GnRH agonists after an initial test dose, but normal responses are seen after repeated injection. This slow response has been attributed to down-regulation of GnRH receptors in response to prolonged GnRH deficiency. Furthermore, with intrinsic pituitary disease the response to GnRH may be absent or normal. Consequently, it is not possible to distinguish between hypothalamic and pituitary disease with a single injection of GnRH. Prolonged infusions or repeated administration of GnRH agonists after hormone replacement therapy priming may aid in the diagnosis or provide therapeutic options for women with Kallmann’s syndrome wishing to become pregnant.

Deficiency of TRH secretion gives rise to hypothyroidism, also called tertiary hypothyroidism. This can occur in hypothyroidism or more rarely as an isolated defect. Molecular genetic analyses have revealed infrequent autosomal recessive mutations in the TRH and TRH receptor genes in the etiology of central hypothyroidism. Hypothalamic and pituitary causes of TSH deficiency are most readily distinguished by imaging methods. Although theoretically reasonable, the TRH stimulation test for the differentiation of hypothyroidism from pituitary disease is of limited value. The typical pituitary response to TRH administration in patients with TRH deficiency is an enhanced and somewhat delayed peak, whereas the response with pituitary failure is subnormal or absent. The hypothalamic type of response has been attributed to an associated GH deficiency that sensitizes the pituitary to TRH (possibly through suppression of somatostatin secretion), but GH also affects T₄ metabolism and may alter pituitary responses as well. In practice, the responses to TRH in hypothalamic and pituitary disease overlap so much that they cannot be used reliably for a differential diagnosis. Persistent failure to demonstrate responses to TRH is good evidence for the presence of intrinsic pituitary disease, but the presence of a response does not mean that the pituitary is normal. Deficient TRH secretion leads to delayed TSH biosynthesis by the pituitary, including impaired glycosylation. Poorly glycosylated TSH has low biologic activity, and dissociation of bioactive and immunoreactive TSH can lead to the paradox of normal or elevated levels of TSH in hypothyroidism.

GHRH deficiency appears to be the principal cause of hGH deficiency in children with idiopathic dwarfism. This condition is frequently associated with abnormal electroencephalograms, a history of birth trauma, and breech delivery. Furthermore, magnetic resonance imaging scans show that a substantial proportion of children with idiopathic hGH deficiency have evidence of a torn pituitary stalk, which is presumed evidence for birth trauma as the cause. Human GH is the most vulnerable of the anterior pituitary hormones when the pituitary stalk is damaged. It can be difficult to differentiate between primary pituitary disease and GHRH deficiency by standard tests of GH reserve. However, a substantial GH secretory response to a single administration of hexarelin occurs only in the presence of at least a partially intact vascular stalk.

In many children with dwarfism, the anatomic abnormalities of the intrasellar contents and pituitary stalk together with the frequent occurrence of other midline defects, such as those in septo-optic dysplasia, are consistent with the alternative hypothesis of a developmental defect occurring in embryogenesis. There has been a remarkable advance in our understanding of the molecular ontogeny of the hypothalamic-pituitary unit, much of it based on mutant mouse models. Parallel genetic analyses have been conducted in children with isolated GH deficiency or combined pituitary hormone deficiencies. These studies have identified autosomal recessive mutations in both structural and regulatory genes including the GHRH receptor, PTP1, PROP1, and HESX1 that are responsible for a sizable proportion of congenital hypothalamic-pituitary disorders once considered idiopathic.

Adrenal insufficiency is another manifestation of hypothyroidism and can be due to CRH deficiency. Isolated ACTH deficiency is uncommon, but there is suggestive evidence in at least one family of genetic linkage to the CRH gene locus. Later investigations have revealed mutations in the TPIT gene, a T box transcription factor expressed only in pituitary corticotrophs and melanotrophs, associated with cases of isolated ACTH deficiency. The CRH stimulation test does not distinguish hypothalamic from pituitary failure as a cause of corticotropin deficiency.

Apart from intrinsic diseases of the hypothalamus such as tumors and granulomas, two environmental causes of central hypophyseotropic deficiencies are of increasing clinical importance. These are trauma to the brain, particularly from motor vehicle accidents, and the sequelae of chemotherapy and radiation therapy for intracranial lesions in children and adults. Improved short-term survival from head injuries associated with coma and CNS malignancies has greatly increased the prevalence of long-term neuroendocrine consequences.
Hypophyseotropic Hormone Hypersecretion

Pituitary hypersecretion is occasionally caused by tumors of the hypothalamus. GnRH-secreting hamartomas can cause precocious puberty. CRH-secreting gangliocytomas can cause Cushing's syndrome, and GHRH-secreting gangliocytomas of the hypothalamus can cause acromegaly. Although they do not arise from the hypothalamus, paraneoplastic syndromes can also cause pituitary hypersecretion, as with CRH-secreting tumors and GHRH-secreting tumors of the bronchi and pancreas. Bronchial carcinoids and pituitary islet cell tumors are the usual causes of this phenomenon.
Precocious puberty is a relatively unusual manifestation of maturation, and some of these tumors have histologic and functional characteristics of choriocarcinomas. Diagnosis is confirmed by the combination of a mass lesion, two or more germ cell layers also occur in the pineal region. Chorionic tissue in teratomas and germinomas may secrete hCG in sufficient amounts to cause gonadal insufficiency, and visual abnormalities. Germinomas may also occur in the anterior hypothalamus or the floor of the third ventricle, where they are often associated with the clinical triad of DI, pituitary dysfunction, and visual abnormalities.

The most common tumors of the pineal gland are actually germinomas (a form of teratoma), so designated because of their presumed origin in germ cells. Pineal cysts, arachnoid cysts, and cavernous hemangioma. Pinealocytes give rise to primitive neuroectodermal tumors, the so-called small blue cell tumors that are oat cell carcinomas of the lung. pineal gland tumors account for only a small percentage of intracranial neoplasms. They occur as a central midline mass with an enhancing lesion on magnetic resonance imaging frequently accompanied by hydrocephalus. Pinealomas cause a variety of neurologic abnormalities. (Table 7-12), Parinaud's syndrome, which consists of paralysis of upward gaze, pupillary areflexia (to light), paralysis of convergence, and a wide-based gait, occurs with about half of pinealomas. Gait disturbances can also occur because of brain stem or cerebellar compression.

Several discrete cytopathologic entities account for mass lesions in the pineal region. The most common non-neoplastic conditions are degenerative pineal cysts, arachnoid cysts, and cavernous hemangioma. Pinealocytes give rise to primitive neuroectodermal tumors, the so-called small blue cell tumors that are immunopositive for the neuronal marker synaptophysin and negative for the lymphocyte marker CD45. True pinealomas can be relatively well-differentiated pineocytomas, middle neuroepithelial differentiation or the less differentiated pineoblastomas, which are typically the same as medulloblastomas, neuroblastomas, and oat cell carcinomas of the lung.

The most common tumors of the pineal gland are actually germinomas (a form of teratoma), so designated because of their presumed origin in germ cells. Germinomas may also occur in the anterior hypothalamus or the floor of the third ventricle, where they are often associated with the clinical triad of DI, pituitary insufficiency, and visual abnormalities. Identical tumors can be found in the testis and anterior mediastinum. Intracranial germinomas have a tendency to spread locally, infiltrate the hypothalamus, and metastasize to the spinal cord and CSF. Extracranial metastases (to the skin, lung, or liver) are rare. Teratomas derived from two or more germ cell layers also occur in the pineal region. Chorionic tissue in teratomas and germinomas may secrete hCG in sufficient amounts to cause gonadal maturation, and some of these tumors have histologic and functional characteristics of chorionicarcinomas. Diagnosis is confirmed by the combination of a mass lesion, cytologic analysis of CSF, and radioimmunoassay detection of hCG in the CSF.

Precocious puberty is a relatively unusual manifestation of...
A. Germ Cell Tumors

1. Germinoma
   a. Posterior third ventricle and pineal lesions
   b. Anterior third ventricle, suprasellar or intrasellar lesions
   c. Combined lesions in anterior and posterior third ventricle, apparently noncontiguous, with or without foci of cystic or solid teratoma

2. Teratoma
   a. Evidencing growth along two or three germ lines in varying degrees of differentiation
   b. Dermoid and epidermoid cysts with or without solid foci of teratoma
   c. Histologically malignant forms with or without differentiated foci of benign, solid, or cystic teratoma-teratocarcinoma, choriocarcinoma, embryonal carcinoma (endodermal-sinus tumor or yolk-sac carcinoma), combinations of these with or without foci of germinoma, chemodectoma

B. Pineal Parenchymal Tumors

1. Pinealocytes
   a. Pineocytoma
   b. Pineoblastoma
   c. Ganglioglioma and chemodectoma
   d. Mixed forms exhibiting transitions between these

2. Glia
   a. Astrocytoma
   b. Ependymoma
   c. Mixed forms and other less frequent gliomas (e.g., glioblastoma, oligodendroglioma)

C. Tumors of Supporting or Adjacent Structures

1. Meningioma

2. Hemangiopericytoma

D. Non-neoplastic Conditions of Neurosurgical Importance

1. "Degenerative" cysts of pineal lined by fibrillary astrocytes

2. Arachnoid cysts

3. Cavernous hemangioma


Table 7.12 -- Pinealomas: Frequency (%) of Presenting Symptoms and Signs

<table>
<thead>
<tr>
<th>Symptom</th>
<th>Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increased intracranial pressure</td>
<td>85</td>
</tr>
<tr>
<td>Spasticity</td>
<td>35</td>
</tr>
<tr>
<td>Ataxia</td>
<td>30</td>
</tr>
<tr>
<td>Parinaud's syndrome</td>
<td>25</td>
</tr>
<tr>
<td>Cerbellar-type nystagmus</td>
<td>25</td>
</tr>
<tr>
<td>Syncope</td>
<td>20</td>
</tr>
<tr>
<td>Vertigo</td>
<td>20</td>
</tr>
<tr>
<td>Cranial nerve palsy (other than cranial nerves VI, VIII)</td>
<td>20</td>
</tr>
<tr>
<td>Intention tremor</td>
<td>15</td>
</tr>
<tr>
<td>Scotoma</td>
<td>10</td>
</tr>
<tr>
<td>Tinnitus</td>
<td>10</td>
</tr>
<tr>
<td>Other</td>
<td>10</td>
</tr>
</tbody>
</table>


Management of tumors in the pineal region is not straightforward. Operative mortality rates can be high, but the rationale for an aggressive approach to the pineal region is based on the need to make a histologic diagnosis, the variety of lesions found in this region, the possibility of cure of an encapsulated lesion, and the effectiveness of chemotherapeutic agents for germinomas and chorocarcinoma. Stereotaxic biopsy of the pineal region provided diagnosis in 33 of 34 cases in one series, suggesting that this is a useful alternative to open surgical exploration for diagnostic purposes. Long-term palliation or cure of many pineal region tumors is possible by combinations of surgery, radiation, gamma knife, or chemotherapy, depending on the nature of the lesion. Hence causes premature puberty almost exclusively in boys. The prevalence of elevated hCG levels in children with premature puberty related to tumors in the pineal region is unknown, but the fact that this phenomenon occurs further challenges the theory that nonparenchymal tumors cause precocious puberty by damaging the normal pineal gland. Rarely, pinealomas cause delayed puberty, raising speculation about a role of melatonin in inhibiting gonadotropin secretion in these cases.

Management of Sexual Precocity

Several groups have reviewed the diagnostic approach to suspected central precocious puberty. Although guidelines differ, the index of suspicion is clearly inversely proportional to the age of the patient. A GnRH stimulation test to assess gonadotropin release and thereby differentiate between primed and inactive gonadotrophs is probably the single most important endocrinologic measure. If LH and FSH levels are not stimulated and there is no evidence of gonadal germ cell maturation, the cause of precocious puberty lies outside the hypothalamic-pituitary axis and the diagnostic process should focus on the adrenal glands and gonads (see Chapter 13 and Chapter 15). Magnetic resonance imaging studies are central to the work-up for exclusion or characterization of organic lesions in the areas of the sella, optic chiasm, suprasellar hypothalamus, and interpeduncular cistern. Precocious puberty is stressful to both the child and...
the parents, and it is essential that psychological support be provided.

Psychogenic Amenorrhea

Menstrual cycles can cease in young nonpregnant women with no demonstrable abnormalities of the brain, pituitary, or ovary in several situations, including pseudocyesis (false pregnancy), anorexia nervosa, excessive exercise, psychogenic disorders, and hyperprolactinemic states. Psychogenic amenorrhea, the most common cause of secondary amenorrhea except for pregnancy, can occur with major psychopathology or minor psychic stress and is often temporary. Psychogenic amenorrhea is probably mediated by excessive endogenous opioid activity because naloxone or naltrexone (opiate receptor blockers) can induce ovulation in some patients with this disorder.

Exercise-induced amenorrhea may be a variant of psychogenic amenorrhea or may result from loss of body fat. The syndrome is associated with intense and prolonged physical exertion such as running, swimming, or ballet dancing. Such women are always below ideal body weight and have low stores of fat. If the activity is begun before puberty, normal sexual maturation can be delayed for many years. The mass of fat may be a regulator of gonadotropin secretion with adipocyte-derived leptin as the principal mediator between peripheral energy stores and hypothalamic regulatory centers. Studies in nonhuman primates showed a direct role of caloric intake in the pathogenesis of amenorrhea associated with long-distance running. Exercise and psychogenic amenorrhea can have adverse effects because of the associated estrogen deficiency and accompanying osteopenia (also see Chapter 23).

Neurogenic Hypogonadism in Males

A discussion of neurogenic hypogonadism in males should begin with an account of Fröhlich's syndrome (adiposogenital dystrophy), originally characterized as delayed puberty, hypogonadism, and obesity associated with a tumor that impinges on the hypothalamus. It was subsequently recognized that either hypothalamic or pituitary dysfunction can induce hypogonadism and the presence of obesity indicates that the appetite-regulating regions of the hypothalamus have been damaged. Several organic lesions of the hypothalamus can cause this syndrome, including tumors, encephalitis, microcephaly, Friedreich's ataxia, and demyelinating diseases. Other important causes of hypogonadotropic hypogonadism are Kallmann's syndrome, a disorder caused by failure of GnRH-containing neurons to migrate normally (see earlier in the section on GnRH and hypophyseotropic hormone deficiency), and a subset of the Prader-Willi syndrome.

However, most males with delayed sexual development do not have serious neurologic conditions. Furthermore, most obese boys with delayed sexual development have no structural damage to the hypothalamus but have constitutional delayed puberty, which is commonly associated with obesity. It is not known whether there is a functional disorder of the hypothalamus in this condition. It is generally believed that psychosexual development of brain maturation depends on the presence of androgens and that hypogonadism in boys (regardless of cause) should be treated by the middle teen years (15 years at the latest).

In adult men, hypogonadism (including reduced spermatogenesis) can be induced by emotional stress or severe exercise, but this abnormality is seldom diagnosed because the symptoms are more subtle than menstrual cycle changes in similarly stressed women. Prolonged physical stress and sleep and energy deficiency can also decrease testosterone and gonadotropin levels. Chronic intrathecal administration of opiates for the control of intractable pain syndromes is strongly associated with hypogonadotropic hypogonadism, and to a lesser extent hypocorticism and GH deficiency, in both men and women. Finally, critical illness with multiple causes is well known to be associated with hypogonadism and ineffectual altered pulsation of GnRH.
Neurogenic Disorders of Prolactin Regulation

Neurogenic causes of hyperprolactinemia include irritative lesions of the chest wall (herpes zoster, thoracotomy), excessive tactile stimulation of the nipple, and lesions within the spinal cord such as ependymoma. Prolonged mechanical stimulation of the nipples by suckling or the use of a breast pump can initiate lactation in some women who are not pregnant, and neurologic lesions that interrupt the hypothalamic-pituitary connection can cause hyperprolactinemia, as discussed earlier. Hyperprolactinemia also occurs after certain forms of epileptic seizures. In one series, six of eight patients with temporal lobe seizures had a marked increase in PRL, whereas only one in eight frontal lobe seizures led to hyperprolactinemia. Agents that block dopamine receptors (such as the phenothiazines) or prevent dopamine release (e.g., reserpine and methyldopa) must be excluded in all cases.

Because the nervous system exerts such profound effects on PRL secretion, patients with hyperprolactinemia (including those with adenomas) may have a deficit of PIF or an excess of PRF activity. In studies of PRL secretion in patients apparently cured of hyperprolactinemia by removal of a pituitary microadenoma, regulatory abnormalities persisted in some but not all patients. Persistence of regulatory abnormalities may be due to incomplete removal of tumor, abnormal function of the remaining part of the gland, or underlying hypothalamic abnormalities.
Neurogenic Disorders of Growth Hormone Secretion

Hypothalamic Growth Failure

Loss of the normal nocturnal increase in GH secretion and loss of GH secretory responses to provocative stimuli occur early in the course of hypothalamic disease and may be the most sensitive endocrine indicator of hypothalamic dysfunction. As noted earlier, anatomic malformations of midline cerebral structures are associated with abnormal GH secretion, presumably related to failure of the development of normal GH regulatory mechanisms. Such disorders include optic nerve dysplasia and midline prosencephalic malformations (absence of the septum pellucidum, abnormal third ventricle, and abnormal lamina terminalis). Certain complex genetic disorders including Prader-Willi syndrome also commonly involve reduced GH secretory capacity. Idiopathic hypopituitarism with GH deficiency was considered earlier in this chapter.

Maternal Deprivation Syndrome and Psychosocial Dwarfism

Infant neglect or abuse can impair growth and cause failure to thrive (the maternal deprivation syndrome). Malnutrition interacts with psychological factors to cause growth failure in children with the maternal deprivation syndrome, and each case should be carefully evaluated from this point of view. Older children with growth failure in a setting of abuse or severe emotional disturbance (termed psychosocial dwarfism) may also have abnormal circadian rhythms and deficient hGH release after insulin-induced hypoglycemia or arginine infusion (see Chapter 8). Deficient release of corticotropin and gonadotropins may also be present. A new variant termed hyperphagic short stature has been identified. These disorders are reversible by placing the child in a supportive milieu where growth and neuroendocrine hGH responses rapidly return to normal. The pathogenesis of altered GH secretion in children in response to deprivation is unknown. In the adult human, furthermore, physical or emotional stress usually causes an increase in hGH secretion, as noted earlier.

Neuroregulatory Growth Hormone Deficiency

The availability of biosynthetic hGH for treatment of short stature has brought into focus a group of patients who grow at low rates (below the third percentile) and have low levels of serum IGF-I but a normal hGH secretory reserve. Studies of 24-hour hGH secretion profiles indicate that many of these children do not have normal spontaneous hGH secretion (abnormal ultradian and circadian rhythms and decreased number or amplitude of secretory bursts, or both). These children with idiopathic short stature may have a functional regulatory disturbance of the hypothalamus and appear to grow normally when given exogenous hGH.

There is considerable uncertainty about the criteria for the diagnosis of neuroregulatory hGH deficiency. Many normally growing children have profiles of hGH secretion that are indistinguishable from those in children with the postulated syndrome. Patterns of hGH secretion do not predict which child will benefit from therapy, and there is a poor correlation between hGH secretion and growth. Furthermore, the results of repeated tests in children show considerable variability. It has been suggested that specific genetic defects may underlie the pathogenesis of a subset of children with this heterogeneous syndrome of growth failure. The prevalence of an hGH neuroregulatory deficiency syndrome is thus unclear, and the decision to treat short children with hGH should be made cautiously.

Neurogenic Hyperscretion of Growth Hormone

Diencephalic Cachexia

Children and infants with tumors in and around the third ventricle frequently become cachectic, which is often associated with elevated hGH levels and paradoxical GH secretory responses to glucose and insulin. GH hypersecretion may be due to a hypothalamic abnormality or to malnutrition. Deficits of pituitary-adrenal regulation are less common.

<table>
<thead>
<tr>
<th>Clinical Feature</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emaciation</td>
<td>100</td>
</tr>
<tr>
<td>Alert appearance</td>
<td>87</td>
</tr>
<tr>
<td>Increased vigor or hyperkinesis, or both</td>
<td>72</td>
</tr>
<tr>
<td>Vomiting</td>
<td>68</td>
</tr>
<tr>
<td>Euphoria</td>
<td>59</td>
</tr>
<tr>
<td>Pallor</td>
<td>55</td>
</tr>
<tr>
<td>Nystagmus</td>
<td>55</td>
</tr>
<tr>
<td>Irritability</td>
<td>32</td>
</tr>
<tr>
<td>Hydrocephalus</td>
<td>33</td>
</tr>
<tr>
<td>Optic atrophy</td>
<td>24</td>
</tr>
<tr>
<td>Tremor</td>
<td>23</td>
</tr>
<tr>
<td>Sweating</td>
<td>15</td>
</tr>
<tr>
<td>Large hands, feet</td>
<td>5</td>
</tr>
<tr>
<td>Large genitalia</td>
<td>5</td>
</tr>
<tr>
<td>Polyuria</td>
<td>5</td>
</tr>
<tr>
<td>Papilledema</td>
<td>5</td>
</tr>
<tr>
<td>Positive pneumoencephalogram results</td>
<td>98</td>
</tr>
<tr>
<td>Endocrine anomalies</td>
<td>90</td>
</tr>
<tr>
<td>Cerebrospinal fluid protein</td>
<td>64</td>
</tr>
<tr>
<td>Cerebrospinal fluid abnormal cells</td>
<td>23</td>
</tr>
</tbody>
</table>

TABLE 7-14 — Tumors Producing Diencephalic Syndrome

<table>
<thead>
<tr>
<th>Tumor</th>
<th>No. of Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gliomas</td>
<td>56</td>
</tr>
<tr>
<td>Astrocytoma</td>
<td>37</td>
</tr>
<tr>
<td>Not subclassified</td>
<td>10</td>
</tr>
<tr>
<td>Spongioblastoma</td>
<td>5</td>
</tr>
<tr>
<td>Astroblastoma</td>
<td>1</td>
</tr>
<tr>
<td>Oligodendrogliona</td>
<td>1</td>
</tr>
<tr>
<td>Mixed astrocytoma-spongioblastoma</td>
<td>1</td>
</tr>
<tr>
<td>Mixed astrocytoma-oligodendrogloma</td>
<td>1</td>
</tr>
<tr>
<td>Ependymoma</td>
<td>2</td>
</tr>
<tr>
<td>Ganglioglioma</td>
<td>1</td>
</tr>
<tr>
<td>Dysgerminoma</td>
<td>1</td>
</tr>
<tr>
<td>No histology</td>
<td>10</td>
</tr>
</tbody>
</table>


A striking feature is an alert appearance and seeming euphoria despite the wasted state. A variety of associated neurologic abnormalities may be present (Table 7-13); the tumors that produce this syndrome are summarized in Table 7-14 and include a high proportion of chiasmatic-hypothalamic gliomas.

Syndrome of Inappropriate Growth Hormone Hypersecretion

Apparent inappropriate hGH hypersecretion (the syndrome of inappropriate somatotropin secretion) occurs with uncontrolled diabetes mellitus, hepatic failure, uremia, anorexia nervosa, and protein-calorie malnutrition. Nutritional factors are probably important in this response because in normal persons obesity inhibits and fasting stimulates episodic GH hypersecretion. In diabetes mellitus cholinergic blockers reverse the abnormality, possibly by inhibiting hypothalamic somatostatin secretion (see earlier in the section on neurotransmitter regulation of GH). Loss of inhibition of GH secretion by IGF-I may also play a role because most disorders in which this syndrome occurs are associated with low IGF-I levels.
Neurogenic Disorders of Corticotropin Regulation

Hypothalamic CRH hypersecretion is the likely cause of sustained pituitary-adrenal hyperfunction in at least two situations: Cushing's syndrome caused by the rare CRH-secreting gangliocytomas of the hypothalamus and severe depression.

Severe depression is associated with pituitary-adrenal abnormalities, including inappropriately elevated corticotropin levels, abnormal cortisol circadian rhythms, and resistance to dexamethasone suppression. The dexamethasone suppression test has, in fact, been used as an aid to the diagnosis of depressive illness. Patients with depression also have diminished responses to CRH, suggesting that depressed individuals hypersecrete CRH (see earlier section on CRH). Another possible example of disordered neurogenic control of CRH associated with stress is the metabolic syndrome. This syndrome is characterized by mild hypercortisolism, blunted dexamethasone suppression of the HPA axis, visceral obesity, and hypertension and may be strongly associated with greater risks for cardiovascular disease and stroke.

A unique syndrome of corticotropin hypersecretion termed periodic hypothalamic discharge (Wolff's syndrome) has been described in one young man. The patient had a recurring cyclic disorder characterized by high fever, paroxysms of glucocorticoid hypersecretion, and electroencephalographic abnormalities.
Genetic Obesity Disorders Involving Hypothalamic Circuits

In the past 5 years there have been a number of important discoveries related to genetic mutations underlying certain human obesity disorders. These clinical advances have closely paralleled the advances of basic research in the neuro-endocrine control of energy homeostasis discussed in an earlier section. Linkage studies and quantitative trait loci analyses have strongly implicated the POMC gene locus as an important determinant of weight homeostasis in humans of many, but not all, different ethnic populations, although specific alleles associated with obesity have not yet been demonstrated. Because no mutations within the coding region of the POMC gene that alter peptide activity have been identified in these populations, a current hypothesis is that mutations in regulatory regions of the gene decrease the level of POMC expression in the brain.

However, a small number of children from consanguineous parents have been found to have null mutations in the POMC gene resulting in absence of detectable circulating ACTH. These children presented with a syndrome of red hair, adrenal insufficiency, and severe, early-onset obesity (see Fig. 7-54). In addition, both dominant and recessive mutations in the MC4R gene have been found in the human population, and MC4R mutations have been proposed to play a role in as many as 5% of pediatric obesity cases. The genetic mirror image may also be true; an association between a polymorphism linked to the gene encoding the MC4R antagonist agouti-related protein and anorexia nervosa has been reported. Taking all these data into account, it is safe to say that obesity in a subpopulation of humans can be considered a genetic disorder of the hypothalamus.
Nonendocrine Manifestations of Hypothalamic Disease

The hypothalamus is involved in the regulation of diverse functions and behaviors. Psychological abnormalities in hypothalamic disease include antisocial behavior; attacks of rage, laughing, and crying; disturbed sleep patterns; excessive sexuality; and hallucinations. Both somnolence (with posterior lesions) and pathologic wakefulness (with anterior lesions) occur, as do bulimia and profound anorexia. The abnormal eating patterns are analogous to the syndromes of hyperphagia produced in rats by destruction of the ventromedial nucleus or of connections to the paraventricular nucleus. Lateral hypothalamic damage causes profound anorexia.

Patients with hypothalamic damage may experience hyperthermia, hypothermia, unexplained fluctuations in body temperature, and poikilothermy. Disturbances of sweating, acrocyanosis, loss of sphincter control, and diencephalic epilepsy are occasional manifestations. Hypothalamic damage also causes loss of recent memory, believed to be due to damage of the mammillothalamic pathways. Severe memory loss, obesity, and personality changes (apathy, loss of ability to concentrate, aggressive antisocial behavior, severe food craving, inability to work or attend school) may occur with suprasellar extension of pituitary tumors, hypothalamic radiation, or damage incurred from surgical removal of parasellar tumors. Hypothalamic tumors grow slowly and may reach a large size while producing minimal disturbance of behavior or visceral homeostasis, whereas surgery of limited extent can produce striking functional abnormalities. Presumably, this is because slowly growing lesions permit compensatory responses to develop. These potential consequences should be weighed carefully with the neurosurgeon, patient, and patient's family in planning the therapeutic approach. Adverse effects of treatment have led to more conservative surgical guidelines for the treatment of craniopharyngioma.

A convergence of functional genomics from two animal species, the dog and mouse, has refocused attention on neuropeptide circuits of the hypothalamus in the control of sleep. Positional cloning was used to identify mutations in the hypocretin-orexin receptor 2 as the cause of canine narcolepsy. Knockout of the gene encoding the hypocretin-orexin peptide precursor produced an equivalent narcoleptic syndrome in mice, further establishing this neuropeptide system as a major component of sleep-modulating neural circuits. Histaminergic neurons of the tuberomammillary nucleus express both forms of the orexin receptor and make reciprocal synaptic connections with orexin neurons in the lateral hypothalamus. Furthermore, orexin is an excitatory transmitter for the histamine neurons, suggesting that the two populations cooperate in the regulation of rapid eye movement sleep. Targeted ablation of orexin neurons in the lateral hypothalamus of rats by means of a hypocretin receptor 2saporin conjugate produced narcoleptic-like sleep behavior, closely paralleling the clinical findings and selected loss of hypocretin-orexin neurons in the lateral hypothalamus of humans with narcolepsy. These new discoveries add to the list of other neuropeptides including GHRH, somatostatin, and cortistatin with established function in modulation of the sleep cycle.

### Table 7-15 — Neurologic Manifestations of Nonendocrine Hypothalamic Disease

<table>
<thead>
<tr>
<th>Disorders of Temperature Regulation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyperthermia</td>
<td></td>
</tr>
<tr>
<td>Hypothermia</td>
<td></td>
</tr>
<tr>
<td>Poikilothermia</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Disorders of Food Intake</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyperphagia (bulimia)</td>
<td></td>
</tr>
<tr>
<td>Anorexia, aphagia</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Disorders of Water Intake</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Compulsive water drinking</td>
<td></td>
</tr>
<tr>
<td>Adipsia</td>
<td></td>
</tr>
<tr>
<td>Essential hyponatremia</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Disorders of Sleep and Consciousness</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Narcolepsy</td>
<td></td>
</tr>
<tr>
<td>Somnolence</td>
<td></td>
</tr>
<tr>
<td>Sleep rhythm reversal</td>
<td></td>
</tr>
<tr>
<td>Akinesis mutism</td>
<td></td>
</tr>
<tr>
<td>Coma</td>
<td></td>
</tr>
<tr>
<td>Delirium</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Periodic Disease of Hypothalamic Origin</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Diencetoic epilepsy</td>
<td></td>
</tr>
<tr>
<td>Kleine-Levin syndrome</td>
<td></td>
</tr>
<tr>
<td>Periodic discharge syndrome of Wolff</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hereditary Hypothalamic Disease</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Laurence-Moon-Biedli syndrome</td>
<td></td>
</tr>
<tr>
<td>Prader-Willi syndrome</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Disorders of Psychic Function</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Rage behavior</td>
<td></td>
</tr>
<tr>
<td>Hallucinations</td>
<td></td>
</tr>
<tr>
<td>Hypersexuality</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Disorders of Autonomic Nervous System</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Pulmonary edema</td>
<td></td>
</tr>
<tr>
<td>Cardiac arrhythmias</td>
<td></td>
</tr>
<tr>
<td>Sphincter disturbance</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Miscellaneous</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Diencephalic syndrome of infancy</td>
<td></td>
</tr>
<tr>
<td>Cerebral gigantism</td>
<td></td>
</tr>
</tbody>
</table>

Table of contents.
ACKNOWLEDGEMENT

The authors are highly indebted to Dr. Seymour Reichlin, not only for text and figures he shared from the ninth edition of this text, but also for the inspiration and mentorship he provided to the current generation of neuroendocrinologists.
References


References


183. Piko G, Koch Y, Ben Anoya N. Release of thyrotropin releasing hormone into hypophysial portal blood is high relative to other neuropeptides and may be related to prolactin secretion. Brain Res 1982; 243:186189.


References


323. Hsu SY, Hueth AJW. Human stresscopin and stresscopin-related peptide are selective ligands for the type 2 corticotropin-releasing hormone receptor. Nat Med 2001; 7:605611.


165


164. Watts AG. The impact of physiological stimuli on the expression of corticotropin-releasing hormone (CRH) and other neuropeptide genes. Front Neuroendocrinol 1996; 17:281326.


Van Cauter E, Leproult R, Plat L. Age-related changes in slow wave sleep and REM sleep and relationship with growth hormone and cortisol levels in healthy men. JAMA 2000; 284:861868.


References


561. Crowder M, Nett T. Pituitary content of gonadotropins and receptors for gonadotropin-releasing hormone (GnRHa) and hypothalamic content of GnRHa during the preovulatory period of the ewe. Endocrinology 1984; 114:23423.


References


Fujikawa K, et al. Significant body composition changes observed in obese subjects receiving chronic subcutaneous administration of a modified form of recombinant human leptin (abstract). North
218:325336.


Neurosci 1997; 17:71667179.


90:76817685.

258:E163E171.


100:27772782.

2000; 85:40034009.


734. Banks WA, Ortiz L, Plotkin SR, Kastin AJ. Human interleukin (IL) 1 alpha, murine IL-1 alpha and murine IL-1 beta are transported from blood to brain in the mouse by a shared saturatable mechanism. J Pharmacol Exp Ther 1991; 256:988996.


References


Chapter 8 - Anterior Pituitary

Shlomo Melmed
David Kleinberg

DEVELOPMENT, ANATOMY, AND OVERVIEW OF CONTROL OF HORMONE SECRETION

The pituitary gland situated within the sella turcica derives its name from the Greek *pituis*, Latin *pituita*, phlegm, reflecting its nasopharyngeal origin. Galen hypothesized that nasal phlegm originated from the brain and drained through the pituitary gland. It is now clear that together with the hypothalamus the pituitary orchestrates the structural integrity and function of endocrine glands, including the thyroid gland, adrenal gland, and gonads, in addition to target tissues including cartilage and breast. The pituitary stalk serves as an anatomic and functional connection to the hypothalamus. Preservation of the hypothalamic-pituitary unit is critical for integration of anterior pituitary control of sexual function and fertility, linear and organ growth, lactation, stress responses, energy, appetite, and temperature regulation and secondarily for carbohydrate and mineral metabolism.

Integration of vital body functions by the brain was first proposed by Descartes in the 17th century. In 1733, Morgagni recorded the absence of adrenal glands in an anencephalic neonate, providing early evidence for functional and developmental connection between the brain and the adrenal glands. In 1849 Claude Bernard set the stage for the subsequent advances in neuroendocrinology by demonstrating that central lesions in the area of the fourth ventricle resulted in polyuria. Subsequent studies led to the identification and chemical isolation of pituitary hormones, and astute clinical observations led to the realization that pituitary tumors were associated with functional hypersecretory syndromes, including acromegaly and Cushing's disease.

In 1948 Geoffrey Harris, the founder of modern neuroendocrinology, in reviewing anterior pituitary gland hormone control, proposed their hypothalamic regulation, predicting the subsequent discovery of specific hypothalamic regulating hormones.

Anatomy

The pituitary gland comprises the predominant anterior lobe, the posterior lobe, and a vestigial intermediate lobe (Fig. 8-1) (Figure Not Available). The gland is situated within the bony sella turcica and is overlain by the dural diaphragma sella, through which the stalk connects to the median eminence of the hypothalamus. The adult pituitary weighs approximately 600 mg (range, 400 to 900 mg) and measures about 13 mm in the longest transverse diameter, 6 to 9 mm in vertical height, and about 9 mm antero-posteriorly. Structural variation may occur in multiparous women, and gland volume also changes during the menstrual cycle. During pregnancy these measurements may be increased in either dimension, with pituitary weight increasing up to 1 g. Normal pituitary hypertrophy without evidence for the presence of an adenoma was described in seven eugonadal women with pituitary height greater than 9 mm and a convex upper gland boundary observed on magnetic resonance imaging (MRI).

The sella turcica located at the base of the skull forms the thin bony roof of the sphenoid sinus. The lateral walls comprising either bone or dural tissue abut the cavernous sinuses, which are traversed by the third, fourth, and sixth cranial nerves and internal carotid arteries (Fig. 8-2). The cavernous sinuses are vulnerable to pressure effects by an expanding pituitary mass, which is likely to follow the path of least tissue resistance by lifting the diaphragma sella (Fig. 8-3). The sinus contents are vulnerable to increased intrasellar expansion. The dural roofing protects the gland from compression by fluctuant cerebrospinal fluid (CSF) pressure. The optic chiasm, located anterior to the pituitary stalk, is directly above the diaphragma sella. The optic tracts and central structures are therefore vulnerable to pressure effects by an expanding pituitary mass, which is likely to follow the path of least tissue resistance by lifting the diaphragma sella (Fig. 8-3) (Figure Not Available). The intimate relationship of the pituitary and chiasm is borne out in optic chiasmal hypoplasia associated with developmental pituitary dysfunction seen in patients with septo-optic dysplasia. The posterior pituitary gland, in contrast to the anterior pituitary, is directly innervated by suprapontocerebellar hypophysial and tubero-hypophysial nerve tracts of the posterior stalk. Hypothalamic neuronal lesions, stalk disruption, or direct systemically derived metastases are therefore often associated with attenuated vasopressin (diabetes insipidus) or oxytocin secretion, or both.

The hypothalamus contains nerve cell bodies that synthesize hypophysiotropic releasing and inhibiting hormones as well as the neurohypophyseal hormones of the posterior pituitary (arginine vasopressin and oxytocin). Five distinct hormone-secreting cell types are present in the mature anterior pituitary gland. Corticotroph cells express pro-opiomelanocortin (POMC) peptides including adrenocorticotropic hormone (ACTH); somatotroph cells express growth hormone (GH); thyrotrroph cells express the common glycoprotein subunit and the specific thyroid-stimulating hormone (TSH) subunit; gonadotrophs express the and subunits for both follicle-stimulating hormone (FSH) and luteinizing hormone (LH).

Figure 8-2 (Figure Not Available) Coronal section of the sellar structures and cavernous sinus showing the relationship of the oculomotor (III), trochlear (IV), trigeminal ophthalmic and maxillary divisions (V1 and V2), and abducens (VI) cranial nerves to the pituitary gland. (From Sliter SL, Sharpe JA. Neuro-ophthalmologic evaluation of pituitary tumors. In Thapar K, Kovacs K, Schilthauer BW, Lloyd RV [eds]. Diagnosis and Management of Pituitary Tumors. Totowa, NJ, Humana Press, 2001, pp 173200.)
Pituitary Development

The pituitary gland arises from within the rostral neural plate. Rathke's pouch, a primitive ectodermal invagination anterior to the roof of the oral cavity, is formed by the fourth to fifth week of gestation and gives rise to the anterior pituitary gland (Fig. 8-4) (Figure Not Available).

Rathke's pouch is directly connected to the stalk and hypothalamic infundibulum and ultimately becomes distinct from the oral cavity and nasopharynx. Rathke's pouch proliferates toward the third ventricle, where it fuses with the diverticulum and subsequently obliterates its lumen, which may persist as Rathke's cleft. The anterior lobe is formed from Rathke's pouch, and the diverticulum gives rise to the adjacent posterior lobe. Remnants of pituitary tissue may persist in the nasopharyngeal midline and rarely give rise to functional ectopic hormone-secreting tumors in the nasopharynx. The neurohypophysis arises from neural ectoderm associated with third-ventricle development.

Functional development of the anterior pituitary cell types involves complex spatiotemporal regulation of cell lineagespecific transcription factors expressed in pluripotential pituitary stem cells as well as dynamic gradients of locally acting soluble factors. Critical neuroectodermal signals for organizing the dorsal gradient of pituitary morphogenesis include infundibular bone morphogenetic protein 4 (BMP4) required for the initial pouch invagination, fibroblast growth factor 8 (FGF-8), Wnt 5, and Wnt 4. Subsequent ventral developmental patterning and transcription factor expression are determined by spatial and graded expression of BMP2 and sonic hedgehog protein (shh), which appears critical for directing early patterns of cell proliferation.

The human fetal Rathke pouch is evident at 3 weeks, and the pituitary grows rapidly in utero. By 7 weeks, the anterior pituitary vasculature begins to develop, and by 20 weeks the entire hypophyseal-portal system is already established. The anterior pituitary undergoes major cellular differentiation during the first 12 weeks, by which time all the major secretory cell compartments are structurally and functionally intact, except for lactotrophs. Totipotential pituitary stem cells give rise to acidophilic (mammosomatotroph, somatotroph, and lactotroph) and basophilic (corticotroph, thyrotroph, and gonadotroph) differentiated pituitary cell types, which appear at clearly demarcated developmental stages. Corticotroph cells are morphologically identifiable at 6 weeks, and immunoreactive ACTH is detectable by 7 weeks. At 8 weeks, somatotroph cells are evident with abundant immunoreactive cytoplasmic GH expression. Glycoprotein hormonesecreting cells express a common subunit, and at 12 weeks differentiated thyrotrophs and gonadotrophs express immunoreactive subunits for TSH, LH, and FSH. Interestingly, gonadotrophs expressing LH and FSH are equally distributed in females, whereas in the male fetus, LH-expressing gonadotrophs predominate. Fully differentiated PRL-expressing lactotrophs are evident only late in gestation (after 24 weeks). Prior to that time, immunoreactive PRL is detectable only in mixed mammosomatotrophs, also expressing GH, reflecting the common genetic origin of these two hormones.
Pituitary Transcription Factors

Determination of anterior pituitary cell type lineages results from a temporally regulated cascade of homeodomain transcription factors. Although most pituitary developmental information has been acquired from murine models, histologic and pathogenetic observations in human subjects have largely corroborated these developmental mechanisms (see Fig. 8-4). Early cell differentiation requires intracellular Rpx and Ptx expression. Rathke’s pouch expresses several transcription factors of the LIM homeodomain family, including Lhx3, Lhx4, and IsI-1, which are early determinants of functional pituitary development. Pitx1 is expressed in the oral ectoderm and subsequently in all pituitary cell types, particularly those arising ventrally. Rieger’s syndrome, characterized by defective eye, tooth, umbilical cord, and pituitary development, is caused by defective related Pitx2.

Ptx behaves as a universal pituitary regulator and activates transcription of -GSU (the -subunit of gonadotroph hormones), POMC and LH (Ptx1), and GH (Ptx2). Lhx3 determines GH-PRL and TSH cell differentiation, and Prop-1 behaves as a prerequisite for Pit-1, which activates GH, PRL, TSH, and growth hormone-releasing hormone (GHRH) receptor transcription. TSH and gonadotropin-expressing cells share a common subunit (GSU) expression under developmental control of GATA-2. These specific anterior pituitary transcription factors participate in a highly orchestrated cascade leading to the commitment of the five differentiated cell types (see Fig. 8-4). The major proximal determinant of pituitary cell lineage derived from a totipotential stem cell is thus Prop-1 expression, which determines subsequent development of Pit-1-dependent and gonadotroph cell lineages.

POU1F1, the renamed Pit-1, is a POU-homeodomain transcription factor that determines development and appropriate temporal and spatial expression of cells committed to GH, PRL, TSH, and GHRH receptor expression. POU1F1 binds to specific deoxyribonucleic acid (DNA) motifs and activates somatotroph, lactotroph, and thyrotroph development and mature secretory function. Signal-dependent coactivating factors also cooperate with Pit-1 to determine specific hormone expression. Thus, in POU1F1-containing cells, high estrogen receptor levels induce a commitment to express PRL, whereas thyrotroph embryonic factor (TEF) favors TSH expression. Selective pituitary cell type specificity is also perpetuated by binding of POU1F1 to its own DNA regulatory elements as well as those contained within the GH, PRL, and TSH genes. Steroidogenic factor-1 (SF-1) and dosage-sensitive sex reversal, adrenal hypoplasia congenita, X-chromosome factor (DAX-1) determine subsequent gonadotroph development.

Corticotroph cell commitment, although occurring earliest during fetal development, is independent of POU1F1-determined lineages, and Tpit protein appears to be a prerequisite for POMC expression. Hereditary mutations arising within these transcription factors may result in isolated or combined pituitary hormone failure syndromes (see later).
Pituitary Blood Supply

The pituitary gland enjoys an abundant blood supply derived from several sources (see Fig. 8-1). The superior hypophyseal arteries branch from the internal carotid arteries to supply the hypothalamus, where they form a capillary network in the median eminence, external to the blood-brain barrier. Long and short hypophyseal portal vessels originate from infundibular plexuses and the stalk, respectively. These vessels form the hypothalamic-portal circulation, the predominant blood supply to the anterior pituitary gland. They deliver hypothalamic releasing and inhibiting hormones to the trophic hormone-producing cells of the adenohypophysis without significant systemic dilution, allowing the pituitary cells to be sensitively regulated by timed hypothalamic hormone secretion.

Vascular transport of hypothalamic hormones is also locally regulated by a contractile internal capillary plexus (gomotoli) derived from stalk branches of the superior hypophyseal arteries. Retrograde blood flow toward the median eminence also occurs, facilitating bidirectional functional hypothalamic-pituitary interactions. Systemic arterial blood supply is maintained by inferior hypophyseal arterial branches, which predominantly supply the posterior pituitary. Disruption of stalk integrity may lead to compromised pituitary portal blood flow, depriving the anterior pituitary cells of hypothalamic hormone access.
Pituitary Control

Three levels of control subserve the regulation of anterior pituitary hormone secretion. Hypothalamic control is mediated by adenohypophysiotropic hormones secreted into the portal system and impinging directly upon anterior pituitary cell surface receptors. G protein-linked cell surface membrane binding sites are highly selective and specific for each of the hypothalamic hormones and elicit positive or negative signals mediating pituitary hormone gene transcription and secretion. Peripheral hormones also participate in mediating pituitary cell function, predominantly by negative feedback regulation of trophic hormones by their respective target hormones. Intrapituitary paracrine and autocrine soluble growth factors and cytokines act locally to regulate neighboring cell development and function.

The net result of these three tiers of complex intracellular signals is the controlled pulsatile secretion of the six pituitary trophic hormones, ACTH, GH, PRL, TSH, FSH, and LH, through the cavernous sinus, petrosal veins, and ultimately the systemic circulation through the superior vena cava. The temporal and quantitative control of pituitary hormone secretion is critical for physiologic integration of peripheral hormonal systems such as the menstrual cycle, which relies on complex and precisely regulated pulse control.
PITUITARY MASSES

Pituitary Mass Effects

An expanding pituitary mass may inexorably alter the sellar size and shape by bone erosion and remodeling. Although the exact time course of this process is unknown, it appears to be slowly progressive over years or decades. The tumor may invaginate soft tissue, and the dorsal sellar roof presents the least resistance to expansion from within the confines of the bony sella. Nevertheless, both suprasellar and parasellar compression and invasion may occur with an enlarging mass, with resultant clinical manifestations. As tumors impinge upon the optic chiasm, they interfere with vision. Because of the anatomy of the chiasm, pressure from below affects temporal visual fields, starting superiority and ultimately extending to the entire temporal field. Loss of nasal fields also occurs and may result in blindness. Long-standing optic chiasmal pressure results in optic disc pallor.

Figure 8-7 Magnetic resonance coronal section of a normal pituitary gland (top). A large pituitary adenoma is seen lifting and distorting the optic chiasm (arrow) and is also invading the sphenoid sinus (middle). A sagittal section of a large macroadenoma with bone invasion and impinging brain structures is shown (bottom).

Lateral invasion of pituitary lesions may invade the dural wall of the cavernous sinus affecting the third, fourth, and sixth cranial nerves as well as the ophthalmic and maxillary branches of the fifth cranial nerve and surround the internal carotid artery. Varying degrees of diplopia, ptosis, ophthalmoplegia, and decreased facial sensation may infrequently occur, depending on the extent of the neural involvement by the cavernous sinus mass. Downward extension into the sphenoid sinus indicates that the parasellar mass has eroded the bony sellar floor. Aggressive tumors may invade the roof of the palate and cause nasopharyngeal obstruction, infection, and CSF leakage. Infrequently, temporal or frontal lobes may be invaded, causing uncinate seizures, personality disorders, and anosmia. In addition to the anatomic lesions caused by the expanding mass, direct hypothalamic involvement of the encroaching mass may lead to important metabolic sequelae discussed in Chapter 7.

Patients with intrasellar tumors commonly present with headaches, even in the absence of demonstrable suprasellar extension. Small changes in intrasellar pressure caused by a microadenoma within the confined sella are sufficient to stretch the dural plate with resultant headache. Headache severity does not correlate with the size of the adenoma or the presence of suprasellar extension. Relatively minor diaphragmatic distortions or dural impingement may be associated with persistent headache. Successful medical management of small functional pituitary tumors with dopamine agonists or somatostatin analogues is often accompanied by a remarkable improvement in or disappearance of headache.

Regardless of their etiology or size, pituitary masses, including adenomas, may be associated with compression of surrounding healthy tissue and resultant hypopituitarism. In 49 patients undergoing transsphenoidal resection of pituitary adenomas, mean intrasellar pressure was elevated twofold to threefold in patients with pituitary failure. Furthermore, prevalence of headache and elevated PRL levels correlated positively with intrasellar pressure levels, suggesting interrupted portal delivery of hypothalamic hormones. Thus, surgical decompression of a sellar mass may lead to recovery of compromised anterior pituitary function. In the patients who do not recover pituitary function postoperatively, ischemic necrosis is likely to have occurred. Stalk compression may result in pituitary failure caused by encroachment of the portal vessels that normally provide pituitary access to the hypothalamic hormones. Stalk compression also usually leads to hyperprolactinemia and concomitant failure of other pituitary trophic hormones.
Pituitary Adenomas

Pathogenesis

Pituitary tumors account for about 15% of all intracranial neoplasms and are commonly encountered at autopsy. The Brain Tumor Registry of Japan reported that 15.8% of 28,424

<table>
<thead>
<tr>
<th>Affected Structure</th>
<th>Clinical Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pituitary</td>
<td>Growth failure</td>
</tr>
<tr>
<td></td>
<td>Adult hyposomatotrophism</td>
</tr>
<tr>
<td></td>
<td>Hypogonadism</td>
</tr>
<tr>
<td></td>
<td>Hypothyroidism</td>
</tr>
<tr>
<td></td>
<td>Hypoadrenalism</td>
</tr>
<tr>
<td>Optic tract</td>
<td>Loss of red perception, bitemporal hemianopsia, superior or bitemporal field defect, scotoma, blindness</td>
</tr>
<tr>
<td>Hypothalamus</td>
<td>Temperature dysregulation, obesity, diabetes insipidus, thirst, sleep, appetite, behavioral and autonomic nervous system dysfunctions</td>
</tr>
<tr>
<td>Cavernous sinus</td>
<td>Plosis, diplopia, ophthalmoplegia, facial numbness</td>
</tr>
<tr>
<td>Temporal lobe</td>
<td>Uncinate seizures</td>
</tr>
<tr>
<td>Frontal lobe</td>
<td>Personality disorder, anosmia</td>
</tr>
<tr>
<td>Central</td>
<td>Headache, hydrocephalus, psychosis, dementia, laughing seizures</td>
</tr>
<tr>
<td>Neuro-ophthalmologic tract</td>
<td>Field defects</td>
</tr>
<tr>
<td></td>
<td>Bitemporal hemianopsia (50%)</td>
</tr>
<tr>
<td></td>
<td>Amaurosis with hemianopsia (12%)</td>
</tr>
<tr>
<td></td>
<td>Contralateral or monocular hemianopsia (7%)</td>
</tr>
<tr>
<td>Scotomas</td>
<td>Junctional; monocular central, arcuate, altitudinal; hemianopic</td>
</tr>
<tr>
<td></td>
<td>Homonymous hemianopsia</td>
</tr>
<tr>
<td>Acuity loss</td>
<td>Snellen</td>
</tr>
<tr>
<td></td>
<td>Contrast sensitivitiy</td>
</tr>
<tr>
<td></td>
<td>Color vision</td>
</tr>
<tr>
<td></td>
<td>Visual evoked potential</td>
</tr>
<tr>
<td>Pupillary abnormality</td>
<td>Impaired light reactivity</td>
</tr>
<tr>
<td></td>
<td>Afferent defect</td>
</tr>
<tr>
<td>Optic atrophy</td>
<td>Papilledema</td>
</tr>
<tr>
<td></td>
<td>Cranial nerve palsyoculomotor, trachlear, abducens, sensory trigeminal</td>
</tr>
<tr>
<td>Nystagmus</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Visual hallucinations</td>
</tr>
<tr>
<td></td>
<td>Postfixation blindness</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>TABLE 8-2 -- Factors Involved in Pituitary Tumor Pathogenesis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Heredity</strong></td>
</tr>
<tr>
<td>MEN-1</td>
</tr>
<tr>
<td>Transcription factor defect (e.g., Prop-1 excess)</td>
</tr>
<tr>
<td>Carney complex</td>
</tr>
<tr>
<td><strong>Hypothalamic</strong></td>
</tr>
<tr>
<td>Excess GHRH or CRH production</td>
</tr>
<tr>
<td>Receptor activation?</td>
</tr>
<tr>
<td>Dopamine deprivation?</td>
</tr>
<tr>
<td><strong>Pituitary</strong></td>
</tr>
<tr>
<td>Signal transduction mutations (e.g., gsp, CREB)</td>
</tr>
<tr>
<td>Disrupted paracrine growth factor or cytokine action (e.g., FGF2, FGF4, LIF, EGF, NGF)</td>
</tr>
<tr>
<td>Activated oncogene or cell cycle disruption (e.g., PTTG; ras; p27)</td>
</tr>
<tr>
<td>Intrapituitary paracrine hypothalamic hormone action (e.g., GHRH, TRH)</td>
</tr>
</tbody>
</table>
cases were historically confirmed pituitary adenomas. They are benign monoclonal adenomas that may express and secrete hormones autonomously, leading to hyperprolactinemia, acromegaly, and Cushing's disease, or may be functionally silent and initially diagnosed as a sellar mass. Although these adenomas are invariably benign, their neoplastic features represent a unique tumor biology that is reflected in their important local and systemic manifestations. These monoclonal neoplasms have a slow doubling time and, if small, may rarely resolve spontaneously. Nevertheless, they can be aggressive and locally invasive or compressive to vital central structures. They usually express a single gene product, but polyclonal expression may reflect a primitive stem cell or mature bimorphic cellular origin. Hypothalamic factors may have a specific role in the pathogenesis of pituitary tumors, in addition to regulating pituitary hormone gene expression and secretion (Table 8-2). Ectopic GHRH-secreting tumors (bronchial carcinoids, pancreatic islet cell tumors, or small cell lung carcinomas) result in GH hypersecretion, acromegaly, somatotroph hyperplasia, and occasionally somatotroph adenoma formation. In transgenic mice overexpressing a GHRH transgene, the pituitary size increased dramatically because of somatotroph hyperplasia, and older mice developed GH-secreting adenomas. However, adenomatous hormonal secretion is usually independent of physiologic hypothalamic control, and the surgical resection of small well-defined adenomas usually results in definitive cure of hormonal hypersecretion. These observations imply that these tumors do not arise because of excessive polyclonal pituitary cell proliferation related to generalized hypothalamic stimulation. However, hypothalamic factors may promote and maintain growth of already transformed pituitary adenomatous cells.

Normal and hyperplastic pituitary tissues are polyclonal, and pituitary adenomas arise as the result of monoclonal pituitary cell proliferation. Using X-chromosomal inactivation analysis, the monoclonal origin of adenomas secreting GH, PRL, and ACTH is confirmed and nonfunctioning pituitary tumors was confirmed in female patients heterozygous for variant alleles of the X-linked genes hypoxanthine phosphoribosyltransferase (HPRT) and phosphoglycerate kinase (PGK). Thus, an intrinsic somatic pituitary cell genetic alteration probably gives rise to clonal expansion of a single cell, resulting in adenoma formation. (Table 8-3).

Activating gsp mutations are present in up to 40% of human GH-secreting adenomas. These somatic heterozygous activating point mutations of the G protein subunit (Gs) gene involving either arginine 201 (replaced by cysteine or histidine) or glutamine 227 (replaced with arginine or leucine) constitutively activate the Gs protein and convert it into an oncogene (gsp). This G protein activation increases cyclic adenosine monophosphate (cAMP) levels and activates protein kinase A, which in turn phosphorylates the cAMP response element binding protein (CREB) and leads to sustained constitutive GH hypersecretion and cell proliferation. The gsp-bearing adenomas are smaller, have mildly lower GH levels and enhanced intratumoral cAMP, do not respond briskly to GHRH, and are extremely sensitive to the inhibitory effect of somatostatin. These gsp activating mutations do not occur in PRL-secreting or in TSH-producing adenomas and are rarely present in nonfunctioning pituitary tumors or ACTH-secreting tumors (<10%).

Similar early postzygotic somatic mutations in codon 201 of Gs were identified in tissues derived from patients with McCune-Albright syndrome, including GH-producing pituitary adenomas. Transgenic mice overexpressing inactive pituitary CREB mutant exhibited a dwarf phenotype and somatotroph hypoplasia. Thus, cAMP probably stimulates somatotroph proliferation by CREB phosphorylation. This was borne out by the observation that 15 human GH-secreting pituitary adenomas contained elevated levels of phosphorylated CREB. However, only four of these tumors also contained the mutant gsp oncogene, and CREB phosphorylation was also demonstrated in adenomas overexpressing wild-type Gs protein, suggesting a role of CREB independent of G protein actions.

Ras mutations are rare in pituitary adenomas. Ras gene mutations were identified in an invasive prolactinoma and in distant metastatic pituitary carcinomas but not in their respective primary pituitary tumors or in noninvasive adenomas. Thus, ras genetic alterations may be important in the rare progression to metastasis formation and growth. Pituitary tumor transforming gene (PTTG) was isolated from experimental pituitary tumors and shown to be highly abundant in all pituitary tumors.

**TABLE 8-3 -- Candidate Genes in Pituitary Tumorigenesis**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Protein</th>
<th>Tumor Type</th>
<th>Mechanism of Overexpression or Inactivation</th>
<th>Function, Defect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gsp</td>
<td>GNAS</td>
<td>40% GH-secreting tumors</td>
<td>Point mutation</td>
<td>Signal transduction, elevated cAMP</td>
</tr>
<tr>
<td></td>
<td>McCune-Albright syndrome</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Minority other types</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTTG</td>
<td>PTTG</td>
<td>All pituitary tumors</td>
<td>Unknown Estrogen?</td>
<td>Chromatid separation, regulates bFGF secretion, disrupted cell cycle, chromosomal instability, bFGF-mediated mitogenesis and angiogenesis</td>
</tr>
<tr>
<td>Hst</td>
<td>FGF4</td>
<td>Large prolactinomas</td>
<td>Unknown</td>
<td>Angiogenesis, overexpression</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Enhanced PRL transcription</td>
</tr>
<tr>
<td>CREB</td>
<td>GH-secreting</td>
<td>Increased Ser-phosphorylated CREB promoted by gsp overexpression</td>
<td>Dimerizes with cAMP response elements</td>
<td></td>
</tr>
<tr>
<td>H-ras</td>
<td>Ras</td>
<td>Metastatic pituitary carcinoma only</td>
<td>Point mutation, amplification</td>
<td>Signal transduction, stimulates tyrosine kinase pathway</td>
</tr>
</tbody>
</table>

Multiple endocrine neoplasia type 1 (MEN-1) is an autosomal dominant hereditary disorder characterized by combined tumor formation or hyperfunction of pancreatic islets and anterior pituitary and, less commonly, parathyroid, thyroid, and adrenal tumors. The MEN-1 syndrome is fully described in Chapter 36. Unlike those in pituitary tumors constituting the MEN-1 syndrome, MEN-1 gene mutations were not identified in familial pituitary adenomas. Patients with sporadic pituitary adenomas do not demonstrate germ line or somatic pathogenic changes in the coding sequence of the MEN-1 gene, even in tumors with loss of heterozygosity of 11q13, and only two cases of sporadic pituitary adenomas (of 94 tumors studied)

<table>
<thead>
<tr>
<th>Syndrome</th>
<th>Clinical Features</th>
<th>Chromosomal Location</th>
<th>Gene</th>
<th>Protein</th>
<th>Proposed Function, Defect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple endocrine neoplasia type 1 (MEN-1)</td>
<td>Parathyroid, endocrine pancreas, anterior pituitary (mostly prolactinomas) tumors</td>
<td>11q13</td>
<td>Men1</td>
<td>Menin</td>
<td>Nuclear, tumor suppressor protein interacts with junD</td>
</tr>
<tr>
<td>Familiar acromegaly</td>
<td>GH-cell adenomas, acromegaly, gigantism</td>
<td>11q13 and other loci</td>
<td>Not Men1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>McCune-Albright syndrome</td>
<td>Polyostotic fibrous dysplasia, pigmented skin patches; endocrine abnormalities: precocious puberty, GH-cell adenomas, acromegaly, gigantism, Cushings syndrome</td>
<td>20q13.2 (mosaic)</td>
<td>GNAS1</td>
<td>Gs</td>
<td>Signal transduction, inactive GTPase results in constitutive cAMP elevation independent of GHRH</td>
</tr>
<tr>
<td>Carney’s syndrome</td>
<td>Skin and cardiac myxomas, Cushings syndrome, acromegaly</td>
<td>2p16</td>
<td></td>
<td></td>
<td>Protein kinase A signaling defect for activating GH</td>
</tr>
</tbody>
</table>


had specific MEN-1 mutations. Thus, MEN-1 gene mutations do not appear to play a role in pituitary tumorigenesis in most sporadic adenomas.

Loss of heterozygosity for chromosomes 11q13, 13, and 9 has been observed in about 15% of spontaneous pituitary adenomas, often correlating with tumor size and invasiveness. However, no distinct tumor suppressor gene has been identified for sporadic pituitary tumors. Loss of heterozygosity in proximity to the retinoblastoma (RB) locus on chromosome 13q14 was detected in malignant or highly invasive pituitary tumors and in their metastases, but immunohistochemical studies have shown the presence of RB protein in these malignant tumors with 13q14 allelic loss, suggesting that the RB gene itself is not involved in pituitary adenoma development and that another suppressor gene located near the RB locus may play a role in invasive or malignant pituitary tumors. p53 gene mutations were not detected in secreting and nonsecreting pituitary tumors or in pituitary carcinomas and their metastases.

Although no mutations in the GHRH, corticotropin-releasing hormone (CRH), thyrotropin-releasing hormone (TRH), or gonadotropin-releasing hormone (GnRH) receptor have been identified in pituitary adenomas, several GH-secreting adenomas express an alternatively spliced truncated GHRH receptor. The insulin-like growth factor I (IGF-I) receptor subunit in GH adenomas exhibited intact regions of the receptor critical for signal transduction. The dopamine D2 receptor gene appears intact in PRL-producing and TSH-producing or nonfunctioning adenomas. Therefore, there is no apparent role of pituitary cell surface receptor mutations of hypothalamic releasing and inhibitory factors in pituitary tumorigenesis.

FGF-2 (basic FGF) is expressed in pituitary tissues and induces basal and stimulated PRL secretion from normal and pituitary adenoma cells. Human pituitary adenomas express FGF-4, and transfected FGF-4 enhances PRL secretion and tumor vascularity. FGF-4 is immunodetected in about a third of prolactinomas and is undetectable in normal pituitaries and other adenoma types.

Carney’s complex is an autosomal dominant disorder comprising benign mesenchymal tumors including cardiac myxomas, schwannomas, and thyroid and pituitary adenomas associated with spotty skin pigmentation (Table 8-4). The disorder has been mapped to chromosome 17q24 and results from a mutated R1 regulatory subunit of the cAMP-dependent protein kinase A (PRKARIA), an apparent tumor suppressor gene.

In summary, multifactorial mechanisms subserve the multistep pathogenetic process of pituitary adenoma formation, including early initiating chromosomal mutations that result in mutated pituitary stem cells (see Table 8-2). The transformed pituitary cell is subjected to signals facilitating clonal expansion, and several permissive factors, including hypothalamic hormone receptor signals, intrapituitary growth factors, and disordered cell cycle regulation, may determine the ultimate biologic fate of the tumor. Autonomous anterior pituitary hormone production and secretion and cell proliferation, which are the hallmarks of pituitary adenomas, result. However, the subcellular events initiating the formation of most secreting and nonfunctional pituitary adenomas have not yet been elucidated.

### Classification

Pituitary adenomas arise from hormone-secreting adenohypophyseal cells, and their secretory products depend on the cell of origin (Table 8-5). Previously clinically inapparent pituitary adenomas are found in about 11% of autopsies (Table 8-6). They are often localized to unique areas of the gland, reflecting relative cell type abundance and intragland distribution (Fig. 8-8) (Figure Not Available). Radiologic and surgical classifications are based on tumor localization, size, and degree of invasiveness (Fig. 8-9) (Figure Not Available). Microadenomas are intrasellar and generally less than 10 mm in widest diameter. Macroadenomas are larger than 10 mm and usually impinge on adjacentellar structures. Specific tumor types are considered subsequently for each respective cell type.

Immunocytochemistry detects pituitary cell gene products at both the light and electron microscopic level and allows classification of pituitary tumors on the basis of their function. Unlike the corticotroph, somatotroph, lactotroph, and thyrotroph cell tumors, which hypersecrete their respective hormones, gonadotroph cell tumors are usually clinically silent and do not secrete their gene products efficiently. Double immunostaining identifies mixed tumors expressing combinations of hormones, which are often macroadenomas secreting GH concomitantly with PRL, TSH, or ACTH. Generally, immunohistochemical identification of pituitary hormones correlates with tumor-specific messenger ribonucleic acid (mRNA) markers measured either in whole tissue extracts by Northern analysis or at the single-cell level by in situ hybridization techniques. With the exception of the glycoprotein subunit, immunohistochemical positivity of more than 5% of cells making up the tumor is usually reflective of peripheral circulating hormone.

### Table 8-5 - Clinical and Pathologic Characteristics of Pituitary Adenomas

<table>
<thead>
<tr>
<th>Adenoma Type</th>
<th>Incidence (%)</th>
<th>Pathologic</th>
<th>Clinical Incidence (new cases/10^6/yr)</th>
<th>Prevalence (total 10^6)</th>
<th>mRNA Expression</th>
<th>Immunohistochemistry</th>
<th>EM Secretory Granules (nm)</th>
<th>Clinical Syndrome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lactotroph</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sparsely granulated</td>
<td>28</td>
<td>29</td>
<td>610</td>
<td>60100</td>
<td></td>
<td></td>
<td></td>
<td>PRL</td>
</tr>
</tbody>
</table>

Hyponogadism, galactorrhea
<table>
<thead>
<tr>
<th>Type</th>
<th>Count</th>
<th>GH</th>
<th>PRL</th>
<th>GH</th>
<th>PRL</th>
<th>Range</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Densely granulated</td>
<td>15</td>
<td>46</td>
<td>4060</td>
<td>15</td>
<td>46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sparsely granulated</td>
<td>5</td>
<td>100250</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Acromegaly or gigantism</td>
</tr>
<tr>
<td>Densely granulated</td>
<td>5</td>
<td>100250</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combined</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GH/PRL cells</td>
<td>5</td>
<td>100600</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hypogonadism</td>
</tr>
<tr>
<td>Mixed GH/PRL</td>
<td>1</td>
<td>3502000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Acromegaly</td>
</tr>
<tr>
<td>Mammosomatotroph</td>
<td>3</td>
<td>50300</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Galactorrhea</td>
</tr>
<tr>
<td>Corticotroph</td>
<td>23</td>
<td>2030</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cushing’s</td>
<td>10</td>
<td>250700</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cushing’s disease</td>
</tr>
<tr>
<td>Silent corticotroph</td>
<td>3</td>
<td>Variable</td>
<td>None</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nelson’s</td>
<td>2</td>
<td>250700</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Local signs</td>
</tr>
<tr>
<td>Thyrotroph</td>
<td>1</td>
<td>50250</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hyperthyroidism</td>
</tr>
<tr>
<td>Plurihormonal</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonfunctioning, null cell,</td>
<td>27</td>
<td>79</td>
<td>7090</td>
<td></td>
<td></td>
<td></td>
<td>Silent or pituitary failure</td>
</tr>
<tr>
<td>gonadotroph</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nononcocytic</td>
<td>14</td>
<td>FSH, LH SU</td>
<td>Glycoprotein</td>
<td>&lt;25% of cells</td>
<td>100250</td>
<td></td>
<td>Silent or pituitary failure</td>
</tr>
<tr>
<td>Oncocytic</td>
<td>6</td>
<td>FSH, LH SU</td>
<td>Glycoprotein</td>
<td>&lt;25% of cells</td>
<td>100250</td>
<td></td>
<td>Pituitary failure</td>
</tr>
<tr>
<td>Gonadotroph</td>
<td>715</td>
<td>FSH, LH</td>
<td>FSH, LH</td>
<td></td>
<td></td>
<td></td>
<td>Silent or pituitary failure</td>
</tr>
</tbody>
</table>

Data are derived from studying a relatively stable 1 million catchment population surrounding Stoke-on-Trent, UK (Clayton RN, Clin Endo & Metab 13:451, 1999), and from Kovacs & Horvath 1986; Scheithauer 1994; Minderman & Wilson, Clin Endo 1994; Asa 1993. In Endocrine Tumours, Blackwell). Levels. Quantification of immunostaining intensity is subjective, and a scale of intensity should also include a description of the extent of staining, that is, whether occasional, scattered, or most tumor cells express the immunodetectable protein.

Electron microscopy is useful for assessing the ultrastructure of hormone secretory granules and their size and distribution. Other subcellular features important for diagnosis include large mitochondria in nonfunctioning oncocytes and the secretory nature of Golgi and endoplasmic reticulum, especially for prolactinomas. Peroxidase or colloidal gold particles of different diameters are also sensitive electron microscopic markers for identifying and localizing intracellular hormone signals. Because even invasive pituitary tumors grow slowly, use of mitotic markers including proliferating cell nuclear antigen (PCNA) and Ki-67 is of limited utility. 

![Bookmark URL](167.html/top)
Other Parasellar Masses

Clinical features of hypothalamic masses are fully described in Chapter 7, and parasellar masses (Table 8-7) are described here. Teramoto

Rathke's Cyst

The anterior and intermediate lobes of the pituitary gland arise embryologically from Rathke's pouch. Inadequate pouch obliteration results in the cysts or cystic remnants at the interface between the anterior and posterior pituitary lobes found in about 20% of pituitary glands at autopsy. Pituitary adenomas may also occasionally contain small cleft cysts. They are lined by cuboidal or columnar ciliated epithelium surrounding mucoid cyst fluid, arising from midline rudiments of failed Rathke's cyst invagination, and account for about 3% of pituitary mass lesions. In contrast, pituitary epidermoid cysts are lined by squamous epithelium, which rarely becomes malignant.

Rathke's cysts vary in size and may also extend to the supra-sellar region. These lesions have heterogeneous MRI characteristics and may arise with panhypopituitarism with or without diabetes insipidus. Most, however, are not symptomatic and should be observed expectantly. The extent of headache or visual disturbance is determined by the size and location of the cyst. Cyst formation is associated with sellar enlargement and hyperdense or hypodense masses seen on either T1-weighted or T2-weighted MR images, and computed tomography (CT) shows homogeneous hypodense areas that may be distinguished from pituitary adenomas. These patients should all be evaluated for hypopituitarism. After surgical resection or drainage, MRI should be performed during long-term follow-up for signs of cyst recurrence.

Arachnoid, epidermoid, and dermoid cysts develop mainly in the cerebellopontine angle but may also arise in the suprasellar region. Dermoid cysts containing greasy sebaceous products or hair follicles are rarely encountered in the pituitary, and the cyst lining may be calcified. Acquired pituitary cysts may be secondary to intrasellar hemorrhage, usually associated with an underlying adenoma, and these rarely cause pituitary failure.

---

**TABLE 8-6 -- Frequency of Pituitary Adenomas Found at Autopsy**

<table>
<thead>
<tr>
<th>Study</th>
<th>No. of Pituitaries Examined</th>
<th>No. of Adenomas Found</th>
<th>Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Susman</td>
<td>260</td>
<td>23</td>
<td>9</td>
</tr>
<tr>
<td>Costello</td>
<td>1000</td>
<td>225</td>
<td>23</td>
</tr>
<tr>
<td>Sommers</td>
<td>400</td>
<td>26</td>
<td>7</td>
</tr>
<tr>
<td>McCormick</td>
<td>1600</td>
<td>140</td>
<td>9</td>
</tr>
<tr>
<td>Kovacs</td>
<td>152</td>
<td>20</td>
<td>13</td>
</tr>
<tr>
<td>Landolt</td>
<td>100</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>Mosca</td>
<td>100</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>Burrow</td>
<td>120</td>
<td>32</td>
<td>27</td>
</tr>
<tr>
<td>Parent</td>
<td>500</td>
<td>42</td>
<td>8</td>
</tr>
<tr>
<td>Muhr</td>
<td>205</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Schwezinger</td>
<td>5100</td>
<td>485</td>
<td>9</td>
</tr>
<tr>
<td>Coulon</td>
<td>100</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Chambers</td>
<td>100</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>Siqueira</td>
<td>450</td>
<td>39</td>
<td>9</td>
</tr>
<tr>
<td>El-Hamid</td>
<td>486</td>
<td>97</td>
<td>20</td>
</tr>
<tr>
<td>Scheltlaufer</td>
<td>251</td>
<td>41</td>
<td>16</td>
</tr>
<tr>
<td>Marin</td>
<td>210</td>
<td>35</td>
<td>16</td>
</tr>
<tr>
<td>Mosca</td>
<td>111</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>Sano</td>
<td>166</td>
<td>15</td>
<td>9</td>
</tr>
<tr>
<td>Teramoto</td>
<td>1000</td>
<td>51</td>
<td>5</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>12,411</strong></td>
<td><strong>1408</strong></td>
<td><strong>11</strong></td>
</tr>
</tbody>
</table>


compression causes internal hydrocephalus, visual disturbances, GH or ACTH deficiency, hyperprolactinemia, and diabetes insipidus. Rarely, squamous cell carcinoma may arise in the cyst.

Granular Cell Tumors

Pituitary choristomas, or schwannomas, usually arise only after the age of 20. Their abundant cytoplasmic granules do not contain pituitary hormones, but these lesions may occur with diabetes insipidus. Pituitary adenomas are occasionally coincidentally associated with these tumors.

Chordomas

These slow-growing cartilaginous tumors arise from midline notochord remnants, are locally invasive, and may metastasize. Most arise from the vertebrae, and about one third involve the clivus region. Chordomas contain a mucin-rich matrix that allows diagnosis by fine-needle aspiration. Patients present with headaches, asymmetric visual disturbances, hormone deficiency, and occasional nasopharyngeal obstruction. The tumor mass is associated with osteolytic bone erosion and calcification, and MRI may allow the normal pituitary gland to be distinguished from the very heterogeneous and often floroculent tumor mass. At surgery, the tumors are rough, heterogeneous, and lobular. Markers for epithelial cells, including cytokeratin and vimentin, are present. Recurrences are common after surgical excision, and mean survival of patients is about 5 years. Rarely, chordomas undergo sarcomatous transformation with an aggressive natural history and require extensive surgical dissection.
Craniopharyngiomas

This parasellar tumor constitutes about 3% of all intracranial tumors and up to 10% of childhood brain tumors. The tumor is commonly diagnosed during childhood and adolescence. Tumors arise from embryonic squamous remnants of Rathke's pouch extending dorsally toward the diencephalon; they may be large (> 10 cm in diameter) and invade the third ventricle and associated brain structures. Over 60% arise from within the sella, with others arising from parasellar cell rests. When intrasellar, they can often be distinguished from pituitary adenomas by a separate visible rim of normal pituitary tissue seen on MRI. The cystic mass is usually filled with cholesterol-rich viscous fluid, which may leak into the CSF and cause aseptic meningitis. The tumors may also contain calcifications and immunoreactive human chorionic gonadotropin (hCG).

Histologically, these tumors comprise two cell populations; cysts are lined with a squamous epithelium containing islands characterized by columnar cells, and a mixed inflammatory reaction may also occur with calcification. Although large craniopharyngiomas may obstruct CSF flow, they rarely undergo malignant transformation. Increased intracranial pressure results in headache, projectile vomiting, papilledema, and somnolence, especially in children. Only about one third of patients are older than 40 years, and they commonly present with asymmetric visual disturbances, including papilledema, optic atrophy, and field deficits. If cavernous sinus invasion is present, other cranial nerves may also be involved.

On CT imaging, most children and about half of all adults exhibit characteristic flocculent or convex calcifications. Rarely, however, pituitary adenomas, other parasellar tumors, and vascular lesions within the sella are also calcified. In contrast to pituitary adenomas, which rarely cause diabetes insipidus, craniopharyngiomas are often associated with this disorder as the earliest feature. These patients may also experience partial or complete pituitary deficiency. GH deficiency, with short stature and diabetes insipidus, and gonadal failure are common. Pituitary stalk compression or damage to hypothalamic dopaminergic neurons results in hyperprolactinemia. Thus, a craniopharyngioma may mimic a prolactinoma in terms of intrapituitary imaging, presence of hyperprolactinemia, and favorable biochemical response to dopamine agonists.

The treatment of these lesions may involve radical surgery, radiotherapy, or a combination of these modalities. Stereotactic radiation has some success. A detailed discussion of the neurosurgical management of this disorder is beyond the scope of this text. Nevertheless, regardless of the form of therapy chosen, ablation of the mass invariably results in anterior or posterior pituitary hormone deficiencies. Postoperative recurrence may occur in about 20% of patients undergoing radical surgical excision, and there is no difference in outcome in those who undergo a subtotal surgical excision followed by radiotherapy. Pure papillary squamous cellular elements in the tumor may portend a higher surgical recurrence rate. Long-term effects of childhood radiation for these tumors are considered elsewhere.

Meningiomas

Meningiomas arise from arachnoid and meningioendothelial cells, and those occurring in the sellar and parasellar region account for about one fifth of all meningiomas. Sellar meningiomas are usually well circumscribed and do not attain the size of craniopharyngiomas. Suprasellar meningiomas may invade the pituitary ventrally, and intrasellar tumor origins are rare. Coexisting functional pituitary adenomas have been described in patients with parasellar meningiomas. Secondary hypopituitarism and visual deficits occur in up to half of patients, who usually present with local mass effects including headache and progressive visual disturbances accompanied by optic atrophy.

The differential distinction between a suprasellar meningioma with downward extension and an upwardly extending pituitary adenoma may be difficult. With MRI, meningiomas are isodense on both T1 and T2 imaging, in contrast to other parasellar lesions, which are usually hyperdense on T2 imaging. Dural calcification may be evident on CT scanning. Because of their rich vascularization, these tumors pose an intracranial risk for hemorrhage and a resultant higher surgical mortality rate than usually encountered for pituitary adenoma resection.

Gliomas

Optic gliomas and low-grade astrocytomas arise from within the optic chiasm or optic tract and often infiltrate the optic nerve; less than one third are intraorbital. Von Recklinghausen's disease is the underlying cause in about one third of the patients. These tumors may occasionally be associated with growth retardation, delayed or precocious puberty, and mass effects including visual disturbances, diencephalic syndrome, diabetes insipidus, and hydrocephalus. Rarely, gliomas arise within the sella associated with hyperprolactinemia, and they should be considered in the uncommon differential diagnosis of a PRL-secreting pituitary adenoma. Important distinguishing features include the young age of these patients (80% are younger than 10 years), relatively intact pituitary function, gross visual disturbances, and localization of the mass as visualized on MRI. Gliomas, unlike hamartomas, are usually enhanced after injection of contrast material.

Mucocles

Mucocles are expanding accumulations of fluid within the sphenoid sinus and may compress parasellar structures. Headaches, visual disturbances (usually unilateral), and exophthalmos are characteristic features. On MRI, the homogeneous sphenoid mass may be quite prominent but may be distinguished from the pituitary gland dorsally.

Parasellar Aneurysms

Parasellar aneurysms may mimic pituitary adenomas and intraoperative rupture may be catastrophic, underlying the absolute need for preoperative diagnosis. Differentiating features of aneurysms from other parasellar masses may be subtle, including eye pain, intense headaches, and relatively sudden onset of cranial nerve palsies. Although imaging techniques usually distinguish blood and hemorrhage from solid tumor or tissue, a highly vascular meningioma may be confused with an aneurysm.

Pituitary Infections

Acute pituitary abscesses and periselar arachnoiditis are encountered with sinus infections, especially after transsphenoidal surgery. Pituitary abscess may develop from hematogenous or direct local spread of infectious agents. Abscesses may arise within a preexisting pituitary adenoma and may be difficult to distinguish from an adenoma because these patients may not have fever or signs of meningitis. On MRI imaging, an isointense central cavity with surrounding ring enhancement is characteristic of an abscess.

Gram-positive streptococci or staphylococci may originate from nasopharyngeal passages. Disseminated Entamoeba histolytica and Pneumocystis carinii may also seed to the pituitary.

Figure 8-9 (Figure Not Available) Classification of pituitary tumors. (Adapted from Thapar K, Laws ER: Growth hormonesecreting pituitary tumors: operative management. In Krisht AF, Tindal GT [eds]: Pituitary Disorders. Philadelphia, Lippincott Williams & Wilkins, 1999; and Aa SL: In Tumors of the Pituitary Gland. Pituitary Adenomas. Atlas of Tumor Pathology. Washington, DC, Armed Forces Institute of Pathology, 1980.)
lead to pituitary damage and insufficiency. Common viral infections, including influenza, measles, mumps, and herpes, are rarely associated with pituitary damage and insufficiency. Although tuberculosis is rarely confined to the pituitary gland, most of the fewer than 20 reported patients exhibited suprasellar extension of the pituitary mass, compromised pituitary function, and visual defects. Although evidence for systemic tuberculosis is usually present, isolated sellar tuberculosis have been described.

**Hematologic Malignancies**

Primary central nervous system lymphomas are usually B-cell non-Hodgkin's types, and nine such patients with pituitary lymphoma have been described. The pituitary mass may be an isolated presentation of the underlying disease. The disorder is usually diagnosed by histologic examination of tissue obtained by excision biopsy. Six of nine patients had headache, and five had cranial nerve abnormalities with varying degrees of hypopituitarism. MRI reveals cavernous sinus invasion and isointense T1-weighted and T2-weighted images that are enhanced by gadolinium. Patients with solitary pituitary plasma-cytomas have been reported who did not have classical multiple myeloma. Acute lymphoblastic leukemia may be associated with periglandular pituitary infiltrates with minimal pituitary dysfunction.

**Pituitary Granulomas**

**Sarcoidosis**

Infiltrative sarcoidosis of the hypothalamic-pituitary region occurs in most patients with central nervous system sarcoid involvement. These patients may present with varying degrees of anterior pituitary failure with or without diabetes insipidus. Hypothalamic granulomatous involvement is commonly encountered in patients with central nervous system sarcoidosis and may be the sole manifestation of the disease. The hypothalamus, pituitary stalk, and posterior pituitary are diffusely invaded by noncaseating granulomas, consisting of giant cells, macrophages, and lymphocytes. Sarcoidosis may be progressive and eventually result in pituitary damage and even an empty sella. Onset of diabetes insipidus with no obvious features of a pituitary disorder should alert the physician to exclude hypothalamic sarcoid deposits, especially in the presence of a thickened stalk on MRI.

**Hand-Schüller-Christian Disease**

Hand-Schüller-Christian disease (histiocytosis X) may comprise sleep disorders, adipsia, and morbid obesity. Other features of granulomatous involvement, including axillary skin rash, history of recurrent pneumothorax, and classical bone lesions, should be sought, especially in young patients with new-onset diabetes insipidus. The disorder may be associated with granulomatous damage to the hypothalamus or posterior pituitary, or both, with characteristic diabetes insipidus. The pituitary lesions consist of dendritic Langerhans cells, and pituitary MRI may reveal stalk thickening or a diminished posterior pituitary bright spot. Adults with the disorder should be carefully evaluated for anterior pituitary hormone deficits, which should be appropriately replaced.

**Metastases to the Pituitary Region**

Pituitary metastases are found in up to 3.5% of patients with cancer. As the vascular supply to the posterior pituitary is derived directly from the systemic circulation through the internal carotid arteries, the posterior pituitary is the preferred site for blood-borne metastatic spread. Carcinomas that metastasize to the pituitary include breast, lung, and gastrointestinal tract cancers. Up to one quarter of patients with metastatic breast cancer have pituitary metastases. Symptomatic

---

**TABLE 8-7 -- Parasellar Masses**

<table>
<thead>
<tr>
<th>Genetic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transcription factor mutations (e.g., PROP-1)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cysts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rathke's</td>
</tr>
<tr>
<td>Arachnoid</td>
</tr>
<tr>
<td>Epidermoid</td>
</tr>
<tr>
<td>Dermoid</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tumors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hormone-secreting or nonfunctional pituitary adenoma</td>
</tr>
<tr>
<td>Granular cell tumor</td>
</tr>
<tr>
<td>Cranopharyngioma</td>
</tr>
<tr>
<td>Chordoma</td>
</tr>
<tr>
<td>Meningioma</td>
</tr>
<tr>
<td>Sarcomas</td>
</tr>
<tr>
<td>Gloma</td>
</tr>
<tr>
<td>Schwannoma</td>
</tr>
<tr>
<td>Germ cell tumor</td>
</tr>
<tr>
<td>Vascular tumor</td>
</tr>
<tr>
<td>Solid or hematologic metastases</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Malformation and Hamartomas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ectopic pituitary, neurohypophyseal, or salivary tissue</td>
</tr>
<tr>
<td>Hypothalamic hamartoma</td>
</tr>
<tr>
<td>Gangliocytoma</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Miscellaneous Lesions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aneurysms</td>
</tr>
<tr>
<td>Hypophysitis</td>
</tr>
<tr>
<td>Infections</td>
</tr>
<tr>
<td>Sarcoidosis</td>
</tr>
<tr>
<td>Giant cell granuloma</td>
</tr>
<tr>
<td>Histiocytosis X</td>
</tr>
</tbody>
</table>

---

Forces Institute of Pathology, 1998, p 51.)
pituitary metastases (usually diabetes insipidus) may be the presenting sign of occult malignancy and of malignancy of unknown origin. Rarely, isolated metastatic stalk deposits may also occur with pituitary failure.

If extensive bone erosion is present and disease onset is rapid, the diagnosis is more readily apparent. However, pituitary imaging may not clearly distinguish metastatic deposits from a pituitary adenoma, these lesions may masquerade as an adenoma, and the diagnosis is made only by histologic study of the resected specimen. When the diagnosis is clear-cut in the presence of a primary cancer, relatively low-dose pituitary radiation may be sufficient to shrink the metastasis and decrease morbidity.

Iron storage diseases, including hemochromatosis and hemosiderosis, result in predominantly gonadotroph cell damage. Idiopathic retroperitoneal fibrosis may also be associated with a suprasellar mass and hypothalamic panhypopituitarism. 

**Primary Hypophysitis**

Pituitary mass lesions composed of inflammatory cells may arise as primary disorders exclusively confined to the hypophysis. At least three distinct clinicopathologic forms have been described.

**Lymphocytic Hypophysitis**

This apparently autoimmune inflammatory disorder occurs predominantly during pregnancy or in postpartum women but has also been reported after menopause, and about 15% of reported cases occurred in men. Of the disorders that develop in association with pregnancy, about 50% occur during the first 6 postpartum months. The disorder is characterized by a lymphocytic and plasma cell pituitary infiltrate, which may be isolated or associated with other recognized endocrinopathies. Circulating antipituitary antibodies have occasionally been reported, and the presence of isolated pituitary hormone deficiency may imply an autoimmune process selectively targeted to pituitary cell types.

Although the natural history is often brief, the few comprehensive pathologic evaluations suggest that secondary adenohypophyseal atrophy, with a resultant empty sella, is a frequent outcome. Pathologic criteria for diagnosis include islands of anterior pituitary cells surrounded by diffuse lymphocytic (T-cell and B-cell) infiltrates. Over 125 cases have been reported, and the diagnosis has been definitively confirmed by histology in only about 10%. Some patients have been reported with lymphocytic, plasma cell, and macrophage infiltrates ultimately resulting in cell destruction, fibrosis, and an irreversible chronic hypophysitis.

**Clinical Features**

More than half the patients present with headache and visual field impairment, and pituitary deficiency accounts for the remaining cases. MRI reveals a

<table>
<thead>
<tr>
<th>TABLE 8-8 -- Features of Lymphocytic Hypophysitis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feature</td>
</tr>
<tr>
<td>Pituitary enlargement</td>
</tr>
<tr>
<td>Headache, visual disturbances</td>
</tr>
<tr>
<td>Hypopituitarism</td>
</tr>
<tr>
<td>Hyperprolactinemia</td>
</tr>
<tr>
<td>Associated autoimmune disease</td>
</tr>
<tr>
<td>Diabetes insipidus</td>
</tr>
</tbody>
</table>

pituitary mass, often indistinguishable from an adenoma. An associated partially empty sella and contrast enhancement of the pituitary mass may be helpful in distinguishing MRI features. The inflammatory process often resolves with time, and pituitary function may be restored or remain chronically compromised. Limited numbers of patients have been reported with histologically proven lymphocytic hypophysitis and documented spontaneous regression of pituitary mass on follow-up imaging. In two patients with histologically proven hypophysitis, spontaneous resolution of the pituitary mass was followed by subsequent successful pregnancies. Diabetes insipidus, encountered in 20% of patients, may be attributed to posterior pituitary or stalk infiltration. In one third of patients, other autoimmune conditions, including thyroiditis, hypoadrenalism, parathyroid failure, atrophic gastritis, systemic lupus erythematous, and Sjögren's syndrome, are also present. The differential diagnosis includes prolactinoma and other sellar masses, and a careful history and demonstrated loss of the posterior pituitary bright spot on MRI are useful for supporting the diagnosis.

**Laboratory Results**

The erythrocyte sedimentation rate is often elevated; antibodies to a 49-kd cytosolic protein were detected in 70% of patients with histologically confirmed lymphocytic hypophysitis and in 10% of control subjects. Although the specificity of this antibody and two additional antibodies to 68- and 43-kd human pituitary membrane antigens is high, all three were detected in only 5 of 13 patients with lymphocytic hypophysitis and 1 of 12 patients with infundibuloneurohypophysitis. PRL levels are usually elevated in both female and male patients, hyperprolactinemia is expected during pregnancy and during the early postpartum period, and the mass effect of the infiltrate may also contribute to stalk compression and secondary hyperprolactinemia. GH and ACTH responses to hypothalamic hormone challenges may be blunted. Rarely, isolated ACTH or TSH deficiencies have been reported.

**Treatment**

If the diagnosis is convincingly supported and compressive visual field disturbances are absent, surgical therapy should be withheld, pituitary hormone deficits appropriately replaced, and spontaneous resolution of the inflammatory mass observed expectantly. Treatment with adrenal steroids is advocated; it often resolves the sellar mass and improves endocrine dysfunction. Steroids are also indicated if adrenal reserve is compromised. Transsphenoidal surgery may be required to confirm the tissue diagnosis and may also relieve compression symptoms, but the degree of surgical resection should be constrained by the need to conserve viable pituitary tissue, particularly in view of frequent spontaneous resolution.

**Granulomatous Hypophysitis**

Granulomatous hypophysitis is not usually associated with pregnancy and has an equal female-male incidence. Rarely, the condition may coexist with lymphocytic
hypophysitis in the same gland. Pituitary histology shows histiocytes, multinucleated giant cells, and other features of chronic inflammation and granuloma. Patients present with headache and may have aseptic meningitis. MRI can reveal a thickened pituitary stalk or a characteristic tongue-shaped extension of the lesion under the hypothalamus. Granulomatous hypophysitis may reflect an underlying systemic disorder such as sarcoidosis or Takayasu's disease.

Xanthomatous Hypophysitis

This least common primary pituitary inflammatory process also occurs with the same frequency in both sexes and consists of lipid-laden macrophages, which resemble postinfectious cell debris. MRI often reveals a highly cystic lesion, leading to the suggestion that this entity reflects an inflammatory response to a damaged or ruptured pituitary cyst.

Hemorrhage and Infarction

Intrapituitary hemorrhage and infarction are usually caused by ischemic damage to the hypophyseal-portal system and may be catastrophic. These acute events cause significant damage to the pituitary gland, and small clinically silent microinfarcts are found in up to 5% of unselected autopsies. Pituitary cells are relatively resilient to vascular insult, and pituitary insufficiency is clinically apparent only when about 75% of the gland is chemically damaged. Ten percent residual functional pituitary cell mass appears sufficient to mask complete pituitary failure. Ischemic damage is limited to the anterior lobe and posterior pituitary function usually remains intact, reflecting the predominant neural control of oxytocin and arginine vasopressin secretion. Acute intrapituitary hemorrhage can cause significant life-threatening damage to the pituitary and its surrounding vital structures.

Postpartum Pituitary Infarction.

During pregnancy, the pituitary gland normally enlarges in response to estrogen stimulation. The hypervascular gland is thus particularly vulnerable to arterial pressure changes and prone to hemorrhage. Sheehan's syndrome, classically described after severe postpartum hemorrhage, is less commonly encountered with modern obstetric care. Development of hypovolemic shock in these women results in adenohypophysial vessel vasospasm and pituitary necrosis.

Pituitary Apoplexy

Pituitary apoplexy may result from spontaneous hemorrhage into a pituitary adenoma or occur after head trauma, skull base fracture, or in association with hypertension and diabetes mellitus, sickle cell anemia, or acute hypovolemic shock.

Clinical Features.

Pituitary apoplexy is an endocrine emergency. The condition may evolve over 1 to 2 days with severe headache, neck stiffness, and progressive cranial nerve damage, cardiovascular collapse, and change in consciousness. Signs include severe hypotension, bilateral visual disturbances, hypoglycemia, fever, central nervous system hemorrhage, and coma. Acute adrenal insufficiency may also be superimposed because of disordered intravascular clotting disorders, heparin administration, or acute effects of central nervous system hemorrhage. Pituitary imaging without contrast usually reveals signs of intrapituitary or intra-adenoma hemorrhage, stalk deviation, compression of normal pituitary tissue, and, in severe cases, signs of parasellar hemorrhage.

Management.

Most patients recover spontaneously but may experience long-term pituitary insufficiency. Patients who are fully alert and conscious with no visual symptoms may be observed. The decision to initiate therapy with high-dose glucocorticoids depends on the clinical status. Ophthalmoplegia, which is common, may resolve spontaneously over time. Failure of optic tract pressure to resolve or signs of progressive pituitary compression are indications for urgent transsphenoidal surgical decompression. Postoperative recovery of visual function correlates inversely with the time elapsed since the acute hemorrhage. Cranial nerve palsies, however, often improve whether or not surgery is undertaken. Pituitary function does not commonly recover after resolution of the acute hemorrhage, and patients require adrenal, thyroid, or gonadal steroid hormone replacement. The subsequent atrophy of infarcted pituitary tissue often results in the development of a complete or partially empty sella evident on MRI.
Evaluation of Pituitary Masses

Approach to the Patient Harboring a Pituitary Mass

Ninety-one percent of 1120 patients undergoing transphenoidal surgery for sellar masses were diagnosed with pituitary adenomas. Thus, the differential diagnosis of a pituitary mass should be aimed at excluding the diagnosis of a pituitary adenoma before considering other rare sellar lesions. The management of and prognosis for anterior pituitary adenomas differ markedly from those for other nonpituitary masses, and an important diagnostic challenge is to distinguish a pituitary adenoma from other parasellar masses.

Several physiologic states are associated with pituitary enlargement. Lactotroph hyperplasia is seen during pregnancy, and thyrotroph or gonadotroph hyperplasia occurs in the presence of long-standing primary thyroid or gonadal failure, respectively. Pituitary enlargement may also occur as a result of ectopic GHRIH or CRH secretion with resultant hyperplasia of somatotroph or corticotroph cells. Autopsy series show that up to 20% of subjects harbor an incidentally clinically silent pituitary adenoma. Incidental pituitary cysts, hemorrhages, and infarctions are also discovered at autopsy.

With the widespread use of sensitive imaging techniques for nonpituitary indications including head trauma, chronic sinusitis, or headaches, previously inapparent pituitary lesions are being identified with increasing frequency. Pituitary abnormalities compatible with the diagnosis of microadenoma are detectable in about 10% of the normal adult population undergoing MRI. Recognizing that approximately 90% of observed pituitary lesions represent pituitary adenomas, initial assessment should determine whether the mass is hormonally functional and whether local mass effects are apparent at the time of diagnosis or likely to develop in the future.

As clinical features associated with disordered hormone secretion have an insidious onset and may be unnoticed for years or decades, endocrine function should always be tested (Table 8-9). Clinical evaluation for changes compatible with hypersecretion or hyposecretion of GH, gonadotropins, PRL, or ACTH may reveal unique long-term sequelae requiring distinct therapies. In the absence of clinical features of a humoral hypersecretory syndrome, cost-effective laboratory screening should be performed. Serum PRL levels higher than 200 µg/L strongly suggest the presence of a prolactinoma, whereas lower PRL levels would indicate secondary stalk interruption by a pituitary mass. Elevated age-matched and gender-matched IGF-I levels indicate the presence of a GH-secreting adenoma, and a high 24-hour urinary free cortisol level is an effective screen for most patients with Cushing's disease. Nevertheless, the incidence of functional hormone-secreting tumors in asymptomatic subjects with incidentally discovered pituitary masses is low.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Test</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acronegaly</td>
<td>IGF-I</td>
<td>Interpret IGF-I relative to age- and gender-matched controls</td>
</tr>
<tr>
<td>Prolactinoma</td>
<td>OGTT with GH obtained at 0, 30 and 60 min</td>
<td>Normal subjects should suppress growth hormone to &lt;1 µg/L</td>
</tr>
<tr>
<td></td>
<td>Serum PRL level</td>
<td>MRI of the sella should be ordered if PRL levels elevated</td>
</tr>
<tr>
<td></td>
<td>Excludes medications</td>
<td></td>
</tr>
<tr>
<td>Cushing's disease</td>
<td>24-hr urinary free cortisol</td>
<td>Ensure urine collection is total and accurate</td>
</tr>
<tr>
<td></td>
<td>Dexamethasone (1 mg) at 11 am and fasting plasma cortisol measured at 8 am</td>
<td>Normal subjects suppress to &lt;5 µg/dL</td>
</tr>
<tr>
<td></td>
<td>ACTH assay</td>
<td>Distinguishes adrenal adenoma from ectopic ACTH or Cushing's disease.</td>
</tr>
</tbody>
</table>

ACTH, adrenocorticotropic hormone; IGF-I, insulin-like growth factor I; MRI, magnetic resonance imaging; OGTT, oral glucose tolerance test; PRL, prolactin.

The presence of, or the potential for, local compressive effects must also be considered. Because the risk of microadenoma enlargement toward a compressive macroadenoma is low, no direct intervention may be warranted. For parasellar masses of uncertain origin, histologic tissue examination may be the best approach to obtain an accurate diagnosis. Although MRI or CT imaging features may be helpful in diagnosing a nonpituitary sellar mass, the final diagnosis may remain elusive until pathologic confirmation is obtained.

Parasellar masses include neoplastic and nonneoplastic lesions and are manifest clinically by local compression of surrounding vital structures or metabolic or hormonal derangements (see Table 8-9). Rarely, sellar masses may be the presenting feature of a previously undiagnosed systemic disorder such as lymphoma or tuberculosis. Fever with or without associated sterile or septic meningitis may rarely be caused by fluid leakage into the subarachnoid space from Rathke's cleft, dermoid and epidermoid cysts, and craniopharyngioma and apoplexy. Patients with pituitary masses may present with hemorhage and infarction, especially during pregnancy when the normal pituitary is adenomatous and swollen; diabetes mellitus; and hypertension or hypotension, which may be found in elderly people with unsuspected pituitary tumors. Rarely, these adenomas occur with CSF leakage, which may predispose to meningitis. Pituitary masses may also undergo silent infarction leading to development of a partially or totally empty pituitary sella with normal pituitary reserve, implying that the surrounding rim of pituitary tissue is fully functional. Large sellar cysts may be mistaken for an empty sella.

Rarely, functional pituitary adenomas may arise within the remnant pituitary tissue, and these tumors may not be visible by sensitive MRI (i.e., <2 mm in diameter) despite their endocrine hyperactivity. Acute or chronic infection with abscess formation rarely occurs within the mass. Compromised pituitary hormone hyposecretion may be due to direct pressure effects of the expanding mass on hormone-secreting cells or to parasellar pressure effects that attenuate synthesis or secretion of hypophyseal hormones, with resultant pituitary failure. Hypophyseal masses (gangliocytomas) may overproduce a specific releasing hormone with resultant stimulation of secretion of a specific pituitary hormone.

Tumors of the pituitary gland are best diagnosed with MRI because it has better resolution than other radiologic modalities for identifying soft tissue changes. When a pituitary tumor or other parasellar mass is suspected, MRI specifically focused on the pituitary should be requested because brain MRI is often inadequate for optimal visualization of pituitary tumors and may miss the tumor completely. High-resolution T1-weighted sections in the coronal and sagittal planes both before and after administration of gadolinium pentetic acid for contrast distinguish most pituitary masses. Slice thickness should be less than 3 mm to obtain a pixel of 1 mm. Contiguous sections are therefore required to diagnose lesions of 1 to 3 mm. If necessary, especially for diagnosing high-signal hemorrhage, T2-weighted images provide additional diagnostic information.

MRI thus clearly delineates the pituitary gland, stalk, optic tracts, and surrounding soft tissues. The gland may be concave, convex, or flat. The posterior pituitary lobe exhibits a discrete bright spot of high signal intensity on T1-weighted images, which declines with age and is absent in diabetes insipidus and most posterior pituitary lesions. This T1 shortening may reflect the presence of arginine vasopressin localized within neurosecretory vesicles.

Tumors of the pituitary gland enlarge transiently during adolescence, during pregnancy, and after childbirth, and teenage girls exhibit increasing gland convexity during the menstrual cycle.
the gland should normally not exceed 10 to 12 mm and the stalk should not exceed 4 mm in diameter. A thickened stalk may indicate the presence of hypophysitis, granuloma, or atypical chondomas.

After gadolinium administration, microadenomas are usually hypodense compared with the normal gland, especially when multiple thin-section echo sequences are examined in the first few minutes after injection of the contrast agent. It has been suggested that this hypointensity may reflect compromised microadenoma vasculature. 

Microadenomas may also cause gland asymmetry or stalk deviation. In contrast, macroadenomas, which are significantly more vascular than microadenomas, have a higher affinity for gadolinium. They often enlarge the sella turcica by remodeling the bony fossa, suggesting a gradual long-term process. These tumors can grow upward toward the optic apparatus and cause draping of the nerves over the tumor, often accompanied by visual field abnormalities. Tumors can also extend into the sphenoid sinus and not infrequently invade connective tissue separating the pituitary from the cavernous sinus.

Radiologically, visible tumor tissue surrounding the carotid artery confirms cavernous sinus invasion. Infrequently, these patients experience palsies of third, fourth, or sixth cranial nerves. MRI readily distinguish pituitary adenomas from other masses, including hyperplasias, craniopharyngiomas, meningiomas, chondomas, cysts, and hypophysitis. Secondary distinguishing features such as visualization of distinct uninvolved pituitary tissue, mass consistency, calcification, hemorrhage, and suprasellar involvement usually allow an imaging diagnosis of these masses, but these can often be confirmed only by direct tissue histology. Preoperative localization of carotid artery aneurysms can also be confirmed by MRI or magnetic resonance angiography.

Pituitary CT allows visualization of bony structures, including the sellar floor and clinoid bones, and their invasion. CT also recognizes calcifications that characterize craniopharyngiomas, meningiomas, and rarely aneurysms. Calcifications are not evident on MRI. Occasionally, pituitary adenomas may calcify. Pituitary CT scanning is indicated for discovery of hemorrhagic lesions, metastatic deposits, and chondomas and evidence of calcification.

**Receptor Imaging**

Because prolactinomas express D2 receptors, they can be imaged with a radiolabeled D2 receptor antagonist by using iodine 123-labeled iodobenzamine single photon emission scanning. Failure to visualize nonfunctioning tumors by this technique has led some to advocate its use to distinguish the two tumor types. Radiolabeled indium pentetreotide has been used for in vivo tumor imaging. Most pituitary adenomas express somatostatin receptor subtypes to a varying degree, thus limiting the specificity of the procedure. Single photon emission CT (SPECT) has a sensitivity of about 1 cm and can detect normal pituitary tissue receptor expression. Because most adenomas and normal tissue are identified by this technique, its utility is limited for tumor detection, but it may be helpful for imaging ectopic ACTH-secreting tumors.

**Neuro-ophthalmologic Assessment of Pituitary Masses**

The optic tracts are particularly vulnerable to compression by expanding pituitary masses, and expanding pituitary tumors affect mostly the chiasm. Accurate neuro-ophthalmologic evaluation is helpful for tumor diagnosis, for determining pretreatment baseline visual status, and for post-treatment monitoring or detection of mass recurrence. The relationships of the optic chiasm and the intracranial components of the optic nerves with the pituitary gland and surrounding vessels are depicted in Figure 8-2 (Figure Not Available). A 10-mm posteriorly angled gap separates the optic chiasm and diaphragma sellae (see Fig. 8-3) (Figure Not Available). Therefore, extensive suprasellar mass extension is required before visual function is compromised. Decussation of neural fibers originating from the nasal half of each retina occurs at the chiasm, and those originating from the temporal retinal halves are situated ipsilaterally. Fibers from the superior and inferior retinal aspect are segregated in the corresponding chiasmal regions. Local vascular compromise or chiasmal stretching contributes to the pathogenesis of selective visual compromise. Reversibility of visual effects may correlate inversely with acuteness of the compressive insult.

**Visual Symptoms.**

An abnormal visual examination may unmask the presence of a pituitary mass in an asymptomatic patient. Before the availability of sophisticated assay and imaging techniques, patients with virtually all pituitary masses presented with visual loss. Currently, less than 10% of patients present with visual loss, and most of these harbor clinically nonfunctioning pituitary adenomas often detected by incidental imaging. Unilateral or bilateral temporal or central visual loss is usually asymmetric and may be quite insidious, remitting, or recurring. Rarely, sudden visual loss occurs in a previously asymptomatic patient. Other symptoms include diplopia, impaired depth perception, and rarely visual hallucinations.

**Clinical Signs.**

Impingement of the inferior crossing chiasmal fibers leads to bitemporal visual loss, especially in the superior field portions, accounting for about half of all pituitary-related visual defects. Pituitary-related defects preferentially marginate at the vertical field midline, whereas bitemporal defects of other causes tend to occur away from the midline. Despite prominent field defects, many of which can be directly correlated with defined tumor location by MRI, visual acuity in the remaining fields is invariably normal in over 95% of patients.
Management of Pituitary Masses

The goals of therapy for masses are to alleviate local compressive mass effects, to suppress hormone hypersecretion, and to relieve hormone hyposecretion while maintaining intact pituitary trophic function. The three modes of therapy available are surgical, radiotherapeutic, and medical approaches. In general, the benefits of each therapy should be weighed against the respective risks, and comprehensive awareness on the part of the physician and the patient is required to individualize treatment approaches.

Surgical Management of Pituitary Tumors and Sellar Masses

Pituitary surgery is indicated for excision of mass lesions causing central pressure effects, primary correction of hormonal hypersecretion (other than prolactinomas), or functional

---

**Figure 8-12** Threshold field test showing bitemporal hemianopsia in a patient with pituitary tumor compressing the optic chiasma (A) and superior bitemporal field cuts (B)

tumor resection in patients resistant to medical treatment. Unusual sellar lesions may require diagnostic tissue evaluation, and primary or secondary parasellar malignancies require excessive excision.

In 1904, Horsley reported the surgical resection of a pituitary tumor by a lateral middle fossa approach. The first successful transsphenoidal approach for pituitary tumor resection was reported by Schloffer in 1907 and subsequently refined by Cushing, who operated on 231 patients harboring pituitary tumors between 1910 and 1925 with a remarkably low mortality rate of 5.6%. Cushing used a sublabial incision to enable an endonasal approach for removing the septum and improved visualization using Kanavel's headlight. Hardy later improved the technique by using the operating microscope and intraoperative fluoroscopy, resulting in markedly reduced morbidity and mortality compared with those usually encountered with craniotomy, and this became the mainstay surgical technique for resecting these tumors.

Craniotomy is indicated only for the rare invasive parasellar mass extending into the frontal or middle cranial fossa or optic nerves or having extensive posterior clival invasion. Suprasellar extension contained by a small diaphragmatic aperture ("hourglass configuration") may also require a transcranial approach. The transsphenoidal approach precludes invasion of the cranial cavity and removes the need for the brain tissue manipulation required during a subfrontal surgical approach. A ventral sphenoid approach for resection of pituitary masses likewise does not violate the cranial fossa.

Thus, transsphenoidal surgery is associated with minimal morbidity and mortality, most patients are ambulatory within 6 to 9 hours, and the hospital stay is generally about 3 days. Furthermore, the transsphenoidal approach allows a clearly visible operative field with high magnification and internal illumination. Normal pituitary can be clearly distinguished from tumor tissue, facilitating microdissection and small tumor resection.

Enhanced MRI sensitivity and precision as well as intraoperative

---

**Figure 8-13** Transsphenoidal pituitary surgery. A, B, Route of the transsphenoidal approach (lateral view) and surgical corridor of the transsphenoidal approach and positioning of the retractor. The extent of removal of bone structures is indicated. C, Parasellar extensions of pituitary adenomas (coronal sections): intrasellar adenoma (a), displacement of the cavernous sinus (b), invasion of the cavernous sinus (c), diffuse invasion of the cavernous sinus by the adenoma (d), E, Extensions of a pituitary (e). Adenoma (sagittal sections): parasellar extension; invasion of the sphenoid sinus and of the clivus (f). (Adapted from Honegger J, Buchfelder M, Fahrbusch R. Surgery for pituitary tumors. In Sheaves R, Jenkins PJ, Wass JAH [eds]. Clinical Endocrine Oncology. Cambridge, Mass, Blackwell Science, 1977, p 179.)

MRI allows clear delineation of the location, size, and invasiveness of the tumor, all critical determinants of surgical success.

Goals of Surgery

The goal of pituitary surgery is total resection limited to the lesion without compromising postoperative endogenous pituitary function. Careful selective mass resection may be difficult for poorly encapsulated lesions, those embedded deeply within the gland body, and those extending into cavernous sinuses or parasellar lesions. Poor operative field visibility also limits resection precision. Normal tissue excision or even gland manipulation should be avoided unless critical for effective dissection. Occasionally, hemihypophysectomy or even nonelective total gland resection may be indicated for multifocal tumors, when the surrounding normal gland is necrotic, or when no mass lesion is discernible despite an accurate clinical and biochemical diagnosis (especially for ACTH cell tumors).

Successful surgery should decompress central visual defects and compromised trophic hormone secretion. For children and young adults, the consideration of adequate normal tissue for subsequent growth patterns and reproductive function is an important determinant in intraoperative decision making. Nevertheless, especially for functional tumors, small residual remnants attached to the dura are difficult to reach but remain hypersecretory with persistent clinical progression. Thus, the skilled neurosurgeon carefully balances maximally effective tumor removal with the requirement to preserve nontumorous pituitary trophic function.

 Advances have enabled improved surgical results, although the long-term outcomes with these new techniques have not yet been rigorously compared with those after standard operations performed by skilled surgeons. Image-guided approaches
enable intraoperative surgical neuronavigation by three-dimensional imaging. Intraoperative ultrasonography and MRI technologies allow real-time assessment of the dimensions and extent of the pituitary mass and the progress of surgery. Intraoperative MRI is performed while the surgical field is still open, allowing the surgeon to assess directly the need for further dissection, and also provides an excellent baseline for postoperative follow-up. In contrast, postoperative image stabilization may not be evident for months after surgery.

Endonasal transsphenoidal endoscopy avoids use of a retractor or speculum, does not require nasal packing, and sometimes leads to a shorter operating time, resulting in decreased postoperative morbidity and a shorter hospital stay (Fig. 8-15). The advantages of the technique include a clear panoramic view of bone landmarks and access to suprasellar and parasellar tumor extensions into the cavernous sinuses. Disadvantages of this relatively new approach include the need for management of perioperative intrasellar bleeding and CSF leaks as well as the added requirement for a preoperative CT scan.

**Indications for Transsphenoidal Surgery**

A pituitary mass that may or may not be compressing local vital structures should be evaluated for surgical resection (Table 8-10). Although surgical resection offers a rapid resolution of hormone hypersecretion and many of the resultant clinical features of functioning adenomas, indications for the procedure differ depending on the tumor type (see later). In general, patients who are intolerant of or resistant to medical therapy require surgery. Surgery is indicated primarily for well-circumscribed GH-secreting adenomas and all ACTH-secreting, TSH-secreting, and nonfunctioning macroadenomas. Surgery may also be indicated when tissue histology is required to determine the nature of an enigmatic sellar mass. Progressive compressive features, including visual field loss, compromised pituitary function, or other central nervous system functional change, are indications for surgical debulking and sellar decompression. Hemorrhage into the enced bony sella turcica, usually occurring within a known or previously unknown adenoma, may require immediate surgical decompression. Urgent surgical decompression is required for acute pituitary hemorrhage that may result in apoplexy related to partial or complete infarction, especially in patients with signs of progressive compressive signs.

When pituitary function after surgery was assessed in 234 patients, 52 patients had new trophic hormone dysfunction and 45 of 93 patients with preoperative evidence for hypopituitarism had recovered between one and three previously suppressed axes. Significant factors determining restoration of postoperative pituitary function were no visible tumor remnants as assessed by MRI (P < 0.001) and no tumor invasion as determined by the neurosurgeon as well as by pathologic examination of surrounding tissue (P < 0.049). Therefore, because about half of all patients with preoperative pituitary failure recover function, depending on the clinical circumstance, patients should be considered for retesting before initiating postoperative substitution therapy except for adrenal steroid replacement, which requires greater caution. Indications for second surgery in the same patient include tumor recurrence, persistent hormonal hypersecretion by tumor remnants, or repair of a CSF leak.

After surgery, patients should be kept in bed rest at an angle of 30 to 45 degrees and urine and serum osmolality and serum electrolytes should be measured every 6 hours. Indications for postoperative vasopressin replacement include urine output greater than 300 mL/hour for 3 consecutive hours with serum osmolality above 285 mOsm/L, elevated serum sodium concentrations, and inappropriately low urine osmolality. Postoperative polyuria alone is not an indication for vasopressin replacement unless it is a reflection of compromised posterior pituitary function.

**Side Effects**

The success of surgery is largely determined by the skill and experience of the neurosurgeon. Tumor size, degree of invasiveness, preoperative hormone levels, and previous pituitary surgery are all determinants of surgical outcome. CSF leakage, transient diabetes insipidus, and inappropriate arginine vasopressin secretion are the most commonly encountered transient side effects, occurring in up to 20% of patients (see Table 8-10). Local damage may also result in arachnoiditis, vascular bleeding, hematoma formation, and epistaxis. Rarely, pulmonary embolism, narcolepsy, and local abscess have been reported. Intracranial hypopituitarism, diabetes insipidus, and syndrome of inappropriate antidiuretic hormone (SIADH) have been reported in up to 10% of patients. Rarely, the central nervous system may be permanently damaged with hemiparesis, cranial nerve palsies, or encephalopathy.

A triphasic postoperative diabetes insipidus has been described in which the transient disorder is followed by an interphase on days 6 to 11 with no polydipsia or polyuria. During this later phase, hyponatremia with features of inappropriate antidiuretic hormone secretion has also been reported. Cognitive dysfunction, including deficits in anterograde memory and executive function, has been reported in several retrospective studies.

**TABLE 8-10 -- Transsphenoidal Pituitary Surgery**

<table>
<thead>
<tr>
<th>Primary Indications</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General</strong></td>
<td></td>
</tr>
<tr>
<td>Visual tract or central nervous compression arising from within sella</td>
<td></td>
</tr>
<tr>
<td>Relief of compressive hypopituitarism by presenting, residual, or recurrent tumor tissue</td>
<td></td>
</tr>
<tr>
<td>Tumor recurrence after surgery or radiation</td>
<td></td>
</tr>
<tr>
<td>Pituitary hemorrhage</td>
<td></td>
</tr>
<tr>
<td>Cerebrospinal fluid leak</td>
<td></td>
</tr>
<tr>
<td>Resistance to medical therapy</td>
<td></td>
</tr>
<tr>
<td>Intolerance of medical therapy</td>
<td></td>
</tr>
<tr>
<td>Personal choice</td>
<td></td>
</tr>
<tr>
<td>Desire for immediate pregnancy with macroadenoma</td>
<td></td>
</tr>
<tr>
<td>Requirement for diagnostic tissue histology</td>
<td></td>
</tr>
</tbody>
</table>

**References**

2. Figure 8-14 Transsphenoidal resection of pituitary adenoma. (Modified from BMI Quarterly 6:5, 1990.)
3. Figure 8-15 Endoscope-assisted microsurgery provides a panoramic view of the sphenoid sinus. Using a 30-degree endoscope, a view "around the corner" is possible. Paranasal structures can be visualized and residual tumor detected and resected. (From Fahlbusch R, Buchfelder M, Kreutzer J, Nomikos P. Surgical management of acromegaly. In Wass JAH [ed]. Handbook of Acromegaly. Bristol, UK, BioScientifica, 2001, p. 48.)
Specific
Acromegaly
Cushing's disease
Clinically nonfunctioning macroadenoma
Prolactinoma (rarely indicated)
Nelson's syndrome
TSH-secreting adenoma

Side Effects
Transient
Diabetes insipidus
Cerebrospinal fluid leak and rhinorrhea
Inappropriate ADH secretion
Arachnoiditis
Meningitis
Postoperative psychosis
Local hematoma
Arterial wall damage
Epistaxis
Local abscess
Pulmonary embolism
Narcolepsy

Permanent (up to 10%)
Diabetes insipidus
Total or partial hypopituitarism
Visual loss
Inappropriate ADH secretion
Vascular occlusion
CNS damage
Oculomotor palsy
Hemiparesis
Encephalopathy
Nasal septum perforation

Surgery-Related Mortality (up to 1%)
Brain, hypothalamic injury
Vascular damage
Postoperative meningitis
Cerebrospinal leak
Pneumocephalus
Acute cardipulmonary disease
Anesthetic
Seizure

ADH, antidiuretic hormone; CNS, central nervous system; TSH, thyroid-stimulating hormone.

studies after transsphenoidal surgery. \[164\] \[165\] Mortality has been reported in less than 1% of patients undergoing pituitary surgery and may be related to direct hypothalamic or cerebrovascular damage, meningitis, pneumocephalus formation, or anesthetic complications. Surgical failure may result from a nonpituitary-related event, including anesthesia-related complication or bleeding disorder. Incomplete tumor removal may also be due to inaccurate preoperative MRI localization or identification. Rarely, a previously undiagnosed functioning pituitary tumor or ectopic source of ACTH may be unmasked after initially unsuccessful pituitary surgery.

Pituitary Radiation

Principles
High-energy ionizing radiation can be delivered to deep tissues by megavoltage techniques. The challenge of this approach is to provide maximal localized necrotizing radiation to the pituitary lesion while minimally exposing surrounding normal structures to radiation damage. Several advances have improved both efficacy and safety, including highly precise tumor localization, a high-voltage (6 to 15 MEv) linear accelerator, and accurate simulation models with isocentric rotational arcing that allow repeated head positioning at exactly the same points in the patient's subsequent visits. Up to a maximum of 4500 rads is administered as 180-rad daily fractions for about 5 to 6 weeks. High-precision techniques such as stereotactic conformal radiotherapy \[166\] and the gamma knife \[167\] allow delivery of high energy to the pituitary lesion while minimizing the mass of normal brain exposed to radiation (Fig. 8-16) (Figure Not Available)

Indications
The use of radiation for treating pituitary tumors is highly individualized and depends on the expertise of the treating center, conviction of the treating physician in weighing the potential benefits and risks of the procedure, and the patient's preference based on informed choice (Table 8-11). In general, radiation techniques are indicated for persistent hormone hypersecretion or residual mass effects after surgery or when surgery for a compressive mass is contraindicated. As GH-secreting and PRL-secreting tumors are generally amenable to medical therapy, indications for radiation are rare. Most indications for radiation are adjuvant to either surgical or medical treatment. Radiation may be indicated after resection of a potentially recurring or inadequately resected pituitary mass, such as a nonfunctioning pituitary adenoma, craniopharyngioma, or chordoma. In acromegaly, use of radiation as primary treatment is generally not recommended, \[168\] but for resistant, aggressively growing prolactinomas the procedure may prevent further local invasion. Recurrent pituitary-dependent Cushing's disease appears to be particularly amenable to radiation, especially in younger patients.
Hypopituitarism.

Pituitary failure occurs commonly in patients who have received pituitary radiation. Within 10 years after radiation, up to 80% of patients may have gonadotroph, somatotroph, thyrotroph, or corticotroph deficits. The mechanism for hypopituitarism appears to involve damage to hypothalamic releasing hormone cells as well as direct pituitary damage. These patients require lifelong endocrine follow-up for pituitary reserve testing and hormone replacement when appropriate.

Second Brain Tumors.

Thirty-two cases of glioma occurring after conventional pituitary radiation for adenomas and craniopharyngioma have been reported with a mean latency period of 11.5 years from initial diagnosis. In a meta-analysis of results of irradiation for pituitary tumors, the standardized incidence ratio for second brain tumors was approximately 6 (confidence interval 3.16 to 10.69). This analysis was based on 12 diagnosed second tumors with a latency of 6 to 24 years in three separate cohorts. Because patients harboring pituitary tumors are more likely to undergo routine brain imaging during follow-up, it is not clear whether observed meningiomas are coincidental findings. As this complication, which occurs in less than 5% of patients, also appears dose-related, fractionated doses not exceeding 4500 rads should be given. Use of conformal radiation techniques to irradiate a smaller tissue volume, including radiosurgery, fractionated stereotactic radiotherapy, and proton beam radiation, may minimize this adverse effect. Nevertheless, prospectively controlled surveillance studies are required to evaluate this critical question rigorously.

Visual Damage.

Approximately 2% of patients experience impaired vision related to optic nerve damage. The risk of visual damage is minimized by fractionating dosages to less than 200 rads per treatment session. Consequent blindness, however, has been reported in two patients who received 4500 rads in 180-rad fractions.

Brain Necrosis.

Dose-related radiation-induced brain necrosis was documented by MRI in 14 of 45 patients, with temporal lobe atrophy and cystic and diffuse cerebral atrophy. Cognitive dysfunction, especially memory loss, has also been reported.

Radiosurgery

The proton beam, the gamma knife using focused cobalt 60 emissions, and the linear accelerator deliver high-dose radiation while sparing surrounding tissue. Delivery of high energy by gamma knife directly targeted at the pituitary tumor minimizes the radiation exposure of surrounding tissues. Early reports indicated more rapid reduction of hormone hypersecretion, although long-term efficacy and safety outcomes are not yet apparent. It appears that this procedure is best suited for intrasellar and cavernous lesions distant from the optic nerves.

Medical

Pituitary tumors often express receptors mediating hypothalamic control of hormone secretion, and appropriate ligands for the dopamine D2 receptor and the somatotropin release-inhibiting factor (SRIF) receptor subtype 2 are employed to suppress PRL or GH hypersecretion, to block tumor growth, and often to shrink tumor size. A novel approach has employed a peripheral receptor antagonist to block GH action without targeting the pituitary tumor source. Medical ablation of target glands, including thyroid and adrenal, may also be useful in mitigating the deleterious impact of pituitary tumor hypersecretion. Each of these medical approaches is fully considered in the following.
PHYSIOLOGY AND DISORDERS OF PITUITARY HORMONE AXES

Prolactin

Lactotroph Cells

Lactotroph cells constitute about 15% to 25% of functioning anterior pituitary cells (Fig. 8-17) (Figure Not Available). Although their absolute number does not change with age, lactotroph hyperplasia does occur during pregnancy and lactation \(^{180}\) and resolves within several months of delivery (Fig. 8-18) (Figure Not Available). Most PRL-expressing cells appear to arise from GH-producing cells. Ablation of somatotrophs by expression of GH-diphtheria toxin and GH-thymidine kinase fusion genes inserted into the germ line of transgenic mice eliminated most lactotrophs, suggesting that the majority of PRL-producing cells arose from postmitotic somatotrophs. \(^{181}\)

Two cell forms expressing the PRL gene are large polyhedral cells found throughout the gland and smaller angulated or elongated cells clustered mainly in the lateral wings and median wedge. Large PRL secretory granules (250 to 800 nm) are present in the evenly distributed cells, and the laterally localized cells are sparsely populated by smaller (200 to 350 nm) granules (Fig. 8-19) (Figure Not Available). Occasional mammosomatotroph cells may also cosecrete both PRL and GH, often stored within the same granule (Fig. 8-20) (Figure Not Available). In animal models, lactotroph cell function is heterogeneous. Thus, dopamine or TRH responsiveness and shifting proportions of PRL-secreting and GH-secreting cells may depend on cell localization within the pituitary as well as the surrounding hormonal milieu, especially that of estrogen. \(^{182}\)

TABLE 8-11 — Pituitary Radiation

<table>
<thead>
<tr>
<th>Indications</th>
<th>Incidence</th>
<th>95% Confidence Intervals</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pituitary adenoma</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acromegaly</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cushing's disease</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonfunctioning adenoma</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prolactinoma</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cranioopharyngioma</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nelson's syndrome</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonadenomatous invasive sellar mass</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tumor recurrence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hormone hypersecretion recurrence</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Side Effects</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypopituitarism</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deficient GH, gonadotropin, TSH, and ACTH reserve</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eye</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optic neuritis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brain</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brain necrosis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temporal lobe deficits</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive dysfunction</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Relative Risk of Second Brain Tumor Post-radiation

<table>
<thead>
<tr>
<th>Second Tumor</th>
<th>Incidence</th>
<th>95% Confidence Intervals</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Astrocytoma (2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meningioma (1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meningeal sarcoma (1)</td>
<td>5 0.53 3.4 3.0521.98</td>
<td>(Brada, 1992)</td>
<td></td>
</tr>
<tr>
<td>Gliomas</td>
<td>4 0.25 16 4.441</td>
<td>(Tsang, 1993)</td>
<td></td>
</tr>
<tr>
<td>Astrocytoma (2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meningioma (1)</td>
<td>3 1.13 2.7</td>
<td>(Erfurth, 2001)</td>
<td></td>
</tr>
<tr>
<td>Meta-analysis</td>
<td>12 1.96 6.1 3.1610.69</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ACTH, adrenocorticotropic hormone; GH, growth hormone; SIR, standardized incidence ratio for person-years at risk; TSH, thyroid-stimulating hormone.


*Excludes patients with acromegaly.

Prolactin History

Shortly after its discovery and partial characterization, \(^{183}\) \(^{184}\) \(^{185}\) PRL was prominently featured in The New York Times on December 3, 1937, indicating that it held the "key to peace in the world." The article, based on a lecture delivered by Prof. C. R. Stockard, proposed that "higher forms of life" were "governed by a 'glandocracy,' with the glands of internal secretion as the supreme rulers [in this instance PRL], exerting absolute control not only over the functioning of the individual from conception to death but also over the relationship of men and other vertebrate animals to each other."
The identification of PRL in humans was elusive until 1970 because human GH is highly lactogenic and active in bioassays used to isolate and measure PRL. Furthermore, GH is present in human pituitary glands in much higher concentrations (5 to 10 mg) than PRL (100 μg). To distinguish human PRL from GH, lactogenic activity was neutralized with GH antisera; sera from postpartum women and patients with galactorrhea had high lactogenic activity in the presence of GH antibodies. Human PRL, bioassayed by stimulating pregnant mouse mammary milk production, was elevated in patients with nonpuerperal galactorrhea resulting from pituitary tumors, exposure to phenothiazines, and withdrawal from oral contraceptives. The purification and isolation of PRL by Frisen and colleagues and the development of a specific radioreimmunoassay underscored the new place of PRL in understanding human disease.

Prolactin Structure

The human PRL gene, located on chromosome 6, apparently arose from a single common ancestral gene giving rise to the relatively homologous PRL, GH, and placental lactogen-related proteins (Fig. 8-21) (Figure Not Available). Several factors influence PRL gene expression, including estrogen, dopamine, TRH, and thyroid hormones. PRL is a 199-amino-acid polypeptide containing three intramolecular disulfide bonds. It circulates in blood in various sizes: monomeric PRL ("little" PRL: 23 kd), dimeric PRL ("big" PRL: 48 to 56 kd), and polymeric forms (also known as "big, big" PRL: > 100 kd). The monomeric form is the most bioactive PRL. In response to TRH, the proportion of the more active monomeric form increases.

Regulation of Prolactin Secretion

PRL secretion is under the inhibitory control of dopamine, which is largely produced by the tuberoinfundibular (TIDA) cells, and the hypothalamic tuberohypophyseal dopaminergic system. DA reaches the lactotrophs through the hypothalamic pituitary portal system and inhibits PRL secretion by binding to the D2 receptors on pituitary lactotrophs. In turn, participates in negative feedback to control its release by increasing tyrosine hydroxylase activity in the TIDA neurons. In PRL-deficient animals, DA was decreased in the median eminence. Mice lacking the D2 receptor experienced hyperprolactinemia and lactotroph proliferation.

Factors other than DA inhibit PRL secretion, including endothelin-1 and transforming growth factor 1, which act as paracrine PRL inhibitors. and calcitonin, which may be derived from the hypothalamus. Several substances act as PRL-releasing factors. Basic FGF and epidermal growth factor induce PRL synthesis and secretion. Vasoactive intestinal polypeptide (VIP) stimulates PRL synthesis through cAMP. A hypothalamic PRL-releasing peptide produced in the hypothalamus acts through a specific receptor in normal pituitary glands and in a subset of PRL-secreting tumors. Oxytocin and pituitary adenylate cyclase activating protein also release PRL. TRH stimulates PRL but probably does not play an important role in PRL secretion. Estrogen stimulates PRL gene transcription and secretion; explaining why women have higher PRL levels and why cycling women have a higher PRL pulse frequency than postmenopausal women and men. Galanin is synthesized in both the pituitary and hypothalamus and may act as a PRL releasing factor. The physiologic role of amnouretic acid, neurotensin, substance P, bombesin, and cholecystokinin in regulating human PRL secretion is unresolved.

Serotonin may be additive with VIP in releasing PRL, and infusion of 5-hydroxytryptophan, a serotonin precursor, elicits PRL release. Nocturnal PRL secretion is attenuated by cyproheptadine. Thus, patients may mediate nocturnal PRL secretion and also participate with VIP in the sucking reflex. Opiates acutely induce PRL release, although naloxone does not consistently suppress PRL levels. GHRH, when administered in high doses, moderately induces PRL secretion, and patients harboring ectopic GHRH-producing tumors have mild to moderate hyperprolactinemia. GHRH also stimulates PRL in women, especially during the periovulatory period. Although posterior pituitary hormones have been shown to regulate rat PRL secretion, the role of vasopressin or oxytocin or other neurohypophyseal molecules in regulating human PRL remains unresolved. Histamine may act on the hypothalamus to regulate PRL, and H₂ blockers induce PRL secretion. A short-loop feedback of PRL has been proposed, and transgenic mice with deleted PRL were found to have a decreased hypothalamic dopaminergic complex.

Prolactin Receptor

The PRL receptor gene is a member of the cytokine receptor superfamily. Localizes to chromosome 5p13, and has 10 exons. The receptor gene has two 5' promoters that direct transcription of a 598-amino-acid peptide comprising an extracellular domain, a hydrophobic transmembrane domain, and an intracytoplasmic region homologous to the GH receptor. Similarly, PRL receptor dimerization occurs with ligand binding and subsequent phosphorylation of intracellular Janus kinases (JAKs) and signal transducer and activator of transcription (STATs) molecules. Two binding sites and activating helices 1 and 4 and helices 1 and 3 on the PRL receptor gene is a member of the cytokine receptor superfamily and has 10 exons. The receptor gene has two 5' promoters that direct transcription of a 598-amino-acid peptide comprising an extracellular domain, a hydrophobic transmembrane domain, and an intracytoplasmic region homologous to the GH receptor. Similarly, PRL receptor dimerization occurs with ligand binding and subsequent phosphorylation of intracellular Janus kinases (JAKs) and signal transducer and activator of transcription (STATs) molecules. Two binding sites and activating helices 1 and 4 and helices 1 and 3 on the PRL receptor gene are critical for formation of the trimeric ligand-receptor complex and subsequent signaling (see Fig. 8-21) (Figure Not Available). The PRL receptor promotes the development of Jak2 kinase and STATs 1 to 5. STATs phosphorylation mediates transcriptional activation of the -caspase gene. PRL receptors are expressed in breast, pituitary, liver, adrenal cortex, kidneys, prostate, ovary, testes, intestine, epidermis, pancreatic islets, lung, myocardium, brain, and lymphocytes. Estrogen also induces liver PRL receptor expression. Regulation of milk production occurs through a cascade of intracellular events. Homozygous mice in which the PRL receptor was inactivated were infertile; heterozygous animals were fertile but unable to nurse their first litters, presumably because of inadequate PRL receptor expression after the first but not subsequent pregnancies.

Functions of Prolactin

PRL is essential for human survival because of its role in milk production during pregnancy and lactation. Additional biologic functions ascribed to PRL include reproductive and metabolic effects, mammary development, pigment carp sac activity, fresh water survival, melanin synthesis, water-seeking behavior of news, motility, and parental behavior. Although PRL and its receptor are clearly crucial in lower animals, the impact of PRL on maternal behavior in humans has not been fully delineated.

Mammary Gland Development and Lactation

Figure 8-20 (Figure Not Available) Normal mammosomatotrophs. Occasional cells resembling densely granulated somatotrophs exhibit atypical features consistent with prolactin secretion; the secreting granules are highly pleomorphic and there is misplaced exocytosis, that is, extrusion of secretory material along the lateral cell border (arrow). (From Asa SL. In Tumors of the Pituitary Gland. Atlas of Tumor Pathology. Washington, DC, Armed Forces Institute of Pathology, 1997, p 17.)

Figure 8-19 (Figure Not Available) Electron micrograph of a normal lactotroph shows a well-developed rough endoplasmic reticulum that forms concentric whorls. A prominent Golgi complex is seen in a juxtaparticular location and harbors forming pleomorphic secretory granules. The cytoplasm is otherwise sparsely granulated. (From Asa SL. In Tumors of the Pituitary Gland. Atlas of Tumor Pathology. Washington, DC, Armed Forces Institute of Pathology, 1997, p 16.)

Figure 8-18 (Figure Not Available) Prolactin cell hyperplasia. In the third trimester of pregnancy, prolactin cell hyperplasia occurs; cells containing immunoreactive prolactin make up almost 50% of the cell population of the gland. (From Asa SL. In Tumors of the Pituitary Gland. Atlas of Tumor Pathology. Washington, DC, Armed Forces Institute of Pathology, 1997, p 15.)

Figure 8-17 (Figure Not Available) Lactotroph cell. Normal prolactin-secreting cells express strong positivity for prolactin within the cytoplasm of polygonal cells, which have elongated cell processes. Some processes surround adjacent immunoreactive cells that correspond to gonadotrophs. (From Asa SL. In Tumors of the Pituitary Gland. Atlas of Tumor Pathology. Washington, DC, Armed Forces Institute of Pathology, 1997, p 15.)

Figure 8-21 (Figure Not Available) Molecular structure of prolactin and its interaction with the dimerized receptor. (From Bole-Feysot C, Goffin V, Edery M, et al. Prolactin (PRL) and its receptors: actions, signal transduction pathways and phenotypes observed in PRL receptor knockout mice. Endocr Rev 1998; 19:225268.)
PRL is not essential for pubertal mammary development, which appears to require GH, whose action is mediated by IGF-I. Studies of mammary development have, for the most part, been carried out in rodents. At birth, the mammary gland consists of a fat pad with small areas of ductal anlagen, which differentiate into pubertal mammary glandular elements under the influence of estrogen, GH, and IGF-I. At puberty, a surge of estrogen begins the process. Terminal end buds form and lead the process of mammary development by branching and extending into the substance of the mammary fat pad, leaving a network of ducts that virtually fill the mouse mammary fat pad. Interestingly, estrogen does not affect this process in the absence of GH and IGF-I.

These hormones are also responsible for most of ductal morphogenesis. Thus, GH acts on the mammary stromal compartment to produce IGF-I, which, in turn, elevates PRL levels after delivery. Suckling also increases milk production after parturition and is essential for continued lactation because of its distal effect on pituitary hormone production and because it empties the mammary gland of milk. Milk accumulation further inhibits milk synthesis, explaining why a certain level of nursing activity is necessary for successful breast-feeding. In the absence of suckling, PRL concentrations, which rise throughout gestation, return to normal by 7 postpartum days. Suckling increases serum PRL levels approximately 8.5-fold in actively nursing mothers, and amniotic fluid PRL concentrations are 100 times those of maternal or fetal blood.

Alveolar formation and milk production also require progesterone, and lobular-alveolar formation does not occur in mice lacking the progesterone receptor.

Mechanisms of milk production are similar in all mammals, but milk composition differs. Active lactation is due in part to a fall off in estrogen and progesterone and elevation of PRL levels after delivery. Suckling also increases milk production after parturition and is essential for continued lactation because of its distal effect on pituitary hormone production and because it empties the mammary gland of milk. Milk accumulation further inhibits milk synthesis, explaining why a certain level of nursing activity is necessary for successful breast-feeding. In the absence of suckling, PRL concentrations, which rise throughout gestation, return to normal by 7 postpartum days. Suckling increases serum PRL levels approximately 8.5-fold in actively nursing mothers, and the milk let-down phenomenon is not associated with increased PRL. As nursing continues, PRL concentrations fall, but each suckling episode causes a subsequent episodic rise in serum PRL. Mean serum concentrations were 162 µg/L at 2 to 4 postpartum weeks, 130 µg/L at 5 to 14 weeks, and 77 µg/L at 15 to 24 weeks.

It is unclear why active milk production continues despite progressively lower PRL levels after parturition. Although PRL is essential for milk production, the milk yield does not closely correlate with serum PRL levels. In addition to its effects on PRL, suckling stimulates posterior pituitary oxytocin release. Unlike those of PRL, oxytocin responses to suckling do not decline as nursing continues for up to 6 months. Mothers who breast-fed exclusively had mean stimulated oxytocin levels significantly higher during late lactation than during early lactation. Oxytocin induces myoepithelial cell contraction, thereby causing milk ejection. Oxytocin also has important effects on alveolar proliferation. Mice deficient in oxytocin are unable to nurse their young, and oxytocin replacement permits dams to nurse.

Lactational Amenorrhea

Lactational amenorrhea is a form of contraception that depends on the frequency and duration of breast-feeding. The Kung hunter-gatherer women were suckled approximately four times an hour and at will during the night and bore a mean of 4.7 children during their reproductive years. In contrast, the Hutterites of North America bore a mean of 10.6 children during their lifetimes, presumably because they nursed according to a rigid schedule, used supplemental feedings, and weaned at 1 year. In Edinburgh, resumption of menses, albent anovulatory, occurred in 28 weeks and the first ovulation occurred at a mean of 34 postpartum weeks because of persistently abnormal LH pulsatile secretion.

Immune Function

Several lines of evidence indicate that PRL is a lymphocyte growth factor and stimulates immune responsiveness. PRL levels change in concert with immune disease, as seen in patients with lupus erythematosus. In immunosuppressed mice, PRL stimulated immune cell functions. Although PRL has been suggested as an immunomodulatory hormone, there is evidence that PRL may not be important for immune function because innate immunity was not altered in mice that lacked either the PRL receptor (PRLR) or the PRL gene (PRL).
levels. In male subjects with hyperprolactinemia, LH and FSH pulsatility is attenuated, testosterone levels are suppressed, and sperm counts and motility are low.

**Prolactin Assays**

The PRL radioimmunoassay (RIA) is highly specific and clearly distinguishes PRL from GH. PRL measurements are standardized using reference preparations provided by the National Institute for Biological Standards and Control in London and the National Hormone and Pituitary Program. Improvements in assay efficiency and turnaround time, reproducibility, and sensitivity have been achieved by immunoradiometric assay (IRMA) and chemiluminescent PRL assays. Because these samples are usually assayed at a single dilution, extremely high PRL concentrations may saturate their ability to detect very high PRL levels, resulting in a falsely low value being reported. This "hook" effect may result in PRL-secreting macroadenomas diagnosed as clinically nonfunctioning adenomas, with "normal" PRL levels reported in about 5% of patients. In patients harboring macroadenomas with clear-cut clinical features of hyperprolactinemia, serum samples should be subjected to at least a 1:100 dilution before assay.

**Prolactin Secretion**

The calculated production rate of PRL ranges from 200 to 536 µg/day/m2, and the metabolic clearance rate ranges from 40 to 71 mL/min/m2. PRL is cleared rapidly with a calculated disappearance half-life ranging from 26 to 47 minutes. PRL secretion occurs episodically in 4 to 14 secretory pulses, each lasting 67 to 76 minutes, over 24 hours. PRL is secreted episodically during the day, with the highest levels achieved during sleep and the lowest occurring between 10 AM and noon. The nocturnal elevation is sleep entrained and a temporal relationship exists between rapid eye movement (REM) and non-REM sleep cycles. PRL and PRL may cause periods of REM. VIP stimulates both REM sleep and PRL, and when VIP was given to rats together with a PRL antiserum, REM sleep was inhibited. PRL levels fall with age in both men and women. In older men, less PRL is produced with each secretory burst than in younger men. Likewise, postmenopausal women have lower mean serum PRL levels and a lower PRL pulse frequency than premenopausal women or men, suggesting a stimulatory effect of estrogen on both of these parameters.

**Hyperprolactinemia**

In the absence of a prolactinoma, hyperprolactinemia may be caused by other pituitary or sellar tumors that inhibit dopamine because of pressure on the pituitary stalk or interruption of the vascular connections between the pituitary and hypothalamus (Table 8-12).

**Idiopathic Hyperprolactinemia**

An elevated circulating PRL level in patients in whom no cause is identified is considered idiopathic, and these patients are relatively resistant to dopamine. The mean serum PRL level in 41 patients with idiopathic hyperprolactinemia was 57 µg/L, with only 3 patients having PRL concentrations over 100 µg/L. Patients were observed for up to 11 years, and 33 had both galactorrhea and amenorrhea or galactorrhea and oligomenorrhea. Hyperprolactinemia ultimately resolved spontaneously in 14 patients.

**Macroprolactinemia**

PRL is a 23-kd single-chain polypeptide but may also be produced in higher molecular mass forms (50 and 150 kd). Macroprolactinemia reflects a predominant larger circulating PRL molecule (particularly the 150-kd variety) with markedly reduced bioactivity, and few of the expected clinical abnormalities usually associated with hyperprolactinemia (sexual dysfunction, galactorrhea, osteoporosis) occur. The high-molecular-weight PRL variant may represent 85% or more of the total PRL, whereas under usual circumstances the 22-kd variety predominates. Screening for macroprolactinemia can be accomplished by polyethylene glycol precipitation of serum samples.

**Other Causes of Hyperprolactinemia**

Mild hyperprolactinemia occurs in up to 30% of women with polycystic ovarian syndrome. No definite cause-and-effect relationship between the two disorders is apparent. Dopamine agonists reduced PRL and LH levels in patients with polycystic ovarian syndrome in the presence or absence of hyperprolactinemia, and indeed a subset of patients with amenorrhea experienced a return of menses after treatment with these drugs. Breast stimulation has only a minimal effect on serum PRL levels. In 18 normal women serum PRL rose from a mean of 10 to 15 µg/L during breast pump stimulation, and no increase was observed in men. Up to 20% of patients with hypothyroidism have elevated PRL levels. Although the cause of this elevation is not known, studies of hypothyroid animals suggest increased pituitary TRH. Treatment of hypothyroidism with thyroid hormone normalizes serum PRL if the hyperprolactinemia is due to thyroid hormone deprivation. PRL is moderately elevated (mean 28 µg/L) in patients with chronic renal failure and those receiving dialysis. The increase is largely a result of an increase in little PRL related in part to a decreased glomerular filtration rate. A specific pituitary defect is also suggested by the observation that TRH failed to evoke PRL in these patients. Sexual dysfunction is common, and reducing PRL with dopamine agonists improved sexual function in men receiving dialysis but did not normalize

### TABLE 8-12 — Etiology of Hyperprolactinemia

| Physiologic  |  |
|---------------|  |
| Pregnancy     |  |
| Lactation     |  |
| Stress        |  |
| Sleep         |  |
| Coitus        |  |
| Exercise      |  |

<table>
<thead>
<tr>
<th>Pathologic</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Hypothalamic-Pituitary Stalk Damage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tumors</td>
</tr>
<tr>
<td>Craniohypophyseoma</td>
</tr>
<tr>
<td>Suprasellar pituitary mass extension</td>
</tr>
<tr>
<td>Meningioma</td>
</tr>
<tr>
<td>Dysgerminoma</td>
</tr>
<tr>
<td>Hypothalamic metastases</td>
</tr>
<tr>
<td>Granulomas</td>
</tr>
<tr>
<td>Infiltrations</td>
</tr>
<tr>
<td>Rathke's cyst</td>
</tr>
<tr>
<td>Irradiation</td>
</tr>
<tr>
<td>Trauma</td>
</tr>
<tr>
<td>Pituitary stalk section</td>
</tr>
<tr>
<td>-------------------------</td>
</tr>
<tr>
<td>Suprasellar surgery</td>
</tr>
</tbody>
</table>

**Pituitary**
- Prolactinoma
- Acromegaly
- Macroadenoma (compressive)
- Idiopathic
- Plurihormonal adenoma
- Lymphocytic hypophysitis or parasellar mass
- Macroprolactinemia

**Surgery**
- Trauma

**Systemic Disorders**
- Chronic renal failure
- Polycystic ovarian disease
- Cirrhosis
- Pseudocystis
- Epileptic seizures
- Cranial radiation
- Chest neurogenic chest wall trauma, surgery, herpes zoster

**Pharmacologic**

**Neuropeptides**
- Thyrotropin-releasing hormone
- PRL-releasing peptide

**Drug-Induced Hypersecretion**
- Dopamine receptor blockers
  - Phenothiazines: chlorpromazine, perphenazine
  - Butyrophenones: haloperidol
  - Thioxanthenes
  - Metoclopramide
- Dopamine synthesis inhibitors
  - Methyldopa
- Catecholamine depleters
  - Reserpine

**Cholinergic Agonists**
- Physostigmine

**Antihypertensives**
- Labetolol
- Reserpine
- Verapamil

**H2 Antihistamines**
- Cimetidine
- Ranitidine

**Estrogens**

**Oral Contraceptives**

**Oral Contraceptive Withdrawal**

**Anticonvulsants**
- Phenytoin

**Anesthetics**

**Neuroleptics**
- Chlorpromazine
- Promazine
- Promethazine
- Trifluoperazine
- Fluphenazine
- Butaperazine
- Perphenazine
- Thiethylperazine
- Thioridazine
- Haloperidol
- Pimozide
- Thiothixene
Amenorrhea, oligomenorrhea, primary amenorrhea, infertility

Signs and Symptoms Associated with Hyperprolactinemia

Severe head trauma also results in hyperprolactinemia, often accompanied by diabetes insipidus or SIADH and other anterior pituitary hormone deficiencies. Fifty percent of patients experienced moderate hyperprolactinemia after cranial and hypophalamic radiation. A variety of medications cause minimal or moderate PRL elevations and may cause galactorrhea, amenorrhea, or reduced male sexual function. Neuroleptic drugs elevate PRL because of their dopamine antagonist properties. Chlorpromazine stimulated PRL acutely after an intramuscular injection and chronically during oral administration. Neuroleptics, which act by antagonizing both serotonin and dopamine receptors, including clozapine and olanzapine, weakly induce PRL, whereas others such as risperidone are potent stimulators of PRL.

Treatment of Drug-Induced Hyperprolactinemia

Unless patients exhibit sexual dysfunction, related osteoporosis, or troublesome galactorrhea, no treatment may be advised. It should not always be assumed that hyperprolactinemia in patients receiving drugs known to elevate PRL is due to those medications. Prolactinoma, other pituitary or hypothalamic lesions, hypothyroidism, or renal failure should be considered. In patients taking neuroleptic medications, if the clinical situation permits, temporary drug withdrawal might be considered to determine whether PRL levels become normal. If not, pituitary MRI should be performed. When neuroleptics elevate PRL, olanzapine may be tried because it does not elevate PRL. In determining to discontinue a drug or use an alternative medication, the benefits should be weighed against the risks of drug replacement or cessation. Although combined use of dopamine antagonists and dopamine agonists is not usually advised because of the increased risk of side effects, such as postural hypotension and worsening of psychosis, some advocate the use of both simultaneously.

Galactorrhea

The Talmud describes a man who nursed his baby after his wife's untimely death, probably representing the first recorded case of male galactorrhea. Galactorrhea and amenorrhea were reported in the 19th century by Chiari, and only in the 1950s did Argonz and Forbes and their colleagues associate galactorrhea and amenorrhea with pituitary tumors and PRL. Galactorrhea, inappropriate secretion of milk-like substances from the nipples of either men or women, may persist after childbirth or discontinuation of nursing for as long as 6 months. Thereafter, continued milk production is considered abnormal and other etiologies for galactorrhea should be investigated.

Galactorrhea can occur either unilaterally or bilaterally, be profuse or sparse, and vary in color and thickness. If blood is present in the galactorrhea fluid, it could be the harbinger of an underlying pathologic process, such as a ductal papilloma or carcinoma, and mammography or sonography is indicated. Blood may also appear in galactorrhea fluid with no underlying tumor, such as during pregnancy. Conversely, the absence of blood does not rule out an underlying tumor, particularly when galactorrhea is unilateral and the fluid emanates from a single duct.

The most common cause of galactorrhea is hyperprolactinemia (Fig. 8-23). It is likely that most patients with so-called idiopathic galactorrhea with amenorrhea harbor microprolactinomas. Fifty percent of patients with acromegaly also have hyperprolactinemia. Even in the absence of hyperprolactinemia, human GH is a potent lactogen and can cause galactorrhea when elevated. Twenty-nine of 48 patients with pituitary tumors and galactorrhea had PRL concentrations less than 200 µg/L, suggesting that they had pituitary tumors other than prolactinomas on the basis of stalk compression.

Idiopathic Galactorrhea with Regular Menses

This diagnosis represents the largest single cause of galactorrhea. In two thirds of patients, galactorrhea begins after parturition persists, despite the resumption of menses, and probably does not represent a pathologic entity. Normal PRL levels may still permit milk production because treatment of such patients with dopamine agonists alleviates galactorrhea.

Chiari-Frommel Syndrome

The syndrome, first described by Chiari and later in the 19th century by Frommel, consists of postpartum galactorrhea, amenorrhea, and "uterine-ovarian atrophy" in patients not nursing. Eighteen such patients had galactorrhea and amenorrhea for up to 11 years after parturition. The mean PRL level was 45 µg/L. This disorder is usually self-limiting, and patients eventually become spontaneously fertile, sometimes without having had an intervening menstrual period. Individual patients with postpartum amenorrhea, hyperprolactinemia, and

<table>
<thead>
<tr>
<th>TABLE 8-13 -- Signs and Symptoms of Prolactinomas</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Signs and Symptoms Associated with Tumor Mass</strong></td>
</tr>
<tr>
<td>Visual field abnormalities</td>
</tr>
</tbody>
</table>

Tricyclic antidepressants

- Amitriptyline

Selective serotonin re-uptake inhibitors

- Fluoxetine

Serotonin and dopamine receptors, including clozapine and olanzapine, weakly induce PRL, whereas others such as risperidone are potent stimulators of PRL.
Decreased libido, impotence, premature ejaculation, erectile dysfunction, oligospermia.

Osteoporosis.

The most frequent ophthalmic complaint in a series of 1000 patients with tumors was loss of vision.

Prolactinomas may be found as a result of tumor size or invasiveness, or both. Microadenomas range from entirely asymptomatic tumors as small as 2 to 3 mm in size to compressive effects on visual structures.

PRL also directly inhibits ovarian and testicular function. Increased opioid LH inhibition has also been implicated as a cause of amenorrhea in hyperprolactinemic patients.

PRL levels and tumor size generally remain stable when followed prospectively. Although most men present with no symptoms, symptoms may become apparent after a latency period of 2 to 3 years.

Management of Galactorrhea

Galactorrhea can be overlooked unless actively elicited. Bone density may decrease in both men and women as a result of hyperprolactinemia. In oophorectomized rats, high PRL decreased LH pulse frequency and amplitude.

Pathology and Pathogenesis

Although more than 99% of prolactinomas are benign and often sharply demarcated without evidence of invasion, about half invade local structures. Invasion into adjacent dura, bone, or venous structures may represent an intermediate form of prolactinoma between the sharply demarcated benign variety and the exceedingly rare malignant tumor. Invasive tumors that do not metastasize are considered benign. Immunostaining for PRL confirms the diagnosis of prolactinoma, which is usually distinct from the adjacent normal pituitary but is not truly encapsulated.

Patients with both large and small PRL-secreting tumors can present with signs and symptoms of hyperprolactinemia. Menstrual irregularities, sexual dysfunction, galactorrhea, osteopenia are attributable to elevated PRL levels. Elevated PRL causes sexual dysfunction through a short-loop feedback effect on gonadotropin pulsatility, presumably inhibiting GnRH (see Fig. 8-22) . In oophorectomized rats, high PRL decreased LH pulse frequency and amplitude.

High PRL also directly inhibits ovarian and testicular function. Increased opioid LH inhibition has also been implicated as a cause of amenorrhea in hyperprolactinemic patients.

Pathologic examination of pituitary tumors reveals a cellular proliferation with minimal nuclear atypia and minimal mitotic activity. Immunochemical studies for PRL and other pituitary hormones confirm the diagnosis of prolactinoma.

Prolactinomas are the most common pituitary tumors associated with MEN-1, occurring in approximately 20% of a large kindred.

Invasive tumors may have higher mitotic activity and are more cellular and pleomorphic. Invasion into adjacent dura, bone, or venous structures may represent an intermediate form of prolactinoma between the sharply demarcated benign variety and the exceedingly rare malignant tumor. Invasive tumors that do not metastasize are considered benign. Immunostaining for PRL confirms the diagnosis of prolactinoma, which is usually distinct from the adjacent normal pituitary but is not truly encapsulated. These tumors have a "pseudocapsule" composed of compressed adenohypophyseal cells and a reticulin fiber network.

To consider a prolactinoma malignant, a distant extracranial metastasis must be demonstrated.

For the most part, prolactinomas grow slowly, arise sporadically, usually occur singly, and are monoclonal. Infrequently, more than one prolactinoma arises within the gland. Prolactinomas are the most common pituitary tumors associated with MEN-1, occurring in approximately 20% of a large kindred, although the occurrence of prolactinomas is not evenly distributed. Familial prolactinomas have been described with no other features of MEN-1.

Clinical Features

Prolactinomas are the most common pituitary tumors associated with MEN-1, occurring in approximately 20% of a large kindred. Although more than 99% of prolactinomas are benign and often sharply demarcated without evidence of invasion, about half invade local structures.

Pathology and Pathogenesis

Although more than 99% of prolactinomas are benign and often sharply demarcated without evidence of invasion, about half invade local structures. Invasion into adjacent dura, bone, or venous structures may represent an intermediate form of prolactinoma between the sharply demarcated benign variety and the exceedingly rare malignant tumor. Invasive tumors that do not metastasize are considered benign. Immunostaining for PRL confirms the diagnosis of prolactinoma, which is usually distinct from the adjacent normal pituitary but is not truly encapsulated. These tumors have a "pseudocapsule" composed of compressed adenohypophyseal cells and a reticulin fiber network.

To consider a prolactinoma malignant, a distant extracranial metastasis must be demonstrated.

For the most part, prolactinomas grow slowly, arise sporadically, usually occur singly, and are monoclonal. Infrequently, more than one prolactinoma arises within the gland. Prolactinomas are the most common pituitary tumors associated with MEN-1, occurring in approximately 20% of a large kindred, although the occurrence of prolactinomas is not evenly distributed. Familial prolactinomas have been described with no other features of MEN-1.

Hyperprolactinemia

Patients with both large and small PRL-secreting tumors can present with signs and symptoms of hyperprolactinemia. Menstrual irregularities, sexual dysfunction, galactorrhea, and osteopenia are attributable to elevated PRL levels. Elevated PRL causes sexual dysfunction through a short-loop feedback effect on gonadotropin pulsatility, presumably inhibiting GnRH (see Fig. 8-22) . In oophorectomized rats, high PRL decreased LH pulse frequency and amplitude.

High PRL also directly inhibits ovarian and testicular function. Increased opioid LH inhibition has also been implicated as a cause of amenorrhea in hyperprolactinemic patients.

Women with prolactinomas may present with primary or secondary amenorrhea, oligomenorrhea, menorrhagia, delayed menarche, or regular menses with a short luteal phase that may cause infertility. Patients may also report changes in libido and vaginal dryness. Sexual dysfunction in men is usually manifest as loss of or decrease in libido, impotence, premature ejaculation, or loss of erection. If men have oligospermia or azoospermia as a result of hyperprolactinemia, it usually occurs only after many years.

Up to 50% of women and 35% of men with prolactinomas have galactorrhea. This gender difference may occur because male mammary tissue is less susceptible to the lactogenic effects of hyperprolactinemia. Galactorrhea can be overlooked unless actively elicited. Bone density may decrease in both men and women as a result of hyperprolactinemia-induced sex steroid deficiency.

Tumor Mass Effects

Prolactinomas may be found as a result of tumor size or invasiveness, or both. Microadenomas range from entirely asymptomatic tumors as small as 2 to 3 mm in diameter found at autopsy to larger ones that are still less than 10 mm in diameter. These tumors can be invasive despite their small size. The incidence of headaches in patients with a microadenoma is twice that of normal control subjects. In contrast, macroadenomas range in size from noninvasive or diffuse tumors approximately 1 cm in diameter to huge tumors that may impinge on paraseptal structures. Signs and symptoms caused by large or invasive tumors are often related to compressive effects on visual structures.

The most frequent ophthalmic complaint in a series of 1000 patients with tumors was loss of vision. The most frequent objective findings were bitemporal hemianopsia, superior bitemporal defects, and decreased visual acuity. Headaches are common, but seizures (a result of extension into the temporal lobe) and hydrocephalus are rare, as is unilateral exophthalmos. Interestingly, many tumors invade the cavernous sinuses and yet cranial nerve palsies are rarely encountered. A sudden insult, such as pituitary apoplexy, is the more common cause of such palsies and may be a presenting symptom. Prolactinomas can also be found inadvertently by MRI or CT performed for another purpose.
Bromocriptine shrinks prolactinomas by reducing tumor cell size, including cytoplasmic, nuclear, and nucleolar areas. An agonist has been reported to normalize PRL further in some cases, although these patients have impressive tumor shrinkage and sometimes improved sexual function. Although higher doses or a change in the form of dopamine agonist (such as bromocriptine) binding to dopamine receptors on cell membranes. Despite high doses of bromocriptine, some patients are entirely or partially resistant to its effects. In bromocriptine-resistant patients, there is reduced bromocriptine elevation. Bromocriptine lowers PRL despite continued tumor expansion, do not become larger after drug withdrawal. Not infrequently, it is difficult to normalize PRL levels completely in patients with initially very high levels, although when tumors grow during dopamine agonist therapy there is usually a simultaneous PRL elevation. Despite high doses of bromocriptine, some patients are entirely or partially resistant to its effects. In bromocriptine-resistant patients, there is reduced bromocriptine elevation. Therefore, serum IGF-I should be measured. Elevated PRL levels are occasionally encountered in patients with TSH-secreting tumors. Other pituitary hormone functions should be ascertained to determine the presence of hypopituitarism. MRI is required to make a definitive diagnosis of a prolactinoma.

**Treatment**

Optimal outcomes of treatment for a prolactinoma include normalization of PRL levels (and associated signs and symptoms) and complete tumor removal or shrinkage with a reversal of tumor mass effects (Table 8-14). Specifically, previously abnormal sexual function and fertility should be restored, galactorrhea stopped, impaired bone density improved, tumor eliminated or reduced in size without impairing pituitary or hypothalamic function, and vision normalized, if impaired.

**Medical Management**

Medical management of prolactinomas with dopamine agonist drugs has been widely recommended as the treatment of choice.

**Bromocriptine.**

Bromocriptine, a semisynthetic ergot alkaloid dopamine agonist, lowers elevated PRL levels, restores abnormal menstrual function in 80% to 90% of patients, and shrinks prolactinomas, restores impaired sexual function, and improves galactorrhea. Improvement in visual field abnormalities occurs in approximately 90% of affected patients. Drug withdrawal can result in rapid tumor expansion. In contrast, occasional tumors that have shrunk during bromocriptine therapy do not become larger after drug withdrawal. In a subset of patients, hyperprolactinemia disappeared spontaneously after long-term observation. Occasionally, bromocriptine lowers PRL despite continued tumor expansion, although when tumors grow during dopamine agonist therapy there is usually a simultaneous PRL elevation.

**Histologic sections appear quite dense as a result of the small cell size and clumping of nuclei** (Fig. 8-25). PRL mRNA and synthesis is inhibited, exocytoses are reduced, PRL secretory.
granules decrease, and rough endoplasmic reticulum and Golgi apparatus involute. The net effect is reduced cell volume. Tumor necrosis may also occur.

Perivascular fibrosis was noted in prolactinomas derived from patients treated with bromocriptine and it was proposed that this led to difficulty in tumor removal. However, others found no effect of prior treatment with bromocriptine on surgical success rates. In contrast, bromocriptine was a helpful adjunct to transphenoidal microsurgery for macroadenomas. Even the largest tumors or those with the highest PRL levels respond well to treatment with 2.5 mg of bromocriptine three times daily. Higher doses are often not more effective. When positive effects on tumor size and amenorrhea and galactorrhea are established, some patients can be satisfactorily maintained with smaller doses but rarely without medication.

CABERGOLINE

CABERGOLINE has a longer duration of action than other available dopamine agonists and is usually administered once or twice weekly. Since its introduction, it has surpassed bromocriptine as the first-line therapeutic choice for most patients. The long half-life of cabergoline is a result of its high affinity for D2 receptors on lactotrophs and a greater propensity of the drug to remain in pituitary tissue.

In pharmacokinetic studies, cabergoline lowered PRL in a dose-related manner. PRL levels were normalized in 83% of 459 women with hyperprolactinemia treated with cabergoline (0.5 to 1 mg twice weekly) and in 52% of women receiving bromocriptine (2.5 to 5 mg twice daily). Cabergoline was also more effective than bromocriptine in restoring ovulatory cycles and fertility (72% versus 52%, P < 0.001), was better tolerated than bromocriptine, and caused fewer but similar side effects (Fig. 8-26) and as evidenced by the success of pergolide in patients intolerant of bromocriptine. Intravaginal bromocriptine administration has been used with some success in patients who are intolerant of bromocriptine but cannot tolerate cabergoline.

PERGOLIDE MESYLATE

Pergolide, a long-acting ergot derivative with dopamine agonist properties, has an estimated potency 100 times that of bromocriptine. Pergolide is administered at an initial dose of 25 µg/day, and then at 50 µg/day with gradual dose escalation depending on the extent of serum PRL normalization. Menses resumed in 76% of women and serum testosterone increased in 10 of 14 men not receiving testosterone, and tumor size decreased in 10 of 13 patients with macroadenomas. In 22 patients with macroadenomas, pergolide lowered PRL from a mean of 2938 to 59 ng/mL. PRL became normal in 15 of 22 patients, and tumor shrinkage was observed in 95% of patients. Nine patients were resistant to cabergoline despite doses of up to 7 mg/week. Despite the continued experience that a subset of hyperprolactinemic patients are resistant to most or all dopamine agonists, a report indicated that cabergoline normalized PRL in 15 of 19 patients with macroadenomas previously resistant to other dopamine agonists.

QUINAGOLIDE

This nonergot dopamine agonist (CV 205502), administered once daily (mean daily dose of 0.09 mg), normalized PRL in 5 of 10 patients previously intolerant of or resistant to bromocriptine. Side effects, principally nausea, occurred in 6 of 10 women. In 26 patients similarly evaluated, quinagolide normalized PRL levels in 13 but the post-treatment mean remained above normal (30 ng/mL). Thirteen had a return of menses, and galactorrhea was reduced in 12 of 15 women. The effect of this medication on tumor shrinkage is similar to that of other dopamine agonists, but it is not available in the United States.

Administration

Attention to administration of dopamine agonists helps avoid or minimize potential adverse effects. Usual starting doses are 1.25 mg of bromocriptine (daily), 0.025 mg of pergolide (daily), and 0.25 mg of cabergoline (weekly). Doses of medication are either increased gradually as tolerated or decreased depending on tolerability, and treatment should be initiated with a small dose with food before bedtime. Patients should initially avoid activities that cause peripheral vasodilatation (e.g., hot showers or baths), thereby decreasing the risk of postural hypotension. If side effects are troublesome, the next dose should be halved and doses subsequently increased gradually to reach effective levels. Switching from one medication to another may be beneficial as evidenced by the success of pergolide in patients intolerant of bromocriptine. Intravaginal bromocriptine administration has been used with some success to reduce adverse events.

Adverse Events of Dopamine Agonists

Side effects of dopamine agonists are common. Nausea occurs in 31% to 50% of patients; nasal stuffiness, depression, and digital vasospasm also occur, the latter more frequently with higher doses, as in patients with Parkinson’s disease. The most serious side effect, postural hypotension, which can cause loss of consciousness, occurs infrequently and can often be avoided by careful dosing. Signs and symptoms of psychosis or exacerbation of preexisting psychosis can be
encountered in up to 1.3% of patients taking bromocriptine. \textsuperscript{[360]} Psychosis also occurs with other dopamine agonists, including cabergoline (personal experience).\textsuperscript{[361]}

A history of present or past psychotic symptoms should raise concerns about using these medications. If psychosis occurs in a patient in whom dopamine agonists are clearly the treatment of choice, a judicious combination of the agent and antipsychotic medication can be effective. A neuroleptic that is not a potent PRL stimulator, such as olanzapine, is preferred. The combined use of dopamine agonists and antagonists may increase the occurrence of side effects, particularly postural hypotension. Other rarely reported serious side effects include CSF rhinorrhea, \textsuperscript{[362]} hepatic dysfunction, \textsuperscript{[363]} and cardiac arrhythmias. \textsuperscript{[364]} Retroperitoneal fibrosis, pleural effusions, and thickening have been reported in patients taking high doses of bromocriptine. \textsuperscript{[365]}

\subsection*{Radiation Therapy}

Linear accelerator radiotherapy is effective in controlling or reducing the size of prolactinomas. \textsuperscript{[366]} \textsuperscript{[367]} However, this therapy takes years to achieve its maximal effect. The usual recommended radiation dose is 4500 to 4600 cGy, and higher doses are associated with a greater complication rate. \textsuperscript{[368]} Normalization of PRL was achieved in 7 of 12 patients during 3 to 8 years after radiotherapy\textsuperscript{[369]} and in 18 of 36 patients at a mean of 7.3 years after treatment in another study. \textsuperscript{[370]}

Hypopituitarism is a side effect of radiation. Of 165 patients after radiotherapy (3750 to 4250 cGy), \textsuperscript{[371]} all patients were GH-deficient, 91% were gonadotropin-deficient, 77% were

\begin{center}
212
\end{center}

ACTH-deficient, and 42% were TSH-deficient by 5 years. Of 36 patients with prolactinomas, 83% of whom had normal GH responses to insulin-induced hypoglycemia before therapy, 34 were GH-deficient 9 to 12 years after radiotherapy. \textsuperscript{[372]} The incidence of other forms of hypopituitarism was lower.

Thus, although radiotherapy is useful in the control of tumor growth, it is not nearly as effective as dopamine agonists on endocrine function. Stereotactic conformal radiotherapy with a linear accelerator can provide greater tumor focus and a smaller radiation field. \textsuperscript{[373]} There are as yet no reports on large-scale studies of treatment of prolactinomas with gamma knife radiotherapy.

\subsection*{Surgery}

Surgical removal of prolactinomas by the transphenoidal route was repopularized in the early 1970s. \textsuperscript{[374]} As with other functioning pituitary tumors, the success rate of surgery correlates inversely with tumor size and serum PRL concentrations. \textsuperscript{[375]} \textsuperscript{[376]} In a compilation of results in 31 published surgical series, serum PRL was normalized in 71% of 1224 patients with microprolactinomas. \textsuperscript{[377]} Although surgical cure rates for microprolactinomas are high, the rate of hyperprolactinemia recurrence is also relatively high. \textsuperscript{[378]} now estimated to be 17% in patients initially considered cured. \textsuperscript{[379]} In contrast, complete removal of macroprolactinomas, especially large invasive ones, is difficult to achieve; postoperative serum PRL was normalized in only 32% of patients with macroadenomas, with a recurrence rate of 19%. The experience of the surgeon is of major importance, as the cure rate is not nearly as favorable for neurosurgeons who perform a limited number of procedures.\textsuperscript{[380]}

Although results of medical therapy are better than those of surgery, there remains a role for surgery in these patients. Patients with prolactinomas who are resistant to dopamine agonist therapy are particularly well suited for surgery. If tumor removal is only partial, adjunctive radiation therapy should be considered. Prophylactic transphenoidal surgery should also be considered in women whose prolactinomas are large enough to be a potential threat to vision during pregnancy. A subset of patients cannot tolerate available dopamine agonists, and others prefer surgery and refuse medication \textsuperscript{[Fig. 8-27]}.

\subsection*{Pregnancy}

The normal pituitary gland enlarges during pregnancy and by the end of pregnancy may increase in size by 136%. \textsuperscript{[381]} Prolactinomas may also increase in size during pregnancy.\textsuperscript{[382]} Pregnancy-associated tumor enlargement, as determined by the development of abnormal visual fields, has been estimated to occur in 1.4% of women with microadenomas and 16% of women with macroadenomas.\textsuperscript{[383]} In other reports, the risk of macroadenoma enlargement has been estimated to be as high as 36%. In a prospective analysis in which 57 patients with microprolactinomas were observed by formal visual field examinations during pregnancy, none experienced visual disturbances. In contrast, six of eight primiparous women with macroadenomas had visual loss.\textsuperscript{[384]} The results for patients with macroadenomas are probably skewed because these patients were recommended for surgery before pregnancy.

Although dopamine agonists have been used during pregnancy to prevent tumor growth \textsuperscript{[Fig. 8-28]}, \textsuperscript{[385]} \textsuperscript{[386]} it seems prudent to reduce fetal exposure to medication if possible. It is recommended that menstrual periods be allowed to occur naturally for a period of time (3 to 4 months) long enough to predict that a missed period might be a result of pregnancy (Table 8-15). Barrier contraception is recommended during this period. Within several days to a week of obtaining a positive hCG test, medication should be discontinued. In 6239 pregnancies of patients managed in this manner, bromocriptine therapy was not associated with increased abortions or terminations, prematurity, multiple births, or infant malformations above those expected in the control population.\textsuperscript{[387]} There is no evidence that other dopamine agonists are less safe, but exposure to the other agonist forms in pregnancy is less comprehensively documented. Treatment options for patients harboring prolactinomas whose vision becomes impaired during pregnancy include bromocriptine, high-dose steroids, and surgery. \textsuperscript{[388]} \textsuperscript{[389]} One study reported that for 53 pregnant women receiving bromocriptine, mean offspring birth weight was normal, congenital abnormalities occurred in four babies, and the physical and intellectual development of the children was normal for up to 9 years.\textsuperscript{[390]}

To avoid neurologic complications of tumor enlargement during pregnancy, it is recommended that women with prolactinomas be tested for sensitivity to dopamine agonists before proceeding with a pregnancy. If tumors are insensitive to dopamine agonist-related tumor shrinkage, prophylactic surgery would be appropriate. If the tumor is a macroadenoma approximating the optic chiasm, the likelihood of visual difficulties is greater and therefore surgery would be prudent before pregnancy. \textsuperscript{[391]}
PHYSIOLOGY AND DISORDERS OF PITUITARY HORMONE AXES (Continued)

Gonadotropins

Gonadotroph cells secreting FSH and LH constitute about 10% to 15% of the functional anterior pituitary cells. Two classes of electrodense secretory granules are evident; large 350- to 450-nm and smaller 150- to 250-nm granules are packaged in vesicles (Fig. 8-29 (Figure Not Available) and Fig. 8-30 (Figure Not Available) ). They contain large, round cell bodies with prominent rough endoplasmic reticulum and Golgi apparatus. LH secretory granules often accumulate peripherally, and their Golgi may be less prominent. SF-1 and DAX-1 orphan nuclear receptors determine gonadotroph-specific gene expression.

Biasynthesis

In concert with peripheral hormones and paracrine soluble factors, FSH and LH function to regulate gonadal steroid hormone biosynthesis and initiate and maintain germ cell development. The four glycoprotein hormones LH, FSH, TSH, and hCG share structural homology, having evolved from a common ancestral gene. Although the homologous LH and FSH molecules are cosecreted by the single gonadotroph cell, their regulatory mechanisms are not uniformly concordant. The and subunits are encoded by different genes located on chromosomes 6, 11, and 19 (Fig. 8-4) (Figure Not Available) . The heterodimeric structure of the common and unique subunit is essential for their biologic activity. Disulfide linkages maintain noncovalent subunit linkages, which also determine the ultrastructure of the mature folded molecule (Fig. 8-31). After processing of hormonal protein precursors, glycosylation occurs by transferring oligosaccharide complexes to asparaginyl residues. Post-translational processing of carbohydrate side chains is critical for hormone signaling and may be species specific and not uniformly similar for both human LH and FSH.

The complex human LH-gene cluster comprises seven CG-like genes, one of which encodes LH-7 whose promoter and transcriptional start site differ from those of hCG. The three exons and two introns encode a 24-amino-acid leader peptide and a 121-amino-acid mature protein. Unlike -LH, hCG is present only in primate and equine species, and the hCG peptide product contains a 24-amino-acid carboxyterminal extension. Cell-specific LH gene expression and GnRH responsiveness of LH are subverted by different transcriptional mechanisms. GnRH induces LH-transcription, as does SF-1. The rat LH-promoter contains an estrogen-responsive motif and a nuclear factor Y binding site, which appears to be important for basal but not GnRH-mediated transcription. The FSH-gene comprises three exons and two introns located on chromosome 11. The gene promoter is dissimilar to that of LH, and structure-function mechanisms for transcriptional regulation of the human gene by GnRH and sex steroids are not well clarified.

Gonadotropin Assays

Because of the high homology of the glycoprotein hormones, development of highly specific assays, especially to distinguish free subunit from intact hormones, has been challenging. Heterogeneity of circulating LH and FSH molecules, insufficient assay sensitivity especially for measurements in normal healthy individuals, and lack of rigorously pure reference preparations have hampered assay development. Immunofluorometric assays detect LH with a sensitivity of 0.1 mIU/mL. Differences in carbohydrate moieties result in isoelectric charge heterogeneity for LH, accounting for some of the disparities in biologic and immunoreactive LH ratios observed with GnRH agonist treatment, acute critical illness, or aging.

LH bioassays include assessing testosterone generation by cell cultures, and FSH bioassays include measuring granulosa cell or Sertoli cell aromatase generation. Because only intact molecules, but not free or subunits, are biologically active, these cumbersome assays are nevertheless useful for measuring bioactive hormone without potential cross-reaction with free subunits.

Subunit Secreton

Both GnRH and TRH increase circulating levels of free subunit derived from either gonadotrophs or thyrotrophs, especially in patients with hypothyroidism, after castration, and during the menopause. GnRH agonist treatment, TSH-secreting tumors, or nonfunctioning pituitary adenomas may result in discordant ratios of free subunit from intact LH dimer secretion.

Regulation of FSH and LH Secretion

FSH and LH secretion patterns reflect the integration of sensitive complex hypothalamic, pituitary, and peripheral signals. Both GnRH pulse amplitude and frequency determine the physiologic patterns of LH and FSH secretion (Fig. 8-32). In patients with hypothalamic GnRH deficiency, intravenous GnRH injections (25 ng/kg) that achieve GnRH levels similar to those present in primate hypophysial-portal

| TABLE 8-15 -- Management of Prolactinomas in Patients Planning Pregnancies |
|------------------|------------------|
| **Microadenoma** | **Macroadenoma** |
| Discontinue dopamine agonist when pregnancy test positive | Consider surgery prior to pregnancy |
| Periodic visual field examinations during pregnancy | Ensure bromocriptine sensitivity prior to pregnancy |
Follow visual fields expectantly and frequently

Administer bromocriptine if vision becomes compromised

Or, continue bromocriptine throughout pregnancy if tumor previously affected vision

Consider high-dose steroids or surgery during pregnancy if vision threatened or adenoma hemorrhage

Postpartum MRI after 6 weeks

*Pituitary MRI may be required during pregnancy if deemed necessary. MRI, magnetic resonance imaging.

---

The magnitude of the LH response exceeds that of FSH. Decreasing GnRH pulse frequency enhances LH pulse amplitudes, whereas increasing GnRH pulse frequency to more than every 2 hours down-regulates the subsequent LH response. The interpulse LH secretory interval is 55 minutes, and the pulse amplitude is approximately 40% of basal tonic secretion. Changes in gonadotropin secretion during infancy, childhood, puberty, and aging are described in Chapter 21 and Chapter 25 (Fig. 8-33). A log-linear relationship is evident between GnRH dose and the amounts of LH, FSH, and free subunit pituitary secretion.

Blood and hypothalamic targets for testosterone signals mediate FSH and LH regulation, and testosterone attenuates gonadotropin secretion in males. Thus, after castration, elevated gonadotropin levels can be partially overcome by testosterone replacement. The mechanisms involved are complex, as testosterone also exerts a stimulatory effect on FSH- mRNA levels. Although estrogen administration decreases LH pulse amplitude in normal and GnRH-deficient male subjects, depending on the clinical situation, estrogen may either stimulate or inhibit pituitary gonadotropin synthesis and secretion and also inhibit GnRH synthesis and or action. This pattern is manifest in the cyclic control of gonadotropin secretion during the menstrual cycle and during puberty.

LH, FSH, free subunit, and testosterone pulses are usually concordant in male subjects (Bhasin 2002). Deconvolution pulse analysis allows estimation of "real-time" hormone secretion rates, with an assumed disappearance rate constant. The characteristic secretory episodes characterized for LH and FSH indicate daily production rates of 1000 and 200 IU, respectively, and a disappearance half-life of 90 and 500 minutes for the respective subunits.

**Gonadal Peptides**

Pituitary gonadotropin secretion is regulated by gonadal peptides including inhibin A, an :A heterodimer, and inhibin B, an :B heterodimer, and follistatin peptides. The activin A (A) and activin AB (B) homodimers stimulate in vitro FSH secretion. These proteins, related to transforming growth factor and müllerian-inhibiting factor, are fully described in Chapter 16.

Female

Luteal cell LH receptors signal to enhance cAMP levels and induce cholesterol availability for ovarian steroidogenesis (Fig. 8-34). The steroidogenic acute regular (STAR) protein is induced by LH and mediates cholesterol delivery to the inner membrane. LH enhances cytochrome P450-linked enzyme activity to synthesize pregnenolone and induces 3-hydroxysteroid dehydrogenase, 17-hydroxylase, and 17,20-ylase synthesis. The FSH receptor, a G protein-linked molecule with seven transmembrane domains shares 50% extracellular domain and 80% transmembrane homology with the LH receptor. FSH regulates ovarian estrogen synthesis by inducing 17-hydroxysteroid dehydrogenase and aromatase and also induces follicular growth. Estrogens are also permissive for FSH action and enhance FSH-induced cAMP levels.

Mole

Leydig cell LH receptor signaling induces intratesticular testosterone synthesis mediated by enhanced cAMP production. FSH function in male subjects is not readily apparent but probably mediates spermatogenesis from spermatids in concert with testosterone, especially as failed spermatogenesis leads to elevated FSH levels.

**Gonadotropin-Releasing Hormone Stimulation Test**

A single bolus of GnRH (25 to 100 µg) dose dependently evokes serum LH and FSH levels within 20 to 30 minutes. LH rises more abundantly than FSH, and peak values range from 8 to 34 mIU/mL. Patients with low testosterone levels exhibit more exuberant responses. In contrast, patients with hypogonadotropic hypogonadism and no demonstrable hypothalamic-pituitary lesion have blunted LH responses and reversal of the LH/FSH ratio. The test, however, cannot adequately distinguish hypothalamic from pituitary lesions, and similar...
patterns are observed in patients with anorexia nervosa. Repetitive GnRH pulses may, in fact, normalize responses, as would be expected from an intact hypothalamic-pituitary unit. GnRH responses may vary during the stages of puberty, reflecting altered pituitary sensitivity.

Gonadotropin Deficiency

Gonadotropin deficiency causes hypogonadism with decreased sex steroid production of varying degree, depending on the severity of the insult (Table 8-16). This disorder may occur at any stage of life. In its complete form (e.g., panhypopituitarism, Kallmann's syndrome), primary amenorrhea or total failure of male sexual development may occur. Later in life a varying spectrum of sexual dysfunction develops, ranging from luteal abnormalities or oligomenorrhea to amenorrhea in women and absence of libido, potency, and fertility in men. Women exhibit secondary amenorrhea, vaginal dryness, hot flushes, decreased bone density, decreased breast tissue, and infertility. Men have impotence, testicular hypoplasia or atrophy, decreased libido, low energy, infertility, loss of secondary sexual characteristics, decreased muscle strength and mass, decreased bone mass, decreased body hair growth, and fine facial wrinkling.

In both men and women, serum gonadotropin levels are inappropriately low in the presence of decreased sex steroids and sexual dysfunction. In women with amenorrhea or oligomenorrhea, serum LH, FSH, and estradiol levels should be measured. A vaginal cytologic study is helpful in determining the adequacy of gonadotropin function. Endogenous estrogen

![Figure 8-32 Effects of pulsatile or continuous administration of gonadotropin-releasing hormone (GnRH) to ovariectomized monkeys rendered GnRH-deficient by placement of a lesion in the hypothalamus. Gonadotropin secretion was restored by hourly GnRH pulses, reduced during a continuous GnRH infusion, and again increased after reinstitution of pulsatile GnRH administration. FSH, follicle-stimulating hormone. (From Belchetz PE, Plant TM, Nakai Y, et al. Hypophyseal responses to continuous and intermittent delivery of hypothalamic gonadotropin-releasing hormone. Science 1978; 202:631633. Copyright 1978 by the American Association for the Advancement of Science.)](image)

![Figure 8-33 Serum luteinizing hormone (LH) levels (open symbols) and follicle-stimulating hormone (FSH) levels (solid symbols) in men as a function of age from three studies. (From Tenover JL. Male hormone replacement therapy including "andropause." Endocrinol Metab Clin North Am 1998; 27:369387; Bhasin. In Melmed S (ed). The Pituitary, 2nd ed. Malden, Mass, Blackwell Scientific, 2002.)](image)


<table>
<thead>
<tr>
<th>Prepubertal Onset</th>
<th>Postpubertal Onset</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-pitched voice</td>
<td>Decreased libido</td>
</tr>
<tr>
<td>Terminal facial hair</td>
<td>Slow beard growth</td>
</tr>
<tr>
<td>Decreased or absent body hair</td>
<td>Decreased body hair</td>
</tr>
<tr>
<td>Eunuchoidal body proportions</td>
<td>Testes atrophic if long-standing</td>
</tr>
<tr>
<td>Female escutcheon</td>
<td>Normal voice pitch</td>
</tr>
<tr>
<td>Testicular volume &lt;6 cm³, hypoplastic</td>
<td>Decreased muscle and bone mass</td>
</tr>
<tr>
<td>Testicular length &lt;2.5 cm</td>
<td>Normal: skeletal proportions, penis length, scrotal rugae, prostate size</td>
</tr>
<tr>
<td>Penile length &lt;5 cm</td>
<td>Normal: skeletal proportions, penis length, scrotal rugae, prostate size</td>
</tr>
<tr>
<td>Smooth scrotum with no rugae</td>
<td>Normal: skeletal proportions, penis length, scrotal rugae, prostate size</td>
</tr>
</tbody>
</table>

**TABLE 8-16 -- Clinical Features of Hypogonadotrophism**

**TABLE 8-17 -- Mutations Affecting Genes of the Hypothalamic-Pituitary-Gonadal Axis**

<table>
<thead>
<tr>
<th>Hypothalamus</th>
<th>Anterior Pituitary</th>
<th>Ovary</th>
<th>Testes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene</td>
<td>Phenotype</td>
<td>Gene</td>
<td>Phenotype</td>
</tr>
<tr>
<td>Kal-Y</td>
<td>HH</td>
<td>GnRH-R</td>
<td>Partial or complete</td>
</tr>
</tbody>
</table>

217
Anosmia | HH | Delayed puberty
---|---|---
Dax-1 | HH | Dax-1 | HH | LH-R | Normal puberty | LH-R | Pseudohermaphroditism
AHC | AHC | PC-1 | HH | FSH- | Primary amenorrhea | DAX-1 | Genital ambiguity
Amenorrhea | Defective spermatogenesis | Amnorrhea | Defective spermatogenesis
Obesity | LH-R | Delayed puberty | AR | Androgen insensitivity
Lep-R | HH | DAZ | OTA | Lep | LH- | Delayed puberty | RBM | OTA | Short stature

AR=Androgen receptor; DAZ=deleted in azoospermia; RBM=RNA binding motif protein; OTA=oligoteratoazoospermia; HH=hypogonadotropic hypogonadism.


sufficiency can also be assessed by the response to a progesterone challenge (100 mg intramuscularly or 10 mg medroxyprogesterone [Provera] orally daily for 5 days). Men should have serum gonadotropin and testosterone levels measured.

Hypogonadotropic hypogonadism may result from hypothalamic or pituitary defects. Hypothalamic damage including Kallmann’s syndrome, radiation, anosmia, or sexual stress may result in deficient GnRH secretion and action. Pituitary damage from tumors, infarction, or hyperprolactinemia may directly or indirectly attenuate FSH and LH secretion patterns. These acquired causes of central hypogonadism are considered fully in Chapter 16 and Chapter 18 , and genetic causes are listed in Table 8-17 .

Because FSH is required for quantitatively normal spermatogenesis, isolated FSH deficiency is associated with oligospermia or azoospermia in normally androgenized men in the presence of normal testosterone and LH levels. Isolated LH deficiency may be manifest by eunuchoidal body proportions and low testosterone levels. Low LH levels in these patients lead to low intratesticular testosterone concentrations with resultant decreased spermatogenesis. Serum testosterone levels are restored in this “fertile eunuch” syndrome by HCG administration. Isolated hypogonadotropic hypogonadism occurring after apparently normal puberty is manifest with relatively mature secondary sex characteristics or even secondary infertility. These patients have abundant gonadotropin responses to pulsatile GnRH therapy, which restores reproductive function and fertility.

Evaluation

In evaluating hypogonadal patients in the absence of an obvious pituitary or gonadal disorder, the primary diagnostic challenge is to distinguish constitutional pubertal delay from other causes of hypogonadotropism. When puberty is delayed beyond 14 years of age, a primary developmental disorder, hypogonadotropic hypogonadism, or acquired disorders of reproductive function should be considered. The presence of midline defects, a pituitary mass lesion, anosmia, a history of radiation or pituitary damage, drug ingestion, or other systemic illness should be excluded. Patients with chronic liver disease or sickle cell disease may present with impaired pubertal reserve as well as primary testicular dysfunction.

No single test clearly distinguishes constitutional delayed puberty and true hypogonadotropic hypogonadism, and expectant follow-up is often helpful as many patients enter puberty spontaneously. To provide androgenization, testosterone replacement should be intermittently provided until age 18 with periodic interruptions to unmask physiologic pubertal advance. In adults presenting with features of hypogonadotropic hypogonadism, hyperprolactinemia, hemochromatosis, and sarcoidosis should also be excluded before the diagnosis of the idiopathic variety. Pituitary MRI, GH, TSH, and ACTH reserve testing should be performed selectively.

Management

Sex steroid replacement therapy is required to induce and maintain primary and secondary sexual functions, minimize cardiovascular risk factors, and maintain normal body composition and integrity of bone mineral density and muscle mass. For patients not desirous of fertility, sex steroid therapy is warranted to correct central hypogonadism. However, monitoring of LH and FSH responses does not accurately reflect adequate steroid hormonal replacement because basal gonadotropin levels are already low or undetectable.

For women, estrogens are administered as a tablet, patch, gel, or implant. For premenopausal women with pituitary deficiency, a combined oral contraceptive (20 to 35 µg ethynyl estradiol) may be used. Conjugated equine estrogen (0.625 mg) and estradiol valerate (2 mg) provide relatively physiologic steroid replenishment. Estrogen patches or transcutaneous gels usually enable daily absorption of 50 to 100 µg of estradiol. Concomitant cycloprogesterone therapy is indicated for women with an intact uterus to prevent unopposed endometrial proliferation and bleeding.

Although early replacement lowers the risk of developing osteoporosis, effects of estrogen replacement on cardiovascular function are unresolved. In patients with hypopituitarism, estrogen replacement should be maintained at least until the age of 50, after which continuation should be determined on an individual basis by assessing risks and benefits especially in terms of bone mineral integrity, cardiovascular function, and cancer risk. In women with ovarian deficiency, combined estrogen and testosterone replacement may improve libido and sexual function. Estrogen treatment may be associated with thromboembolic disease, breast tenderness, and possibly an enhanced risk of breast cancer.

For males not desiring fertility, intramuscular injection of testosterone 17-hydroxy esters (testosterone enanthate and testosterone cypionate) at 200 mg intramuscularly every 2 or 3 weeks is effective but may be associated with fluctuations in

sexual potency, energy level, and mood reflecting dynamic changes in circulating testosterone concentrations. Administration of lower doses more frequently (e.g., 100 mg weekly or 150 mg every 14 days) may stabilize hormone fluctuations. Elderly men require lower doses, as do boys with delayed puberty. Scrotal and nonscrotal transdermal testosterone patch systems deliver 4 to 6 mg and sustain testosterone profiles. These preparations require adequate shaved scrotal skin for application. Nonscrotal patch sites may develop skin irritation, blisters, and vesicles in about 25% of patients. Some patients require combined patches or low-dose injection to maintain adequate potency and energy levels. There is no apparent cost-benefit advantage of patch delivery over intramuscular injection. Oral androgen replacement therapy with 17-hydroxy ester testosterone undecanoate requires frequent dosing (two to four times daily) and, because absorption is not uniform, testosterone levels may not be adequately maintained. Oral 17-methyltestosterone is associated with hepatotoxicity and not recommended. Testosterone may cause acne, gynecomastia, rarely urine retention related to prostatic obstruction, and polycythemia. Although there is no compelling evidence that testosterone replacement causes prostate cancer, benign prostatic hypertrophy may be exacerbated, especially in elderly patients. Testosterone replacement should not be administered to men with diagnosed prostate cancer.

In patients with hypogonadotropic hypogonadism, fertility may be achieved with gonadotropin or GnRH therapy. In males, coexistence of primary testicular dysfunction precludes the success of direct gonadotropin replacement or GnRH, although the relatively low sperm counts induced may be adequate for impregnation when fertility is induced by gonadotropin or GnRH. Because testosterone therapy may suppress spermatogenesis, the steroid should be discontinued prior to initiating treatment. To induce spermatogenesis, NCG is administered subcutaneously or intramuscularly (500 to 2000 IU two to three times weekly). Lower doses may also be effective. If necessary, after 6 months, human menopausal gonadotropin or purified FSH (75 IU three times weekly) should be added to improve sperm quantity, and doses may be doubled after a further 6 months. If testosterone levels are increased, subsequent conversion to estradiol may be enhanced, resulting in
**TABLE 8-18** -- Presentation of Gonadotroph Adenomas

<table>
<thead>
<tr>
<th>Common</th>
<th>Uncommon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinically nonfunctioning macroadenomas</td>
<td>Intact gonadotropin overproduction</td>
</tr>
<tr>
<td>Immunostain for gonadotropin subunits (usually more than one)</td>
<td>Immunostain for subunits or intact hormone being hypersecreted</td>
</tr>
<tr>
<td>Usually discovered because of space-occupying effects or inadvertently</td>
<td>Usually discovered because of space-occupying effects or inadvertently</td>
</tr>
<tr>
<td>Pituitary deficiency</td>
<td>May cause clinical syndrome related to hormone overproduction</td>
</tr>
<tr>
<td></td>
<td>Other pituitary hormones may be deficient</td>
</tr>
</tbody>
</table>

but are not often associated with specific endocrine syndromes. High serum FSH, usually with low LH levels, is usually the only sign that a pituitary tumor secretes FSH. Paradoxically, these patients may present with hypogonadism related to gonadal down-regulation. Female patients with such tumors may present with pelvic pain caused by ovarian hyperstimulation. High gonadotropin levels associated with menopause or testicular failure may complicate interpretation of gonadotropin levels, but both LH and FSH are high in primary gonadal failure. LH-producing tumors are exceedingly rare and in males cause elevations of serum testosterone with acne and skin oiliness.

**Evaluation**

MRI, visual field examination, and pituitary hormone evaluation should be performed, the last not only to detect hypopituitarism but also to exclude hormone overproduction that may not be clinically apparent. LH, FSH, subunit, PRL, T4, triiodothyronine (T3), TSH, cortisol, and IGF-I levels should be measured. A 24-hour urinary free cortisol measurement by RIA is useful to exclude inapparent ACTH hypersecretion. The extent of hormonal evaluation requires clinical judgment. When LH or FSH is elevated, the values must be interpreted in light of the patient's physiologic state.

**Treatment**

Clinical judgment should be used in determining appropriate therapy including surgery, surgery followed by radiotherapy, radiotherapy alone, or expectant observation. Unfortunately, no reliable tumor marker is predictive of mass growth or recurrence (Fig. 8-35).

If tumors threaten vision or are macroadenomas whose size threatens vital structures, transphenoidal surgery is recommended. Vision improved in approximately 75% of patients whose vision was impaired. Of 100 patients undergoing transphenoidal surgery, 72 had visual disturbances, 81 hypopituitarism, and 36 headache. Vision improved in 53 of 72 patients after surgery, and headache improved in all. Of 50 patients who underwent surgery followed by radiotherapy, 9 had tumor recurrences at a mean of 73 months after radiotherapy and there were 5 recurrences in 42 patients who did not receive radiotherapy. An expectant follow-up of 65 patients after pituitary surgery for nonfunctioning adenomas showed that 32% of tumors grew during a mean follow-up period of 76 months. In a retrospective comparison of 126 patients undergoing surgery alone or surgery with radiation, early postoperative radiotherapy (within 12 months) significantly reduced the risk of tumor regrowth by about 15% at 10 years. Despite the relatively high incidence of postoperative tumor regrowth, even after apparently complete resection, most neurosurgeons avoid routine postoperative radiation therapy. Radiation can be offered if the tumor mass reexpands.

This approach requires advising careful
follow-up with periodic annual MRI studies and visual evaluations. Because patients experience tumor regrowth even after
radiation therapy, all should undergo periodic post-treatment MRI, although less frequently.

**Expectant Observation.**

For nonfunctioning microadenomas or small macroadenomas (incidentalomas), patients may be observed expectantly. Some tumors do not grow over years or even decades. However, regular follow-up with MRI is necessary because these tumors may grow insidiously and are usually asymptomatic until they are large enough to affect vision. Periodic but less frequent endocrine evaluation is also suggested according to the clinical situation. Microadenomas rarely impair vision during pregnancy, whereas macroadenomas do so with greater frequency. Because macroadenomas do not respond to medical therapy, the risks of visual impairment arising during a pregnancy must be weighed carefully and resection prior to pregnancy may be indicated.

**Medications.**

Medications are not effective in reducing tumor size and visual compromise. Although dopamine agonists, GnRH antagonists, and somatostatin analogues modestly reduce tumor size in a few patients, they are not sufficiently effective to be recommended as therapy.
Growth Hormone

Somatotroph Cells

Mammosomatotroph cells expressing both PRL and GH arise from the acidophilic stem cell and immunostain mainly for PRL. Somatotrophs are located predominantly in the lateral wings of the anterior pituitary gland and constitute 35% to 45% of pituitary cells (Fig. 8-36) (Figure Not Available). These ovoid cells contain prominent secretory granules up to 700 µm in diameter. Juxtanuclear Golgi is particularly prominent with secretory granules

Figure 8-36 (Figure Not Available) Normal somatotroph. A somatotroph in the nontumorous pituitary is large, round to ovoid, and contains numerous electron-dense secretory granules whose diameter ranges from 250 to 700 µm. Short profiles of rough endoplasmic reticulum are scattered throughout the cytoplasm. The prominent juxtanuclear Golgi complex harbors forming secretory granules. (From Asa SL. In Tumors of the Pituitary Gland. Atlas of Tumor Pathology. Washington, DC, Armed Forces Institute of Pathology, 1997, p 14.)

in formation. The gland contains a total of 5 to 15 mg of GH.

Growth Hormone Biosynthesis

The human GH (hGH) genome locus spans approximately 66 kb and contains a cluster of five highly conserved genes located on the long arm of human chromosome 17q22-24. These are hGH-N, hCS-L, hCS-A, hGH-V, and hCS-B all of which consist of five exons separated by four introns. The hGH-V gene is selectively transcribed in pituitary somatotrophs and encodes a 22-kd (191-amino-acid) protein. The hCS-A and hCS-B genes are expressed in placental trophoblasts. Approximately 10% of pituitary GH is a 20-kd variant lacking amino acid residues 32 to 46. hGH-V, expressed in placental sycloctrophoblasts, encodes a 22-kd protein detected in the maternal circulation from midpregnancy and a minor form, IGH-V2. Elevated maternal IGH-V serum concentrations are accompanied by a decline in hGH-N, suggesting feedback regulation of the maternal hypothalamic-pituitary axis. After childbirth, circulating GH-V levels drop rapidly and are undetectable after 1 hour.

The hGH promoter region contains cis elements that mediate both pituitary-specific and hormone-specific signaling. The POUF1 transcription factor confers tissue-specific GH expression, and a second, ubiquitous factor binds to a distal Pit-1 site containing a consensus sequence for the Sp1 transcription factor. Pit-1 and Sp1 both contribute to GH promoter activation because mutation of the Sp1 binding site attenuates promoter activity. Deoxyribonucleosine hypersensitive sites of a locus control region of the hGH gene determine somatotroph and lactotroph GH expression, which involves regulation of a chromatin domain in these pituitary cells.

GH synthesis and release are under control of a variety of hormonal agents, including GHRH, somatostatin, ghrelin, IGF-I, thyroid hormone, and glucocorticoids. GH stimulates GH synthesis and release mediated by cAMP, CREB binding protein is phosphorylated by protein kinase A and is a cofactor for Pit-1-dependent human GH activation. IGF-I attenuates basal and stimulated GH gene expression.

The GH molecule, a single-chain polypeptide hormone consisting of 191 amino acids, is synthesized, stored, and secreted by somatotroph cells. The crystal structure of human GH reveals four alpha helices. Circulating GH molecules comprise several heterogeneous forms: 22- and 20-kd monomers, acetylated 22K, and two des-amino GH molecules. The 22-kd peptide is the major physiologic GH component, accounting for 75% of pituitary GH secretion. Amino acids 32 to 46 are deleted by alternative splicing of the gene to yield 20-kd GH, accounting for about 10% of pituitary GH. The 20-kd GH has slower metabolic clearance, accounting for the 20-kd/22-kd ratio being higher in plasma than in the pituitary gland. The 22-kd peptide retains growth-promoting activity but lacks diabetogenic effects, which are more pronounced with the 20-kd form.

GHRH and SRIF Interaction in Regulating Growth Hormone Secretion

The somatotroph cell expresses specific receptors for GHRH, GH secretagogues, and SRIF receptor subtypes 2 and 5 that mediate GH secretion. Hypothalamic SRIF and GHRH are secreted in independent waves and interact together with additional GH secretagogues to generate pulsatile GH release. GHRH selectively induces GH gene transcription and hormone release and does not induce other anterior pituitary or gut hormones. SRIF suppresses both basal and GHRH-stimulated GH pulse amplitude and frequency but does not affect GH biosynthesis. GH stimulation to normal adults elicits a prompt rise in serum GH levels, with higher levels occurring in female subjects. Although mature GH comprises 44 amino acids, GH-releasing activity resides in shorter proteolyzed forms involving amino acids 1 to 37 and 1 to 40 and the N-terminus. GH also is a determinant of somatotroph mitotic activity.

The rat hypothalamus releases GHRH and SRIF 180 degrees of out phase every 3 to 4 hours, resulting in pulsatile GH levels. SRIF antibody administration elevates GH levels, with intact intervening GH pulses, implying that hypothalamic SRIF secretion generates GH troughs. Similarly, GHRHR antibodies eliminate spontaneous GH surges. In humans, GH pulsatility persists when GH is tonically elevated as with ectopic tumor GHRH production or during GHRIH infusion, suggesting that hypothalamic SRIF is largely responsible for GH pulsatility. Preexposure to SRIF enhances somatotroph sensitivity to GHRH stimulation. Hence, during a normal GH trough period, the high SRIF level probably primes the somatotroph to respond maximally to a subsequent GHRH pulse, thus optimizing GH release. SRIF also inhibits central GHRIH release through direct synaptic connections with hypothalamic SRIF-containing neurons.

Chronic GHRH stimulation, by either continuous infusion or repeated bolus administration, eventually desensitizes GH release in vitro and in vivo, possibly through depletion of a GHRH-sensitive pool of GH. GHRH pretreatment also decreases somatotroph GHRIH binding sites. GH stimulation hypothalamic SRIF, GHRH and SRIF autoregulate their own secretion, and GHRIH also stimulates SRIF release. GH secretion is further regulated by its target growth factor, IGF-I, which participates in a hypothalamic-pituitary peripheral regulatory feedback system. GH stimulates IGF-I, which exerts a negative feedback effect on the hypothalamus and pituitary. IGF-I stimulates hypothalamic SRIF release and inhibits pituitary GH gene transcription and secretion.

Growth Hormone Secretagogues and Ghrelin

The isolation of ghrelin indicates a control system in addition to GHRH and SRIF in regulation of GH secretion (see Chapter 23). Ghrelin is a 28-amino-acid peptide that binds the GH secretagogue (GHS) receptor to induce hypothalamic GHRH and pituitary GH. A unique n-octanoylated serine 3 residue confers GH-releasing activity to the molecule. Ghrelin
is synthesized primarily in peripheral tissues, especially gastric mucosal neuroendocrine cells. Ghrelin administration dose dependently evokes GH release and also induces food intake and obesity development. It is thought that ghrelin controls GH secretion and peripheral sources may have additional nutritional effects requiring further elucidation. Current evidence suggests that the dual control of GH secretion postulated for GHRH and SRIF should be expanded to incorporate ghrelin.

Synthetic hexapeptides (artificial GHSs) recognize the GHS receptor, induce potent and reproducible GH release, and are useful for the diagnosis of GHD. GHSs stimulate GH secretion, and GHRH and GHSs act through distinct receptors and different intracellular signaling pathways on somatotroph subpopulations. GHSs require the presence of a functional hypothalamus to evoke GH, as evidenced in patients with an intact pituitary but disordered hypothalamic function, in whom GHS does not induce GH. GHSs potentiate GH release in response to a maximal stimulating dose of exogenous GHRH, and after a saturating dose of GHRH, when subsequent GHRH administration is ineffective, GHSs remain fully effective.

Functional GHS receptors are expressed in the human fetal pituitary by the fifth week of gestation. GHS-mediated GH release is demonstrable at birth, continues through infancy, increases at puberty, and decreases thereafter. Estradiol and testosterone increase GH-mediated GH release in childhood. Because GHS-evoked GH secretion is minimally altered by age, sex, or adiposity and is devoid of potential side effects (unlike insulin-induced hypoglycemia), GHSs may become a useful diagnostic tool in the diagnosis of adult GHD. Slight PRL and ACTH or cortisol increases have been reported with some GHSs, leading to the development of novel GHSs with more selective somatotroph actions.

Regulation of Growth Hormone Secretion

Multiple factors regulate the integrated secretion of GH. In some patients with acromegaly or chronic depression, CRH modestly increases GH. GnRH stimulates GH secretion in about one third of patients with GH responses to TRH are evoked in patients with liver disease, renal disease, ectopic GHRH-releasing carcinoid tumors, anorexia nervosa, and depression. Intravenous CRH modestly increases GH in some patients with acromegaly or chronic depression. GnRH stimulates GH secretion in about one third of patients with acromegaly.

Other Hormones Facilitating Growth Hormone Secretion

Acute glucocorticoid administration stimulates GH secretion, whereas chronic steroid treatment inhibits GH. Three hours after acute glucocorticoid administration, GH levels rose and remained elevated for 2 hours. However, supraphysiologic glucocorticoid exposure retarded growth, and Cushings disease was also associated with growth retardation, decreased serum GH, and decreased pituitary content surrounding the adenoma. Glucocorticoids administered to normal subjects do not inhibit GHRH-stimulated GH secretion in a manner similar to that seen in Cushing's syndrome. Furthermore, cortisol antagonizes peripheral GH action.

GH levels are decreased in hyperthyroid patients but become normal when patients are rendered euthyroid, suggesting that thyroid hormone suppresses GH secretion. Elevated circulating gonadal steroids observed during puberty may also account for higher pubertal GH levels. Estradiol stimulates GH secretory rates, and testosterone increases GH secretory mass per pulse with resultant IGF-I induction. Elevated circulating gonadal steroids may also account for higher pubertal GH levels. Estradiol stimulates GH secretory rates, and testosterone increases GH secretory mass per pulse with resultant IGF-I induction. Higher pubertal GH levels are associated with elevated GH pulse frequency and amplitude.

Neuropeptides, neurotransmitters, and opiate impinge on the hypothalamus and modulate GHRH and somatostatin (SRIF) release. Integrated effects of these complex neurogenic influences determine the final secretory pattern of GH. Apomorphine, a central dopamine receptor agonist, stimulates GH release. As does levodopa treatment. Oral levodopa administration evokes a brisk serum GH response within an hour in healthy young subjects. Norepinephrine increases GH secretion through -adrenergic pathways and inhibits GH release through -adrenergic pathways. Insulin-induced hypoglycemia, clonidine, arginine administration, exercise, levodopa, and arginine vasopressin facilitate GH release through -adrenergic effects. -Adrenergic blockade increases GHRH-induced GH release, possibly by a direct pituitary action or by decreasing hypothalamic somatostatin release. Endorphins and enkephalins stimulate GH and may account for GH release during severe physical stress and extreme exercise. Galanin, a 29-amino-acid neuropeptide, induces GH release and responses to GHRH. Cholinergic and serotoninergic neurons and several neuropeptides stimulate GH, including neuropeptide Y and somatostatin reverse starvation-induced GH release. In GH-deficient hypopituitary adults, elevated concentrations are higher than would be expected from their body fat mass.

Neuropeptides, neurotransmitters, and opiate impinge on the hypothalamus and modulate GHRH and somatostatin (SRIF) release. Integrated effects of these complex neurogenic influences determine the final secretory pattern of GH. Apomorphine, a central dopamine receptor agonist, stimulates GH release. GHRH is synthesized primarily in peripheral tissues, especially gastric mucosal neuroendocrine cells. Ghrelin administration dose dependently evokes GH release and also induces food intake and obesity development. It is thought that ghrelin controls GH secretion and peripheral sources may have additional nutritional effects requiring further elucidation. Current evidence suggests that the dual control of GH secretion postulated for GHRH and SRIF should be expanded to incorporate ghrelin.

Synthetic hexapeptides (artificial GHSs) recognize the GHS receptor, induce potent and reproducible GH release, and are useful for the diagnosis of GHD. GHSs stimulate GH secretion, and GHRH and GHSs act through distinct receptors and different intracellular signaling pathways on somatotroph subpopulations. GHSs require the presence of a functional hypothalamus to evoke GH, as evidenced in patients with an intact pituitary but disordered hypothalamic function, in whom GHS does not induce GH. GHSs potentiate GH release in response to a maximal stimulating dose of exogenous GHRH, and after a saturating dose of GHRH, when subsequent GHRH administration is ineffective, GHSs remain fully effective.

Functional GHS receptors are expressed in the human fetal pituitary by the fifth week of gestation. GHS-mediated GH release is demonstrable at birth, continues through infancy, increases at puberty, and decreases thereafter. Estradiol and testosterone increase GH-mediated GH release in childhood. Because GHS-evoked GH secretion is minimally altered by age, sex, or adiposity and is devoid of potential side effects (unlike insulin-induced hypoglycemia), GHSs may become a useful diagnostic tool in the diagnosis of adult GHD. Slight PRL and ACTH or cortisol increases have been reported with some GHSs, leading to the development of novel GHSs with more selective somatotroph actions.

Regulation of Growth Hormone Secretion

Multiple factors regulate the integrated secretion of GH. In some patients with acromegaly or chronic depression, CRH modestly increases GH. GnRH stimulates GH secretion in about one third of patients with GH responses to TRH are evoked in patients with liver disease, renal disease, ectopic GHRH-releasing carcinoid tumors, anorexia nervosa, and depression. Intravenous CRH modestly increases GH in some patients with acromegaly or chronic depression. GnRH stimulates GH secretion in about one third of patients with acromegaly.

Other Hormones Facilitating Growth Hormone Secretion

Acute glucocorticoid administration stimulates GH secretion, whereas chronic steroid treatment inhibits GH. Three hours after acute glucocorticoid administration, GH levels rose and remained elevated for 2 hours. However, supraphysiologic glucocorticoid exposure retarded growth, and Cushings disease was also associated with growth retardation, decreased serum GH, and decreased pituitary content surrounding the adenoma. Glucocorticoids administered to normal subjects do not inhibit GHRH-stimulated GH secretion in a manner similar to that seen in Cushing's syndrome. Furthermore, cortisol antagonizes peripheral GH action.

GH levels are decreased in hyperthyroid patients but become normal when patients are rendered euthyroid, suggesting that thyroid hormone suppresses GH secretion. Elevated circulating gonadal steroids observed during puberty may also account for higher pubertal GH levels. Estradiol stimulates GH secretory rates, and testosterone increases GH secretory mass per pulse with resultant IGF-I induction. GH responses to TRH are evoked in patients with liver disease, renal disease, ectopic GHRH-releasing carcinoid tumors, anorexia nervosa, and depression. Intravenous CRH modestly increases GH in some patients with acromegaly or chronic depression. GnRH stimulates GH secretion in about one third of patients with acromegaly.

Growth Hormone Binding Proteins

Circulating GH binding proteins (GHBPs) include a 20-kd low-affinity BP and a 60-kd high-affinity BP, which corresponds to the extracellular domain of the hepatic GH receptor and binds half of the circulating 22-kd GH form. The 20-kd GH binds preferentially to the low-affinity BP, which is unrelated to the GH receptor. The GHBPs function to damp acute oscillations in serum GH levels associated with pulsatile pituitary GH secretion, and the plasma half-life of GH is prolonged by...
decreased renal clearance of bound GH. The high-affinity BP also prevents GH binding to surface GH receptors by competing for the GH ligand. Patients with hypopituitarism or acromegaly have normal BP concentrations. GH resistance, as demonstrated in malnutrition, chronic liver disease, short stature, Laron dwarfism, and some African pygmies, is characterized by decreased BP levels in plasma. High BP levels are encountered in obese or pregnant subjects or those receiving estrogens or undergoing refeeding.

**Peripheral Growth Hormone Action**

GH acts to mediate growth and metabolic functions (Fig. 8-40). GH elicits intracellular signaling though a peripheral receptor and initiates a phosphorylation cascade involving the JAK/STAT pathway. The liver contains abundant GH receptors, and several peripheral tissues also express modest amounts of receptor, including muscle and fat (Fig. 8-41). The GH receptor is a 620-amino-acid, 70-kd protein of the class I cytokine-hematopoietin receptor superfamily consisting of an extracellular ligand-binding domain, a single membrane-spanning domain, and a cytoplasmic signaling domain. The GH receptor superfamily is homologous with receptors for PRL, interleukins 2 to 7, erythropoietin, interferon, and colony-stimulating factor.

GH complexes with two GH receptor components leading to receptor dimerization critical for subsequent GH signaling. Dimerization is followed by rapid JAK2 tyrosine kinase activation leading to phosphorylation of intracellular signaling molecules, including the signal-transducing activators of transcription proteins (STATs 1, 3, and 5), critical signaling components.

**IGF-I, a critical growth factor induced by GH, is probably responsible for most growth-promoting activities of GH** and also directly regulates GH receptor function. Paracrine IGF-I produced in extrathoracic tissues appears critical for growth because growth persists even when hepatic IGF-I is depleted in mice. GH receptor mutations are associated with partial or complete GH insensitivity and growth failure. These syndromes are associated with normal or high circulating levels, decreased circulating GHB levels, and low levels of circulating IGF-I. Multiple homozygous or heterozygous exonic and intronic GHR mutations have been described. Tissue responses to GH signaling are determined by the pattern of GH secretion in addition to the absolute amount of circulating hormone. Gender-specific patterns of GH secretion profiles determine sex-specific expression of cytokrome P450 enzymes. In turn, circulating steroids regulate neuroendocrine release of GH. SRIF, by suppressing interpulse GH levels, serves to masculinize the ultradian GH rhythm. In mice harboring a disrupted SRIF gene, plasma GH secretory patterns were elevated and liver enzyme induction lost its gender-specific dimorphism but the animals retained sexually dimorphic growth patterns.

Linear growth patterns and GH dose induction are phenotypically gender-specific because of higher GH pulse frequency rates and also show gender-specific STAT5b activity. Sexually dimorphic patterns of GH secretion and tissue targeting appear to be determined by STAT5b, which is sensitive to repeated pulses of injected GH, whereas other GH-induced responses are desensitized by repeated administration. Disruption of STAT5b in transgenic mice caused impaired male-pattern body growth associated with female-pattern IGF-I and testosterone levels. Appropriate GH pulsatility is also required to determine body growth mediated by STAT5b but not for metabolic effects of GH on carbohydrate metabolism.

Intracellular GH signaling is abrogated by suppression of cytokine signaling (SOCS) proteins, which disrupt the JAK/STAT pathway and thus disrupt GH action. In transgenic mice with deletion of SOCS-2, gigantism develops, presumably because of unrestrained GH action. Because SOCS proteins are also induced by proinflammatory cytokines, critically ill patients or those with renal failure may experience GH resistance related to cytokine-induced SOCS proteins. Unraveling STAT-SOCS regulation in syndromes associated with disordered GH signaling should yield mechanistic insights into dysregulated GH action. The impact of GH on growth is fully reviewed in Chapter 23. Extrapolitarily GH and GH secretagogues may complement the classical endocrine action between the GH-releasing factors, GH, and target tissues. Although GH immunoreactivity and mRNA expression have been documented in placenta, mammary gland, muscle, spleen, and lymphocytes, their physiologic role is not yet apparent.
**GH** is anabolic and causes urinary nitrogen retention, decreased plasma urea levels, and increased muscle mass. GH increases fat mobilization, decreases fat deposition, and activates hormone-sensitive lipase, resulting in increased triglyceride hydrolysis to free fatty acids and glycerol (lipolysis) and also decreased fatty acid reesterification. Replacement of GH in GH-deficient adults leads to decreased body fat and decreased adipocyte size and lipid content. As GH is degraded in the kidney, GH levels are elevated in patients with chronic renal failure and GH rises paradoxically in response to a glucose load in these patients.

**Growth Hormone Secretion Patterns**

Frequent serum sampling for measurements of GH concentrations has revealed a pattern of pulsatile secretion separated by troughs, during which GH is undetectable (<0.4 µg/L) (Table 8-19). Using currently available assays, random GH levels are undetectable in 50% of samples obtained from healthy subjects. GH secretion is high in the fetal circulation, peaking at about 150 µg/L during midgestation. Neonatal levels are lower (30 µg/L), possibly reflecting negative feedback control by rising levels of circulating IGF. GH levels during childhood (up to 7 µg/L) are characterized by enhanced pubertal GH pulse amplitude and mass with unchanged GH pulse frequency. GH pulse amplitudes decline inexorably with age, and GH levels in middle age are about 15% of pubertal levels. Healthy adult males produce GH at about 0.25 to 0.52 mg/m² per 24 hours. Obesity is associated with decreased GH pulse frequency and blunted evoked GH responses to secretagogues. In contrast, fasting is associated with enhanced GH pulse frequency and amplitude, possibly reflecting altered feedback regulation by nutritionally mediated changes in IGF binding protein concentrations and free IGF-I availability.

**Measurement of Spontaneous Growth Hormone Secretion**

Because pituitary GH secretion occurs episodically, accurate quantification of integrated GH secretion requires continuous measurement of secretion over 24 hours. This procedure requires insertion of a continuous withdrawal pump or patent indwelling catheter with unrestricted food intake and physical activity. Increasing sampling frequencies from every 20 minutes to 5-minute or 30-second sampling intervals enhances the threshold for detecting more pulses per hour. Although cumbersome and expensive, this method eliminates the error of isolated peak or trough measurements that might otherwise be obtained by single or multiple random GH samplings.

The discriminating power of continuous 24-hour GH measurement in the diagnosis of GHD in children has been disputed. With no clear diagnostic advantage over GH stimulation tests, integrated GH levels in young healthy subjects may overlap those of patients with organic hypopituitary disorders, which limits the utility of the measurement in the diagnosis of acquired adult GHD. However, fasted subjects may exhibit a clear distinction of deficient integrated GH levels measured over 8 hours.

**Urinary Growth Hormone Measurement**

Immunoassay methods for urinary GH measurement do not reliably reflect pharmacologic GH testing or adequately discriminate between normal and abnormal GH secretion. Clinical utility of urinary GH measurements requires rigorous agematched and gender-matched control subjects and standardized expression of GH concentrations relative to body weight or creatinine excretion.

**Variability of Growth Hormone Assays**

Plasma GH is measured by RIA (polyclonal or monoclonal) or by IRMA (dual monoclonal), but comparative GH measurements obtained using 11 commercial immunoassays varied by a factor of 3. Measured GH concentrations are antibody-dependent, and different antibodies bind to a heterogeneous spectrum of GH isoforms. Furthermore, GH isoform patterns vary between individuals and not all circulating GH forms are routinely detectable in GH assays adding further variation to comparison of results from different GH immunoassays. Monomeric 22-kd GH, the most abundant circulating form, is the only GH standard of sufficient purity and quantity and is used as the basis for GH measurement; however, it accounts for only about 25% of circulating immunoreactivity. Other GH forms are recognized to varying and largely unknown degrees. Polyclonal antibodies used in earlier RIAs recognized several molecular forms of GH as compared with newer immunometric assays employing highly specific monoclonal antibodies.

GH standards also affect comparison of GH values. In 1994, the first World Health Organization international standard for somatotropin, IRP 88/624, used recombinant technology, whereas previous standards were prepared from pituitary extracts. GHBPs may also interfere because approximately 50% of GH is bound by GHBP, spuriously high GH values may be reported. The heterogeneity of GH immunoassay results poses a challenge in the definition of accepted standards for diagnosis of GHD. However, the RIA is now infrequently used and clinicians should be aware of the nature of the GH assay employed and how values compare with those previously obtained by polyclonal RIA. New GH assays based on measuring GH bioactivity have been developed, including the eluted stain assay (ESTA) and the immunofunctional assay (IFA). The 22-kd exclusion assay (GHEA) also measures circulating GH isoforms.

**Growth Hormone Deficiency**

GHD in adults is recognized as a distinct adult syndrome (Table 8-20). GH is the most frequently deficient of the pituitary hormones in patients with pituitary disease, and GH has negative effects on body composition, cardiovascular risk factors, and quality of life. Life expectancy is reduced in hypopituitary patients with GHD and largely as a consequence of cardiovascular and cerebrovascular events, especially in female subjects. Although neither estrogen nor thyroid deficiency accounts for these risk factors and reduced survival, it has not yet been rigorously confirmed that the observed increased mortality and morbidity are solely a result of GHD.

**Pathophysiology**

Adult GHD is most frequently encountered in patients with pituitary or hypothalamic disease. Although pituitary tumors and craniopharyngiomas are associated with GHD, these patients may also experience GHD as a result of head or neck radiotherapy. GHD in children is typically isolated GHD or results from rare genetic causes (see Chapter 23) or other central structural abnormalities. Isolated GHD may be complete or partial, and up to 67% of children initially diagnosed with “idiopathic” GHD had normal GH responses when retested for GHD after cessation of GH treatment as adults. Children with GHD should be retested before GH treatment is continued into adulthood unless they have clearly documented panhypopituitarism or a defined genetic or developmental abnormality that causes complete and irreversible GHD. Mutations in the GH and GHRH receptor genes and GH insensitivity as a result of primary GH receptor dysfunction result in GHD. Other

---

**TABLE 8-19 -- Adult Growth Hormone Secretion.**

<table>
<thead>
<tr>
<th>Observation</th>
<th>Young Adult (µg/24 hr)</th>
<th>Fasting</th>
<th>Obesity</th>
<th>Middle-Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>24-hr secretion</td>
<td>540±44</td>
<td>2171±333</td>
<td>77±20</td>
<td>196±65</td>
</tr>
<tr>
<td>Secretary bursts</td>
<td>12±1</td>
<td>32±2</td>
<td>10±1</td>
<td></td>
</tr>
<tr>
<td>Growth hormone burst</td>
<td>45±4</td>
<td>64±9</td>
<td>24±5</td>
<td>10±6</td>
</tr>
</tbody>
</table>

*Deconvolution analysis of growth hormone (GH) secretion in adult males.*

---

Increased T
Increased maximum power
3
Decreased lipogenesis
Social isolation improved
Increased left ventricular wall mass
Increased IGF-I levels
Increased stroke volume
Improved physical mobility
Social isolation improved
Increased body fat mass with altered distribution
Increased lean body mass
Increased waist-hip ratio
Decreased fat mass
Decreased lean body mass
Increased bone mass
Reduced maximum O₂ uptake
Increased maximum O₂ uptake
Impaired cardiac function
Increased maximum power
Reduced muscle mass
Cardiovascular risk factors
Cardiovascular structure and function impaired
Increased stroke volume
Abnormal lipid profile
Increased diastolic volume
Decreased fibrinolytic activity
Increased left ventricular wall mass
Atherosclerosis
Atherosclerosis impact
Omental obesity
Insulin resistance
Sialic acid increased
Lower exercise frequency and duration
Imaging
Pituitary: mass or structural damage
Decreased adipocyte size
Bone: reduced density
Increased lipolysis
Abdomen: excess omental adiposity
Decreased lipogenesis
Laboratory
Evoked GH response (see Table 8-21)
Increased IGF-I levels
Increased BMR
IGF-I and IGF-BP3 low or normal
Decreased LDL with probable increased HDL
Lipid disorders
Transient hyperglycemia
Concomitant gonadotrophin, TSH, and/or ACTH reserve deficits
Increased T₃ levels
Salt and water retention
ACTH, adrenocorticotrophic hormone; BMR, basal metabolic rate; GH, growth hormone; HDL, high-density lipoprotein; IGF-I, insulin-like growth factor 1; IGF-BP3, IGF-binding protein 3; LDL low-density lipoprotein; T₃, triiodothyronine; T₄, thyroxine.

specific GH secretagogues. The insulin tolerance test (ITT) has remained the "gold standard" test for GHD. Although some GH secretagogues reliably differentiate normal GH reserve from GHD, others are not as effective in clearly distinguishing between normal and deficient because of overlapping GH values (Table 8-21).

Differences in published responses can also be accounted for by gender, weight, or possibly age. For example, the mean peak GH response to arginine plus GHRH, GHRH plus hexarelin, or GHRH plus GH-releasing peptide 6 was administered than when insulin-induced hypoglycemia was employed. Although a GH response of less than 3 µg/L has been considered consistent with GHD according to a current consensus, the recommended cutoff points below which adult patients should be considered to have GHD vary according to the test employed (Fig. 8-42). Not unexpectedly, patients with pituitary disease with no or one pituitary hormone deficiency have a lower incidence of GHD than patients with multiple hormone deficiencies. When 3 or more pituitary hormones are deficient, or if patients had pituitary disease with no or one pituitary hormone deficiency, a GH stimulation test may not be required to diagnose adult GHD.

Use of ghrelin, an endogenous ligand for the GH-releasing peptide receptor, may also be proposed as a test for GHD because ghrelin is a potent GH secretagogue. The requirement to test patients with proven panhypopituitarism has been questioned because virtually all patients with more than two trophic hormone deficiencies are also GH-deficient. Although mean serum IGF-I levels are low in adults with GHD and very low IGF-I levels may indicate GHD, IGF-I is not useful in a screening test in adults because about 60% of GH-deficient adults have normal IGF-I levels for their age and gender. Forty percent of patients older than 60 years with GHD had IGF-I concentrations that were normal for their age and gender. Measurement of IGF-BP3 is also not a reliable screening procedure for adult GHD.

Symptoms of GHD are nonspecific and may include fatigue, lack of energy, social isolation, poor concentration, and memory loss. Signs include increased fat mass (especially abdominal and visceral), decreased lean body mass, decreased body water, and decreased bone density, particularly in patients with more severe GHD and long-standing childhood-onset GHD. Other features include hyperlipidemia, reduced exercise capacity, and an increase in cardiovascular

---

**TABLE 8-20 -- Adult Somatotropin Deficiency**

<table>
<thead>
<tr>
<th>Clinical</th>
<th>Effects of Growth Hormone Replacement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impaired quality of life</td>
<td>Mood and energy uplift</td>
</tr>
<tr>
<td>Decreased energy and drive</td>
<td></td>
</tr>
<tr>
<td>Poor concentration</td>
<td>Enhanced vitality</td>
</tr>
<tr>
<td>Low self-esteem</td>
<td>Improved physical mobility</td>
</tr>
<tr>
<td>Social isolation</td>
<td>Social isolation improved</td>
</tr>
<tr>
<td>Body composition changes</td>
<td></td>
</tr>
<tr>
<td>Increased body fat mass with altered distribution</td>
<td>Increased lean body mass</td>
</tr>
<tr>
<td>Increased waist-hip ratio</td>
<td>Decreased fat mass</td>
</tr>
<tr>
<td>Decreased lean body mass</td>
<td>Increased bone mass</td>
</tr>
<tr>
<td>Cardiovascular risk factors</td>
<td></td>
</tr>
<tr>
<td>Cardiovascular structure and function impaired</td>
<td>Increased stroke volume</td>
</tr>
<tr>
<td>Abnormal lipid profile</td>
<td>Increased diastolic volume</td>
</tr>
<tr>
<td>Decreased fibrinolytic activity</td>
<td>Increased left ventricular wall mass</td>
</tr>
<tr>
<td>Atherosclerosis</td>
<td>Atherosclerosis impact</td>
</tr>
<tr>
<td>Omental obesity</td>
<td></td>
</tr>
<tr>
<td>Insulin resistance</td>
<td></td>
</tr>
<tr>
<td>Sialic acid increased</td>
<td></td>
</tr>
<tr>
<td>Lower exercise frequency and duration</td>
<td></td>
</tr>
</tbody>
</table>

---

As GH is secreted in a pulsatile manner, the diagnosis of GHD requires tests of evoked GH sufficiency in response to
risk factors including abdominal adiposity, insulin resistance, and increased carotid intimal thickness. GHD may also be associated with heart abnormalities including reduced left ventricular mass. Cardiovascular parameters of GHD are often more pronounced in adults who had childhood-onset GHD than in those who acquired the deficiency during adulthood, who have more pronounced disorders involving quality of life, lipids, and body composition.

Many symptoms of GHD are nonspecific, and clinical judgment should be exercised in selecting patients for testing. Patients with pituitary or hypothalamic disease should be considered for diagnosis of GHD and replacement with GH even in the absence of other pituitary hormone deficiencies. Deficiencies in sex hormones also lead to body composition changes similar to those observed in GHD. For example, testosterone deficiency causes decreased lean body mass and increased fat mass, each of which can be partially improved when testosterone is administered. That GH has additional positive effects on body composition is suggested by observations that many patients with GHD whose body composition parameters improved after GH replacement were already receiving sex steroids. A study of healthy volunteers also demonstrated inhibition of catabolic effects when hGH was given along with prednisone.

**Growth Hormone Replacement Therapy**

See Table 8-20. Treatment of GH-deficient adults with recombinant hGH for 4 to 6 months increased lean body mass and decreased fat mass (Fig. 8-43 and Fig. 8-44). In a review of nine placebo-controlled trials, hGH (in doses ranging from 2.6 to 26 µg/kg/day) increased lean body mass by a mean of 3.4 kg and reduced fat mass by 4.4%. GH also increased bone density and parameters of both bone formation and bone resorption. The GH-induced reduction in abdominal and visceral fat suggests an associated improvement in cardiovascular risk factors.

### TABLE 8-21 -- Responses to Growth Hormone Stimulation Tests.

<table>
<thead>
<tr>
<th>Test (Reference)</th>
<th>Controls, Growth Hormone Deficiency</th>
<th>Sex</th>
<th>Mean Age</th>
<th>Mean Body Mass Index (kg/m²)</th>
<th>Dose</th>
<th>Serum Sampling</th>
<th>Controls, Peak Mean (Range) (µg/L)</th>
<th>Growth Hormone Deficiency Peak Mean (RANGE) (µg/L)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITT</td>
<td>M:12</td>
<td>F:21</td>
<td>34.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arginine-GHRH</td>
<td>M:40</td>
<td>F:37</td>
<td>28.1</td>
<td></td>
<td>0.5 g/kg IV</td>
<td>q15 min</td>
<td>69.5</td>
<td>(13.8171)</td>
<td>(0.0257.5)</td>
</tr>
<tr>
<td>ITT</td>
<td>M:10</td>
<td>F:9</td>
<td>39.9</td>
<td>24.5</td>
<td>0.10.15 U/kg IV</td>
<td>-15 to 90</td>
<td>(384)</td>
<td>(0.111.8)</td>
<td></td>
</tr>
<tr>
<td>Arginine-GHRH</td>
<td>M:20</td>
<td>F:13</td>
<td>47.2</td>
<td>30.3</td>
<td>Insulin (R)</td>
<td>2030 min</td>
<td>17.8</td>
<td>(0.02562)</td>
<td>(0.0257.9)</td>
</tr>
<tr>
<td>ITT</td>
<td>M:26</td>
<td>F:14</td>
<td>48.9</td>
<td>30.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arginine-GHRH</td>
<td>M:20</td>
<td>F:14</td>
<td>47.2</td>
<td>30.3</td>
<td>30 g over</td>
<td>q30 min for</td>
<td>18.4</td>
<td>(1.2127)</td>
<td>(0.0257.7)</td>
</tr>
<tr>
<td>Hexarelin-GHRH</td>
<td>M:18</td>
<td>F:7</td>
<td>28.5</td>
<td></td>
<td>0.25 µg/kg IV</td>
<td>q15 min</td>
<td>83.6</td>
<td>(49124)</td>
<td>(0.111.1)</td>
</tr>
<tr>
<td>GHRH-GHRP-6</td>
<td>M:65</td>
<td>F:60</td>
<td>39.9</td>
<td>23.5</td>
<td>1 µg/kg IV</td>
<td>q1530</td>
<td>59.2</td>
<td>(0.0115)</td>
<td></td>
</tr>
<tr>
<td>GHRH-GHRP-6</td>
<td>M:73</td>
<td>F:52</td>
<td>39.9</td>
<td>28.9</td>
<td>1 µg/kg IV</td>
<td>-30 to 120</td>
<td>4.1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Recommended test sensitivity by 95% confidence limits* or recommended by investigator** to diagnose adult GH deficiency

<table>
<thead>
<tr>
<th>Test</th>
<th>Value (µg/L)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITT</td>
<td>5.1*</td>
<td>Biller, 2002</td>
</tr>
<tr>
<td>Arginine + GHRH</td>
<td>4.1*</td>
<td>Biller, 2002</td>
</tr>
<tr>
<td>Arginine + GHRH</td>
<td>9.0**</td>
<td>Gasperi, 1999</td>
</tr>
<tr>
<td>Arginine + -Dopa</td>
<td>1.7*</td>
<td>Biller, 2002</td>
</tr>
<tr>
<td>Hexarelin + GHRH</td>
<td>9.0**</td>
<td>Gasperi, 1999</td>
</tr>
<tr>
<td>GHRH + GHRP-6</td>
<td>10.0**</td>
<td>Popovic, 2000</td>
</tr>
</tbody>
</table>

* Results from two studies on the effects of insulin-induced hypoglycemia and the combination of arginine and GHRH illustrate the effect of body weight on GH responses.

The effect of hGH replacement on lipid abnormalities is variable. The hGH had a significant effect on raising high-density lipoprotein cholesterol and overall the most consistent change was an improvement in the ratio of cholesterol to high-density lipoprotein cholesterol. Some have reported increases in cardiac output, reduction in intima media thickness, and improved energy, mood, and quality of life, but not all investigators concur. Improved quality of life, including a significant reduction in sick days, hospitalization length, and physician visits, has been reported in some studies. There may be a latency period up to 3 months before patients recognize the benefits of hGH replacement, which are most obvious in those with the most profound symptoms and signs of GHD. Beneficial effects...
of GH replacement persist for at least 10 years. 

GH Administration.

GH is administered by nightly subcutaneous injection, the recommended dose for adults is much lower than for children, and children experience fewer side effects of GH than adults. Men, particularly older ones, are more sensitive to GH and require lower GH doses than women. The maintenance dose of hGH in 665 adult patients with GHD was 0.43 mg/day for men and 0.53 mg/day for women. Women with GHD require higher doses of hGH with oral than with transdermal estrogen (Fig. 8-45). It is recommended that replacement be initiated with relatively low doses. Most adults tolerate a starting dose of 300 µg/day or lower, which is then titrated according to serum IGF-I concentrations and side effects of the medication. If side effects occur, the dose should be reduced. If no side effects are reported, the therapeutic goal is to maintain IGF-I levels in the normal range for age and gender while avoiding levels in the upper quintile.

Precautions and Caveats of Treating with Human Growth Hormone

The most common side effects of hGH are edema, arthralgias, and myalgias which occurred in up to one third of patients when the drug was administered at a weight-based dosage. Using total daily doses titrated for IGF-I levels, the incidence of side effects is much lower. Patients with active malignancies should not be treated with GH, nor should patients with active carpal tunnel syndrome or other fluid retention disorders. The possibility that hGH might initiate new cancers or stimulate growth of preexisting benign tumors is an important theoretical issue. An epidemiologic association between higher, albeit normal, IGF-I levels and later risk of development of prostate cancer, breast cancer in premenopausal women, and colon and lung cancer has been reported. In contrast, patients with acromegaly, who have very high serum levels of IGF-I, do not have an increased incidence of either breast or prostate cancer or cancer in general. In fact, the overall risk for cancer in acromegaly is lower than expected. However, these patients have significantly increased mortality from colon cancer.

The possibility that hGH treatment might cause new or recurrent malignancies has been best examined in children in whom GHD developed as a result of treatment. When the relative risk of brain tumor recurrence in 180 children treated with hGH versus 891 who did not receive hGH was analyzed, the risk of recurrence after a mean of 6.4 years was lower in the treated group than those not receiving hGH. Nevertheless, long-term surveillance with adequate control groups and avoidance of high IGF-I levels in adults being treated for GHD are required to ensure that GH replacement in adults does not increase the incidence of new cancers or growth of existing benign tumors. Blood glucose levels should also be monitored carefully, especially in patients also being treated for diabetes.

Growth Hormone Treatment of Catabolic States

The well-recognized anabolic actions of GH have prompted use of GH in catabolic states including those associated with surgery, trauma, burns, parenteral nutrition, and organ failure. These potential indications for GH are not approved in the United States. The negative nitrogen balance in critically ill patients is partly attributable to GH resistance as well as to decreased IGF-I production and action. GH administered to postsurgical patients as well as to normal subjects receiving hypocaloric intravenous alimentation results in reversion to positive nitrogen balance.

Beneficial effects of GH have been reported in patients with extensive burns, chronic high-dose glucocorticoid treatment, chronic obstructive pulmonary disease, cancer, and cardiac failure. Nevertheless, published end points for these studies have not been definitive. When GH was administered to elderly malnourished patients, it was found to be an effective adjuvant for dietary augmentation. A study in which critically ill patients...
Figure 8-46 Management of adult somatotropin deficiency. Patients older than 60 years require lower maintenance doses. Women receiving transdermal estrogen require lower doses than those receiving oral estrogen preparations. GH, growth hormone; IGF-I, insulin-like growth factor I; Rx, treatment.

<table>
<thead>
<tr>
<th>TABLE 8-22</th>
<th>Side Effects of Growth Hormone Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edema</td>
<td></td>
</tr>
<tr>
<td>Arthralgias</td>
<td></td>
</tr>
<tr>
<td>Myalgias</td>
<td></td>
</tr>
<tr>
<td>Muscle stiffness</td>
<td></td>
</tr>
<tr>
<td>Paresthesias</td>
<td></td>
</tr>
<tr>
<td>Carpal tunnel syndrome</td>
<td></td>
</tr>
<tr>
<td>Hypertension</td>
<td></td>
</tr>
<tr>
<td>Atrial fibrillation</td>
<td></td>
</tr>
<tr>
<td>Headache</td>
<td></td>
</tr>
<tr>
<td>Tinnitus</td>
<td></td>
</tr>
<tr>
<td>Benign intracranial hypertension</td>
<td></td>
</tr>
<tr>
<td>Increase in melanocytic nevi</td>
<td></td>
</tr>
</tbody>
</table>

received very high doses of GH (up to 7 mg/day) was prematurely terminated because of unexplained increased mortality. It was suggested that GH may have had an adverse effect on acute phase protein synthesis in these patients.

Growth Hormone Treatment for Osteoporosis

As declining GH secretion has been implicated in the pathogenesis of osteoporosis, GH was administered to otherwise healthy subjects with idiopathic osteoporosis in an attempt to decrease bone loss. GH increased indices of bone formation and resorption, but a modest increase in spine bone mineral density was observed only in male subjects. A longer study in osteoporotic female subjects showed that GH together with calcitonin increased spine and total hip bone mineral density after 2 years of treatment, although the response was less marked than that observed with estrogen or bisphosphonate therapy. Limited proven efficacy, unclear side effects, and lack of comparative studies with other beneficial therapies for osteoporosis indicate a need for further study of the potential use of GH in treating osteoporosis.

Growth Hormone Treatment in Human Immunodeficiency Virus Infection

GH is approved by the Food and Drug Administration for administration to adult patients with human immunodeficiency virus (HIV) associated cachexia. The GH treatment resulted in positive nitrogen balance, increased lean body mass, decreased body fat, and improved work output. Ten HIV-infected subjects with fat redistribution syndrome associated with protease inhibitor therapy received GH at 6 mg/day subcutaneously for 12 weeks and showed decreased weight/hip ratios and enhanced mid thigh circumference. However, long-term beneficial effects of GH on survival and quality of life in HIV infection have not yet been reported.

Growth Hormone Use in Competitive Sports

The public policy issues of GH abuse in competitive sports have received much attention. GH has been used by athletes to enhance muscle mass. Whether persistent GH use is accompanied by increased muscle strength is unclear. Continued use of pharmacologic GH doses by athletes could result in adverse effects of acromegaly, which would decrease performance.

Decreased Insulin-like Growth Factor I Levels

Short-term fasted normal healthy subjects have moderately elevated basal GH levels, and protein-calorie malnutrition, starvation, and anorexia nervosa are associated with low IGF-I and markedly elevated GH levels. This observation may reflect uncoupling of IGF-I feedback regulation of GH secretion. Low IGF-I levels in normal fasting subjects are normalized by caloric refeeding to a greater degree than by protein intake alone.
Acromegaly

In 1866 Pierre Marie published the first clinical description of disordered somatic growth and proportion and proposed the name acromegaly. He also recognized cases previously described by others. When the relation of this syndrome to a pituitary tumor was later recognized, Benda showed in 1900 that these tumors comprise mainly adenohypophyseal eosinophilic cells, which he proposed to be hyperfunctioning. Cushing, Davidoff, and Bailey documented the clinicopathologic features of acromegaly and demonstrated clinical remission of soft tissue signs after adenoma resection. Evans and Long induced gigantism in rats injected with anterior pituitary extracts, confirming the association of a pituitary factor with somatic growth. Establishment of the unequivocal pathophysiologic link between hyperfunctioning adenoma and acromegaly represented the earliest example of a pituitary disorder being clinically and pathologically recognized and appropriately managed by surgical excision of a hypersecreting source.

Incidence

The prevalence of acromegaly is estimated to range from 38 to 69 cases per million, and the annual incidence of new patients is 3 to 4 cases per million. On the basis of these largely Western European studies, it is estimated that over 1000 new cases of acromegaly are diagnosed annually in the United States.

Pathogenesis

GH and IGF-I act both independently and dependently in inducing features of hypersonsomatotropism. Acromegaly is caused by pituitary tumors secreting GH or rarely by extrapituitary disorders (Fig. 8-47) (Figure Not Available). Regardless of the etiology, the disease is characterized by elevated levels of GH and IGF-I with resultant signs and symptoms of hypersonsomatotropism.

Pituitary Acromegaly

More than 95% of patients with acromegaly harbor a GH-secreting pituitary adenoma (Table 8-23). Pure GH cell adenomas contain either densely or sparsely staining cytoplasmic GH granules, and these two variants grow slowly (densely granulated) or rapidly (sparsely granulated). The former arise insidiously and occur during or after middle age, whereas the latter arise in younger subjects with more florid disease. Mixed GH cell and PRL cell adenomas are composed of distinct somatotrophs expressing GH and lactotrophs expressing PRL. Monomorphic acidophilic stem cell adenomas arise from the common GH and PRL stem cell and also often contain giant mitochondria and misplaced GH granule exocytosis. They grow rapidly, are invasive, and arise with predominant features of hyperprolactinemia. Monomorphic mammomatosomatotroph cell adenomas express both GH and PRL from a single cell, and plurihormonal tumors may express GH with any combination of PRL, TSH, ACTH, or subunit. These patients present with clinical features of acromegaly as well as hyperprolactinemia, Cushing's disease, or rarely hyperthyroxinemia.

Somatotroph hyperplasia is difficult to distinguish from a GH cell adenoma, and silver staining displays a well-preserved reticulin network without a surrounding pseudocapsule. The rigorous morphologic diagnosis of GH cell hyperplasia is usually associated with stimulation by ectopic GHRH derived from an extrapituitary tumor causing acromegaly. Silent somatotroph adenomas immunostain positively for GH and are apparently clinically nonfunctional, although GH or PRL levels, or both, may be modestly elevated in over 50% of these patients.

Pathogenesis of Somatotroph Cell Adenomas

Both pituitary and hypothalamic factors influence pituitary tumor pathogenesis. Even when exhibiting marked nuclear pleomorphism, mitotic activity, and invasiveness, these tumors are usually benign.

Disordered GHRH Secretion or Action.

Adenomas express receptors for GHRH, ghrelin, and SRIF, but activating mutations of the GHRH or SRIF receptor have not been reported. GHRH directly stimulates GH gene expression and also induces somatotroph mitotic activity. Transgenic GHRH expression causes somatotroph hyperplasia and ultimately adenoma. Clinically, GHRH production by hypothalamic, abdominal or chest neuroendocrine tumors causes somatotroph hyperplasia and occasionally adenoma with resultant unrestrained GH secretion and acromegaly. However, histologic examination of most pituitary GH cell adenoma tissue specimens does not show hyperplastic somatotroph tissue surrounding the adenoma, implying no generalized hypothalamic overstimulation. Failure to down-regulate GH secretion during prolonged GHRH stimulation also points to a role for GHRH in maintaining persistent GH hypersecretion. Expression of intra-adenomatous GHRH correlates with tumor size and activity, implying a paracrine role for GHRH in mediating adenoma pathogenesis. GH modestly stimulates PRL secretion, and up to 40% of patients with acromegaly also have hyperprolactinemia.

Complete surgical resection of well-defined GH-secreting microadenomas usually results in a definitive cure of excess hormone secretion with low postoperative tumor recurrence rates, strongly suggestive of intact hypothalamic function in these patients. Although basal GH levels are usually high in acromegaly, the episodic pulsatile pattern of GH release is intact and the nocturnal GH surge usually preserved. Patients treated with SRIF analogues also retain GH pulsatility, and GH pulse amplitude and sensitivity to GHRH appear intact.

Disordered Somatotroph Cell Function.

A somatotroph mutation may be a prerequisite for the abnormal growth response to disordered GH secretion or action. The monoclonal origin of somatotroph adenomas was determined by X-chromosome inactivation analysis of somatotroph tumor DNA. An altered Gs protein identified in a subset of GH-secreting pituitary adenomas led to high levels of intracellular cAMP and GH hypersecretion. Point mutations in two critical sites, Arg201, the site for adenosine diphosphate ribosylation, and Gly227, the guanosine triphosphate binding domain of Gs proteins, prevent guanosine triphosphatase activity and result in constitutive adenyl cyclase activation. This dominant gain mutant mimics GHRH effects, results in elevated cAMP levels, and is present in about 30% of GH-secreting tumors. Loss of heterozygosity has been observed for chromosomes 11, 13, and 9, especially in larger, more invasive macroadenomas. However, no defined tumor suppressor gene has been isolated for these sporadic nonfamilial tumors. An activating pituitary tumor transforming gene (PTTG) isolated from pituitary tumors is overexpressed in GH-secreting tumors, and its abundance correlates with tumor size and invasiveness. PTTG participates as a securin protein, regulating sister chromatid separation during the cell cycle, and its overexpression may lead to cell aneuploidy.

The sequence of events leading to somatotroph clonal expansion appears multifactorial. An activated oncogene may be required for initiating tumorigenesis, and promotion of tumor growth may require GHRH and other growth factor stimulation. The cellular mutation may not by itself be sufficient to provide a growth advantage for a GH-secreting adenoma without additional disordered hypothalamic or paracrine growth factor signaling.

Extrapituitary Acromegaly

Microadenomas usually result in a definitive cure of excess hormone secretion with low postoperative tumor recurrence rates, strongly suggestive of intact hypothalamic function in these patients. Although basal GH levels are usually high in acromegaly, the episodic pulsatile pattern of GH release is intact and the nocturnal GH surge usually preserved. Patients treated with SRIF analogues also retain GH pulsatility, and GH pulse amplitude and sensitivity to GHRH appear intact.
Excess GH secretion in acromegaly may not necessarily be pituitary in origin. Because management of ectopic acromegaly differs from that of pituitary GH hypersecretion, rigorous clinical and biochemical criteria should be fulfilled to confirm the diagnosis of ectopic acromegaly. These criteria include demonstration of elevated circulating GHRH or GH levels in the absence of a primary pituitary lesion, a significant arteriovenous hormone gradient across the ectopic tumor source, biochemical and clinical cure of acromegaly after resection of the ectopic hormone-producing tumor, and normalization of the GHRH-GH-IGF-I axis. Finally, GHRH or GH gene product expression should be shown. Patients with inconclusive imaging, biochemical, or clinical features of pituitary acromegaly may inadvertently be diagnosed as harboring a nonpituitary source of excess GH secretion and be inappropriately treated.

**GHRH Hypersecretion.**

Hypothalamic tumors, including hamartomas, choristomas, gliomas, and gangliocytomas, may produce GHRH with subsequent somatotroph hyperplasia or even a pituitary GH cell adenoma and resultant acromegaly. Primary mammosomatotroph hyperplasia with no evidence for pituitary adenoma or an extrapituitary tumor source of GHRH has been described in gigantism. The structure of hypothalamic GHRH was, in fact, elucidated from material extracted from pancreatic GHRH-secreting tumors in patients with acromegaly.

### TABLE 8-23: Causes of Acromegaly

<table>
<thead>
<tr>
<th>Cause</th>
<th>Prevalence (%)</th>
<th>Hormonal Product/s</th>
<th>Clinical Features</th>
<th>Pathologic Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excess Growth Hormone</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secretion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pituitary</td>
<td>98</td>
<td>GH</td>
<td>Slow growing</td>
<td>Resemble normal somatotrophs, numerous, large secretory granules</td>
</tr>
<tr>
<td>Densely granulated GH cell adenoma</td>
<td>30</td>
<td></td>
<td>Clinically insidious</td>
<td>Cellular pleomorphism</td>
</tr>
<tr>
<td>Sparsely granulated adenoma</td>
<td>30</td>
<td>GH</td>
<td>Rapidly growing</td>
<td>Characteristic ultrastructure</td>
</tr>
<tr>
<td>Mixed GH cell and PRL cell adenoma</td>
<td>25</td>
<td>GH and PRL</td>
<td>Variable</td>
<td>Densely granulated somatotrophs sparsely granulated lactotrophs</td>
</tr>
<tr>
<td>Mammosomatotroph cell adenoma</td>
<td>10</td>
<td>GH and PRL</td>
<td>Common in children. Gigantism, mild hyperprolactinemia</td>
<td>Both GH and PRL in same cell, often same secretory granule</td>
</tr>
<tr>
<td>Acidophil stem cell adenoma</td>
<td></td>
<td>PRL and GH</td>
<td>Rapidly growing invasive, hyperprolactinemia dominant</td>
<td>Distinctive ultrastructure</td>
</tr>
<tr>
<td>Plurihormonal adenoma</td>
<td></td>
<td>GH (PRL w/GSU, FSH/LH, TSH, or ACTH</td>
<td>Often secondary hormonal products are clinically silent</td>
<td>Variable epithelial monomorphous or plurimorphous</td>
</tr>
<tr>
<td>GH cell carcinoma or metastases</td>
<td></td>
<td>GH</td>
<td>Usually aggressive</td>
<td>Documented metastasis</td>
</tr>
<tr>
<td>Multiple endocrine neoplasia-type 1 (adenoma)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McCune-Albright syndrome (rarely adenoma)</td>
<td></td>
<td>GH, PRL</td>
<td>Classic triad</td>
<td></td>
</tr>
<tr>
<td>Ectopic sphenoid or parapharyngeal sinus pituitary adenoma</td>
<td></td>
<td>GH</td>
<td>Ectopic mass</td>
<td>Adenoma</td>
</tr>
<tr>
<td>Familial acromegaly (adenoma)</td>
<td></td>
<td>GH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carney's syndrome (adenoma)</td>
<td></td>
<td>GH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extrapituitary tumor</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pancreatic islet cell tumor</td>
<td>&lt;1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Excess Growth Hormone</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Releasing Hormone</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secretion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central</td>
<td>&lt;1</td>
<td>Hypothalamic mass</td>
<td>Somatotroph hyperplasia</td>
<td></td>
</tr>
<tr>
<td>Hypothalamic hamartoma, choristoma, ganglieneuroma</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peripheral</td>
<td>1</td>
<td>GH, PRL</td>
<td>Systemic features</td>
<td>Somatotroph hyperplasia</td>
</tr>
<tr>
<td>Bronchial carcinoid, pancreatic islet cell tumor, small cell lung cancer, adrenal adenoma, medullary thyroid carcinoma, pheochromocytoma</td>
<td></td>
<td></td>
<td>Rarely adenoma</td>
<td></td>
</tr>
</tbody>
</table>


GHRH immunoreactivity is detectable in about 25% of carcinoid tumor samples. Acromegaly in these patients is uncommon, however. In a retrospective survey of 177 patients with acromegaly, only a single patient was identified with elevated plasma GHRH levels. Bronchial carcinoids make up most tumors associated with ectopic GHRH secretion. Pancreatic cell tumors, small cell lung cancers, adrenal adenoma, pheochromocytoma, and medullary thyroid, endometrioid, and breast cancers have rarely been described to express GHRH and cause acromegaly. Surgical resection of the tumor secreting ectopic GHRH should reverse the GH hypersecretion, and pituitary surgery is not required in these patients. Carcinoid syndrome with ectopic GHRH secretion can also be managed with somatostatin analogues, which lower GH and IGF-I levels and suppress ectopic tumor elaboration of GHRH.

**Ectopic Pituitary Adenomas.**

GH-secreting adenomas may arise from ectopic pituitary remnants in the sphenoid sinus, petrous temporal bone, or nasopharyngeal cavity. Rarely, pituitary carcinoma may spread to the meninges, CSF, or cervical lymph nodes, resulting in functional GH-secreting metastases that may be diagnosed by radiolabeled octreotide imaging (indium In 111 pentetreotide [OctreoScan]).
Peripheral GH-Secreting Tumors.

Lung adenocarcinoma, breast cancer, and ovarian tissues contain immunoreactive GH without clinical evidence of acromegaly. Rarely, a GH-secreting intramesenteric pancreatic islet cell tumor causes acromegaly; these patients present with an intra-abdominal mass, a normal-sized or small pituitary gland on MRI, no GH response to TRH injection, and normal levels of circulating plasma GHRH.

Acromegolism.

Rarely, conditions exhibiting soft tissue and skin changes usually associated with acromegaly and normal baseline and dynamic GH and IGF-I with no demonstrable pituitary or extrapituitary tumor have been termed acromegolism. Pachydermoperiostosis should be considered in the differential diagnosis. Insulin resistance and defective IGF-I binding have been demonstrated in cells derived from some patients with acanthosis nigricans, and treatment is symptomatic.

McCune-Albright Syndrome.

This rare hypersecretory syndrome consists of polyostotic fibrous dysplasia, cutaneous pigmentation, sexual precocity, hyperthyroidism, hypercortisolism, hyperprolactinemia, and acromegaly. Although few patients have definitive evidence of a pituitary adenoma, Gs mutations have been detected in both endocrine and nonendocrine tissues. GH hypersecretion can be controlled by somatostatin analogues or pituitary irradiation.

Multiple Endocrine Neoplasia.

GH cell pituitary adenoma is a well-documented component of the autosomal dominant MEN-1 syndrome, which also includes parathyroid and pancreatic tumors (see Chapter 36). MEN-1, associated with germ cell inactivation of the MENIN tumor suppressor gene located on chromosome 11q13, appears intact in sporadic GH cell adenomas. Rarely, functional pancreatic tumors in patients with MEN-1 also express GHRH.

Familial Acromegaly.

Familial acromegaly may occur in association with the Carney complex, which maps to chromosome 2p. Several families with isolated familial acromegaly comprise related cases of acromegaly and gigantism and harbor loss of heterozygosity in chromosome 11q13, distinct from MENIN (see Table 8-4).

Clinical Features

Manifestations of acromegaly are caused by either central pressure effects of the pituitary mass or peripheral actions of excess GH and IGF-I. Central features of the expanding pituitary mass are common to all pituitary masses and have already been described. In acromegaly, headache is often severe and debilitating. Local signs are especially important presenting features because a preponderance of macroadenomas (>65%) is encountered in acromegaly, compared with mostly microadenomas for PRL-secreting tumors.

Gigantism

Tall stature may be caused by a GH-secreting pituitary tumor or hyperplasia. About 20% of patients have the McCune-Albright syndrome with somatotroph hyperplasia or rarely pituitary adenomas. Somatotroph hyperplasia and acidophilic stem cell adenomas may cause gigantism in infancy or early childhood, suggesting early hypersecretion of GHRH or disordered pituicyte cell differentiation. Pituitary gigantism should be considered in children who are more than 3 standard deviations (SD) above normal mean height for age or more than 2 SD above their adjusted mean parental height.

The biochemical diagnosis is similar to that for acromegaly; that is, GH levels are in excess of 1 µg/L after a glucose load and serum IGF-I concentrations are elevated. In children undergoing pubertal growth spurts, GH responses to glucose may vary. A summary of the clinical presentation of acromegaly is shown in Table 8-24.

TABLE 8-24 -- Presentation of Acromegaly

<table>
<thead>
<tr>
<th>Presenting Chief Complaint</th>
<th>Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Menstrual disturbance</td>
<td>13</td>
</tr>
<tr>
<td>Change in appearance, acral growth</td>
<td>11</td>
</tr>
<tr>
<td>Headaches</td>
<td>8</td>
</tr>
<tr>
<td>Paresthesias, carpal tunnel syndrome</td>
<td>6</td>
</tr>
<tr>
<td>Diabetes mellitus, impaired glucose tolerance</td>
<td>5</td>
</tr>
<tr>
<td>Heart disease</td>
<td>3</td>
</tr>
<tr>
<td>Visual impairment</td>
<td>3</td>
</tr>
<tr>
<td>Decreased libido, impotence</td>
<td>3</td>
</tr>
<tr>
<td>Arthopathy</td>
<td>3</td>
</tr>
<tr>
<td>Thyroid disorder</td>
<td>2</td>
</tr>
<tr>
<td>Hypertension</td>
<td>1</td>
</tr>
<tr>
<td>Gigantism</td>
<td>1</td>
</tr>
<tr>
<td>Fatigue</td>
<td>0.3</td>
</tr>
<tr>
<td>Hyperhidrosis</td>
<td>0.3</td>
</tr>
<tr>
<td>Somnolence</td>
<td>0.3</td>
</tr>
<tr>
<td>Other</td>
<td>5</td>
</tr>
<tr>
<td>Chance (detected by unrelated physical or dental examination or radiograph)</td>
<td>40</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
</tr>
</tbody>
</table>

Causes of Death

Cardiovascular                              60
Respiratory                                  25
Malignancy                                   15


be paradoxical and serum IGFl concentrations are often physiologically elevated. Thus, the diagnosis requires clear-cut MRI evidence for a pituitary lesion. The differential diagnosis includes familial tall stature, redundancy of Y chromosomes, Marfan's syndrome, and homocystinuria.

**Clinical Features of Acromegaly**

Effects of hypersomatotropism on acral and soft tissue growth and metabolic function occur insidiously over several years ([Table 8-24; Fig. 8-48 and Fig. 8-49]). The slow onset and elusive symptoms often result in a delay in diagnosis ranging from 6.6 to 10.2 years, with a mean delay of almost 9 years. Patients may seek care for dental, orthopedic, rheumatologic, or cardiac disorders. Only 13% of 256 patients diagnosed during a 20-year period presented with primary symptoms of altered facial appearance or enlarged extremities.

In a review of several hundred patients presenting with acromegaly worldwide, 98% had acral enlargement and hyperhidrosis was prominent in 70%.

When patients present early, facial and peripheral features are usually not obvious and a serial review of old photographs often reveals the progress of subtle physical changes. Characteristic features include large fleshy lips and nose, spade-like hands, frontal skull bossing, and cranial ridges. Enlarged tongue, bones, salivary glands, thyroid, heart, liver, and spleen are the effects of generalized visceromegaly. Clinically apparent hepatosplenomegaly, however, is rare. Increases in shoe, ring, or hat size are commonly reported. Progressive acral changes may lead to facial and skeletal disfigurement, especially if excess GH secretion begins prior to epiphyseal closure. These include mandibular overgrowth with prognathism, maxillary widening, teeth separation, jaw malocclusion and overbite, and nasal bone hypertrophy. Sonorous voice deepening occurs in association with laryngeal hypertrophy and enlarged panarosal sinuses.

Up to half of the patients may experience joint symptoms severe enough to limit daily activities. Arthropathy occurs in about 70% of patients, most of whom exhibit joint swelling, hypermobility, and cartilaginous thickening. Local periarticular fibrous tissue thickening may cause joint stiffening or deformities and nerve entrapment. Knees, hips, shoulders, lumbosacral joints, elbows, and ankles are affected with monoartricular or polyartricular arthritides, but joint effusions rarely develop. Spinal involvement includes osteophytes, disc space widening, and increased anteroposterior vertebral length, which may result in dorsal kyphosis.

Neural enlargement and wrist tissue swelling may lead to carpal tunnel syndrome in up to half of all patients. Chondrocyte proliferation with an increased joint space occurs early, and ulcerations and fissures of weight-bearing cartilage areas are often accompanied by new bone formation. Debulking osteoarthritides may result in bone remodeling, osteophyte formation, subchondral cysts, narrowed joint spaces, and lax periarticular ligaments. Osteophytes commonly occur at the phalangeal tufts and over the anterior aspects of spinal vertebrae. Ligaments may ossify, and periarticular calcium pyrophosphate deposition occurs. Although the duration of hypersomatotropism correlates with clinical severity of the joint changes, it is unclear whether higher GH levels correlate with increased articular disease activity. Therapeutic responses usually depend on the degree of irreversible bone changes already in place.

Hyperhidrosis and malodorous oily skin are common early signs, occurring in up to 70% of patients. Facial wrinkles, nasolabial folds, and heel pads thicken and body hair may become coarse. Hyperhidrosis and malodorous oily skin are common early signs, occurring in up to 70% of patients. Facial wrinkles, nasolabial folds, and heel pads thicken and body hair may become coarse. Hyperhidrosis and malodorous oily skin are common early signs, occurring in up to 70% of patients. Facial wrinkles, nasolabial folds, and heel pads thicken and body hair may become coarse. Hyperhidrosis and malodorous oily skin are common early signs, occurring in up to 70% of patients. Facial wrinkles, nasolabial folds, and heel pads thicken and body hair may become coarse. Hyperhidrosis and malodorous oily skin are common early signs, occurring in up to 70% of patients. Facial wrinkles, nasolabial folds, and heel pads thicken and body hair may become coarse. Hyperhidrosis and malodorous oily skin are common early signs, occurring in up to 70% of patients. Facial wrinkles, nasolabial folds, and heel pads thicken and body hair may become coarse. Hyperhidrosis and malodorous oily skin are common early signs, occurring in up to 70% of patients. Facial wrinkles, nasolabial folds, and heel pads thicken and body hair may become coarse. Hyperhidrosis and malodorous oily skin are common early signs, occurring in up to 70% of patients. Facial wrinkles, nasolabial folds, and heel pads thicken and body hair may become coarse.

Symptomatic cardiac disease is present in about 20% of patients and is a major cause of morbidity and mortality. Hypertension is present in about 50% of patients with active acromegaly. Left ventricular hypertrophy is also observed in about half of normotensive patients with acromegaly. Asymmetric septal hypertrophy is common, and cardiac failure may occur with early or mild cardiomegaly. Subclinical left ventricular diastolic dysfunction is due to myocardial hypertrophy, interstitial fibrosis, and lymphocytic myocardial infiltrates. Resting electrocardiograms are abnormal in about 50% of patients, with S-T segment depression, T-wave abnormalities, conduction defects, and arrhythmias. Plasma renin levels are suppressed, and endogenous plasma digitalis-like activity with chronic volume expansion has been identified in acromegaly. Cardiovascular disease accounts for approximately 60% of deaths in patients with acromegaly, and the presence of cardiovascular disease at the time of diagnosis portends high mortality rates despite improved cardiac function after effective GH and IGF-I control.

Prognathism, thick lips, macrognathia, and hypertrophied nasal structures may obstruct airways. Irregular laryngeal mucosa, cartilage hypertrophy, tracheal calcification, and cricoarytenoid joint arthropathy lead to unilateral or bilateral vocal cord fixation or laryngeal stenosis with voice changes and upper airway obstruction or stridor. Tracheal intubation may be particularly difficult in patients undergoing anesthesia, and tracheostomy may be required. Both central respiratory obstruction and airway obstruction lead to paroxysmal daytime sleep (narcolepsy), sleep apnea, and habitual excessive snoring. Obstructive sleep apnea, characterized by excessive daytime sleepiness with at least five episodes of apnea per hour of sleep, causes daytime somnolence, especially in men with acromegaly, who may also have a ventilation-perfusion defect with hypoxemia. Sleep apnea may also be central in origin and associated with higher GH and IGF-I levels.

Synovial edema leads to hyperplastic wrist ligaments and tendons that contribute to painful median nerve compression. Chondrocyte proliferation with an increased joint space and the presence of cardiovascular disease at the time of diagnosis portends high mortality rates despite improved cardiac function after effective GH and IGF-I control. Synovial edema leads to hyperplastic wrist ligaments and tendons that contribute to painful median nerve compression. Chondrocyte proliferation with an increased joint space and the presence of cardiovascular disease at the time of diagnosis portends high mortality rates despite improved cardiac function after effective GH and IGF-I control. Synovial edema leads to hyperplastic wrist ligaments and tendons that contribute to painful median nerve compression. Chondrocyte proliferation with an increased joint space and the presence of cardiovascular disease at the time of diagnosis portends high mortality rates despite improved cardiac function after effective GH and IGF-I control. Synovial edema leads to hyperplastic wrist ligaments and tendons that contribute to painful median nerve compression. Chondrocyte proliferation with an increased joint space and the presence of cardiovascular disease at the time of diagnosis portends high mortality rates despite improved cardiac function after effective GH and IGF-I control.
tags in patients older than 50 years may be peripheral markers for the presence of adenomatous colon polyps, unrelated to GH or IGF-I serum levels. Hypertrophic mucosal folds and colonic hypertrophy are commonly present; colonoscopy is warranted every 3 to 5 years after diagnosis, depending on the presence of other risk factors. Mortality from colon cancer is largely related to GH levels rather than the enhanced incidence of the disease in acromegaly (Table 8-26).

Analysis of nine retrospective reports (1956 to 1998) encompassing 21,470 person-years of risk yielded no significant increase in cancer incidence. Cancer incidence was, in fact, lower than expected in 1362 patients with acromegaly in the United Kingdom, and the enhanced colon cancer mortality observed in this study correlated with GH levels. Thus, although disordered cell proliferation and increased risk for promotion of coexisting neoplasms could be anticipated, there is little evidence for an increased cancer incidence in acromegaly (Table 8-27). Although elevated IGF-I levels may correlate with colon polyph prevalence when patients are retested, a controlled prospective study showed no increased colon polyp incidence in acromegaly. Patients are now living longer with improved biochemical control, and long-term prospective controlled studies are required to resolve this question in an aging population.

Endocrine Complications

About 30% of patients exhibit elevated serum PRL levels (up to 100 µg/L or more) with or without galactorrhea. Functional pituitary stalk compression by a pituitary mass prevents lactotroph access of hypothalamic dopamine, releasing the cell from tonic hypothalamic inhibition. GH-secreting adenoma subtypes may also concomitantly secrete PRL. Because GH behaves as an agonist for breast PRL binding sites, the

### Table 8-25 -- Colon Polyps in Acromegaly

<table>
<thead>
<tr>
<th>n</th>
<th>MF</th>
<th>Mean Age</th>
<th>Adenoma</th>
<th>Hyperplastic</th>
<th>Total</th>
<th>Carcinoma</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>10/7</td>
<td>49</td>
<td>5</td>
<td>3</td>
<td>8</td>
<td>2</td>
<td><img src="image1" alt="" /></td>
</tr>
<tr>
<td>12</td>
<td>11/11</td>
<td>56</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td><img src="image2" alt="" /></td>
</tr>
<tr>
<td>29</td>
<td>n.a.</td>
<td>47</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td><img src="image3" alt="" /></td>
</tr>
<tr>
<td>23</td>
<td>12/11</td>
<td>47</td>
<td>8</td>
<td>1</td>
<td>9</td>
<td>0</td>
<td><img src="image4" alt="" /></td>
</tr>
<tr>
<td>54</td>
<td>26/26</td>
<td>47</td>
<td>5</td>
<td>11</td>
<td>19</td>
<td>0</td>
<td><img src="image5" alt="" /></td>
</tr>
<tr>
<td>50</td>
<td>25/25</td>
<td>2570</td>
<td>11</td>
<td>12</td>
<td>23</td>
<td>1</td>
<td><img src="image6" alt="" /></td>
</tr>
<tr>
<td>49</td>
<td>30/19</td>
<td>54</td>
<td>11</td>
<td>5</td>
<td>16</td>
<td>0</td>
<td><img src="image7" alt="" /></td>
</tr>
<tr>
<td>31</td>
<td>11/20</td>
<td>52</td>
<td>11</td>
<td>8</td>
<td>16</td>
<td>0</td>
<td><img src="image8" alt="" /></td>
</tr>
<tr>
<td>103</td>
<td>49/54</td>
<td>51</td>
<td>23</td>
<td>25</td>
<td>48</td>
<td>0</td>
<td><img src="image9" alt="" /></td>
</tr>
<tr>
<td>129</td>
<td>68/60</td>
<td>57</td>
<td>33</td>
<td>42</td>
<td>75</td>
<td>6</td>
<td><img src="image10" alt="" /></td>
</tr>
<tr>
<td>115</td>
<td>63/69</td>
<td>54.8</td>
<td>27</td>
<td>18</td>
<td>45</td>
<td>3</td>
<td><img src="image11" alt="" /></td>
</tr>
<tr>
<td>66</td>
<td>NA</td>
<td>32.7</td>
<td>25</td>
<td>18</td>
<td>43</td>
<td>1</td>
<td><img src="image12" alt="" /></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>678</td>
<td>(24%)</td>
<td>(21%)</td>
<td>(45%)</td>
<td>(2.5%)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


*Incidence of colonic lesions in 678 patients prospectively evaluated in 12 studies. Of note, up to 45% of asymptomatic males age older than 50 years harbor colon adenomas. (Liebermann. Use of colonoscopy to screen asymptomatic adults for colon cancer. N Eng J Med 2000; 343:162168.)

Median age.

Repeated colonoscopy.

tumor may cause galactorrhea in the presence of normal PRL levels. Tumor mass compressing surrounding normal pituitary tissue may also cause hypopituitarism. Over half of all patients have amenorrhea or impotence, and secondary thyroid or adrenal failure are common in about 20% of patients. Gonadal dysfunction may result in reduced bone.

The direct anti-insulin effects of GH cause carbohydrate intolerance, and insulin-requiring diabetes mellitus may also develop. Carbohydrate intolerance and insulin requirements improve rapidly with lowering of GH after surgery or somatostatin analogue therapy. Hypertriglyceridemia (type IV), hypercalciuria, and hypercalcemia also occur. Thyroid dysfunction in acromegaly may be caused by diffuse or nodular toxic or nontoxic goiter or Graves’ disease, especially as IGF-I is a major determinant of thyroid cell growth. Associated MEN-1 features may be present in affected individuals, including hypercalcemia with hyperparathyroidism or pancreatic tumors. Benign prostatic hypertrophy has been documented in acromegaly with no apparent increase in prostate cancer rates.

Mortality and Mortality

Cardiovascular disease, respiratory disorders, diabetes, and malignancy account for the threefold enhanced mortality in acromegaly. In a retrospective study reported in 1966, cardiovascular disease was the leading cause of

### Table 8-26 -- Post-treatment Growth Hormone Levels and Mortality in Acromegaly

<table>
<thead>
<tr>
<th>Mortality Ratio</th>
<th>Post-treatment Growth Hormone (ng/mL)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;2.5 (n = 541)</td>
<td>2.59.9 (n = 493)</td>
<td>&gt;10 (n = 207)</td>
</tr>
<tr>
<td>Overall</td>
<td>1.10 (0.891.15)</td>
<td>1.41 (1.161.69)</td>
</tr>
<tr>
<td>Cancer-related</td>
<td>0.96 (0.631.41)</td>
<td>0.81 (0.501.24)</td>
</tr>
</tbody>
</table>


*Post-treatment GH levels correlate with mortality in acromegaly. Standardized mortality ratios are depicted for overall mortality and for cancer-related mortality.

death and 50% of patients died before the age of 50. Life expectancy was reduced in 194 patients with acromegaly, with cardiovascular disorders accounting for 24% of deaths followed by respiratory (18%) and cerebrovascular (14%) disease. Diabetes mellitus, occurring in 20% of patients, was associated with 2.5 times the predicted mortality, and hypertension was present in about 50% of all patients. The most significant mortality determinants are GH levels and the presence of coexisting cardiac disease. Moreover, control of GH levels to less than 2.5 µg/L after surgery or medical treatments significantly reduces both morbidity and mortality (Fig. 8-50).
The diagnosis of acromegaly requires measurement of a random GH higher than 0.4 µg/L or a GH nadir greater than 1 µg/L during an oral glucose tolerance test. In healthy subjects, serum GH levels initially fall after oral glucose and subsequently increase as plasma glucose declines. However, in patients with acromegaly, oral glucose fails to suppress GH; GH levels may increase, remain unchanged, or fall modestly in approximately one third of patients. Basal morning (**) and random GH levels are usually elevated in acromegaly. Because of the episodic nature of GH secretion, however, serum concentrations may normally fluctuate from undetectable up to 30 µg/L. In contrast to the largely undetectable nadir GH levels in normal subjects, detectable levels of GH (>2 µg/L) were found in those with acromegaly sampled over 24 hours. (**)

Elevated IGF-I levels are also encountered during pregnancy and late puberty. A high IGF-I level is thus highly specific for acromegaly and correlates with clinical indices of disease activity. IGF-BP3 levels are also elevated but provide little added diagnostic information. GH-secreting adenomas exhibit discordant GHI responses to TRH and GnRH administration in up to 50% of patients, but these adjunctive tests are rarely required to confirm the diagnosis.

Differential Diagnosis of Acromegaly

The overwhelming majority of patients with acromegaly harbor a GH cell pituitary adenoma; rarely, extrapituitary acromegaly should be considered. Nevertheless, distinguishing between pituitary and extrapituitary acromegaly is important for planning effective management. Regardless of the cause of unrestrained GH secretion, IGF-I levels are invariably elevated and GH levels are not suppressed (<1 µg/L) after an oral glucose load. When clinical features of acromegaly are associated with normal GH and IGF-I levels, "burned out" acromegaly associated with an infarcted pituitary adenoma, often with a secondary empty sella, should be considered. About 5% of consecutive patients with proven GH cell adenomas have normal GH and elevated IGF-I levels. It is likely that improved GH assay sensitivity will unmask abnormal GH secretion in these patients.

Dynamic pituitary testing (with TRH, dopamine) does not differentiate patients with pituitary adenomas from those harboring extrapituitary tumors. Plasma GHRH levels are invariably elevated in patients with peripheral GHRH-secreting tumors but are normal or low in patients with pituitary adenomas. GHRH plasma level measurement is precise and cost-effective for diagnosis of ectopic acromegaly. Peripherial GHRH levels are not elevated in patients with hypothalamic GHRH-secreting tumors, presumably because eulotropic hypothalamic GHRH secreted into the hypophyseal portal system does not appreciably enter the systemic circulation.

Unique or unexpected clinical features, including respiratory wheezing or dyspnea, facial flushing, peptic ulcers, or renal stones, sometimes indicate the diagnosis of a nonpituitary endocrine tumor. Hypoglycemia, hyperinsulinemia, hypergastrinemia, and rarely hypercortisolism, all not usually encountered in pituitary acromegaly, should justify an evaluation for an extrapituitary source of GH excess. MRI and CT scanning are employed to localize a pituitary or extrapituitary tumor. Routine abdominal or chest imaging of all patients yields a low incidence of true positive cases of ectopic tumor, and such screening is not recommended as cost-effective.

A normal-sized or small pituitary gland or clinical and biochemical features of other tumors known to be associated with extrapituitary acromegaly and elevated circulating GHRH levels are indications for extrapituitary imaging. An enlarged pituitary is, however, often present in patients with peripheral GHRH-secreting tumors, and the radiologic diagnosis of a pituitary adenoma may be difficult to exclude. The McCune-Albright syndrome should be considered after definitive exclusion of a nonpituitary endocrine tumor. Hypoglycemia, hyperinsulinemia, hypergastrinemia, and rarely hypercortisolism, all not usually encountered in pituitary acromegaly, should justify an evaluation for an extrapituitary source of GH excess. MRI and CT scanning are employed to localize a pituitary or extrapituitary tumor. Routine abdominal or chest imaging of all patients yields a low incidence of true positive cases of ectopic tumor, and such screening is not recommended as cost-effective.

A comprehensive strategy for treating patients with acromegaly should aim to manage the pituitary mass, suppress GH and IGF-I hypersecretion, and prevent long-term clinical sequelae of hypogonadotropism while maintaining normal anterior pituitary function. As elevated GH levels per se are associated with threefold increased morbidity and account for the single most important determinant of mortality, it is important to reverse the mortality rate to that of age-matched healthy subjects by aiming for tight GH control.

---

**Table 8-27 -- Acromegaly and Cancer Incidence: Multicenter Analyses.**

<table>
<thead>
<tr>
<th>Gender</th>
<th>n</th>
<th>Person-Years at Risk</th>
<th>Cancers</th>
<th>Observed</th>
<th>O/E</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Females</td>
<td>95</td>
<td>1531</td>
<td></td>
<td>8</td>
<td>1.33</td>
<td>NS</td>
</tr>
<tr>
<td>Males</td>
<td>128</td>
<td>1630</td>
<td></td>
<td>5</td>
<td>1.30</td>
<td>NS</td>
</tr>
<tr>
<td>Total</td>
<td>223</td>
<td>2981</td>
<td></td>
<td>13</td>
<td>1.3</td>
<td></td>
</tr>
</tbody>
</table>

GH levels should be suppressed to less than 1 µg/L after an oral glucose load and serum IGF-I levels normalized for age and gender. A patient whose condition is controlled should also have a normal 24-hour integrated secretion of GH (<2.5 µg/L). GH may not be measurable for most of the day, but the tumor may still be hypersecreting as reflected by elevated IGF-I levels. Current therapeutic modes for acromegaly management, including surgery, irradiation, and medical treatment, do not comprehensively fulfill these goals.

**Surgical Management**

Well-circumscribed somatotroph cell adenomas should preferably be resected by transphenoidal surgery. Successful resection alleviates preoperative compression effects and compromised trophic hormone secretion, and the skilled surgeon balances the extent of maximal tumor tissue removal with preserving anterior pituitary function. Within 2 hours of successful resection, metabolic dysfunction and soft tissue swelling start to improve, and GH levels are often controlled within an hour. Surgical outcome correlates well with adenoma size and preoperative serum GH levels and particularly with the experience of the surgeon.

Smaller tumors (less than 5 mm), those totally confined within the sella, and preoperative serum GH levels lower than 40 µg/L portend a favorable surgical outcome. Up to 90% of patients with microadenomas achieved postoperative GH levels less than 2.5 µg/L, and less than 50% of those with all-sized macroadenomas had postoperative GH levels below 2 µg/L after glucose administration (Table 8-16). Less than one third of all patients achieve control after resection of adenomas larger than 10 mm, and about 75% of patients with preoperative GH less than 5 µg/L have normalized IGF-I. Overall, in 17 studies of 1284 patients published between 1995 and 1999, 82% of patients harboring macroadenomas had normalized IGF-I levels compared with less than 50% of those with macroadenomas (Table 8-28).

A review of 2665 patients from a single center showed that 72% of patients with microadenomas and 50% harboring macroadenomas had GH levels below 1.0 µg/L during glucose load and normal serum IGF-I levels. Less than one third of these patients had recurrences after 10 years. Endoscopic transnasal surgery offers promise as a less invasive procedure for resection of pituitary tumors and access to cavernous sinus tumor mass, although long-term comparative results are not yet available. Difficulties in endotracheal intubation related to macroglossia or severe kyphosis, or both, may necessitate tracheostomy for anesthesia.

**Side Effects.**

Although often transient, surgical complications may require lifelong pituitary hormone replacement. New hypopituitarism develops in up to 20% of patients, reflecting operative damage to the surrounding normal pituitary tissue. Permanent diabetes insipidus, CSF leaks, hemorrhage, and meningitis occur in up to 10% of patients (Table 8-10). The extent and prevalence of local complications depend on tumor size and invasiveness. Experienced pituitary surgeons report more favorable postoperative complication rates. Biochemical or anatomic recurrence (7% or 10 years) or postoperative tumor persistence may indicate incomplete resection of adenomatous tissue, surgically inaccessible cavernous sinus tissue, or nesting of functional tumor tissue within the dura.

**Radiation**

Primary or adjuvant radiation of GH-secreting tumors may be achieved by conventional external deep X-ray therapy as well as heavy-particle (proton beam) therapy. Maximal tumor radiation should ideally be attained with minimal soft tissue damage. Precise MRI localization, accurate simulation and isocentral rotational techniques, and high-voltage (6 to 15 MEV) delivery have improved radiation efficacy. Radiation is a highly individual choice, depending on the expertise and experience of the treating radiotherapist as well as the physician's and patient's choice of the benefits of therapy weighed against potential risks.

For radiation treatment, up to 5000 rads are administered in split doses of 180-cGy fractions divided over 6 weeks. Radiation arrests tumor growth, and most pituitary adenomas ultimately shrink. GH levels fall gradually during the first year after treatment, and levels are less than 10 µg/L in 70% of patients after 10 years. The National Institutes of Health experience is that over 90% of patients have GH levels less than 5 µg/L after 20 years (Fig 8-51). When pretreatment GH levels were greater than 100 µg/L, only 60% of patients had GH less than 5 µg/L after 18 years. During the first 7 years after irradiation less than 5% of patients had normal GH <2 µg/L (<4 mU/I) after OGTT and/or normal IGF-I levels. GH <5 µg/L (<10 mU/I) and/or normal IGF-I levels

<table>
<thead>
<tr>
<th>Series</th>
<th>Number of Cases</th>
<th>Total Cure Rate (%)</th>
<th>Microadenomas (%)</th>
<th>Macroadenomas (%)</th>
<th>Definition of &quot;Cure&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ross and Wilson, 1988</td>
<td>153</td>
<td>56</td>
<td>NA</td>
<td>NA</td>
<td>GH &lt;5 µg/L (&lt;10 mU/I)</td>
</tr>
<tr>
<td>Losa et al., 1989</td>
<td>29</td>
<td>55</td>
<td>NA</td>
<td>NA</td>
<td>GH &lt;1 µg/L (&lt;2 mU/I) and normal IGF-I levels</td>
</tr>
<tr>
<td>Fahrbuscher et al., 1992</td>
<td>222</td>
<td>57</td>
<td>72</td>
<td>49</td>
<td>GH &lt;2 µg/L (&lt;4 mU/I) after OGTT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GH &lt;5 µg/L (&lt;10 mU/I) and/or normal IGF-I levels</td>
</tr>
<tr>
<td>Tindall et al., 1993</td>
<td>91</td>
<td>82</td>
<td>NA</td>
<td>NA</td>
<td>GH &lt;5 µg/L (&lt;10 mU/I) and/or normal IGF-I levels</td>
</tr>
<tr>
<td>Davis et al., 1993</td>
<td>174</td>
<td>52</td>
<td>NA</td>
<td>NA</td>
<td>GH 2 µg/L (&lt;4 mU/I; basal or OGTT)</td>
</tr>
<tr>
<td>Sheaves et al., 1996</td>
<td>100</td>
<td>42</td>
<td>61</td>
<td>23</td>
<td>GH 2.5 µg/L (&lt;5 mU/I)</td>
</tr>
<tr>
<td>Aboesch et al., 1998</td>
<td>254</td>
<td>76</td>
<td>75</td>
<td>71</td>
<td>GH &lt;5 µg/L (&lt;10 mU/I)</td>
</tr>
<tr>
<td>Freda et al., 1998</td>
<td>115</td>
<td>61</td>
<td>88</td>
<td>53</td>
<td>GH &lt;2 µg/L (&lt;4 mU/I) OGTT or normal IGF-I levels</td>
</tr>
<tr>
<td>Swearingen, 1998</td>
<td>162</td>
<td>57</td>
<td>91</td>
<td>48</td>
<td>Normal IGF-I levels</td>
</tr>
<tr>
<td>Lissett, 1998</td>
<td>73</td>
<td>18</td>
<td>39</td>
<td>12</td>
<td>GH &lt;5 µg/L OGTT</td>
</tr>
<tr>
<td>Ahmed et al., 1999</td>
<td>97</td>
<td></td>
<td></td>
<td></td>
<td>Basal GH 2.5 µg/L (&lt;5 µU/I), OGTT GH &lt;1 µg/L (&lt;2 mU/I), normal IGF-I levels</td>
</tr>
<tr>
<td>Laws et al., 2000</td>
<td>117</td>
<td>67</td>
<td>87</td>
<td>51</td>
<td>Basal GH &lt;2.5 µg/L (&lt;5 µU/I), OGTT GH &lt;1 µg/L (&lt;2 mU/I), normal IGF-I levels</td>
</tr>
<tr>
<td>Fahrbuscher, 2001</td>
<td>490</td>
<td>56</td>
<td>78</td>
<td>50</td>
<td>Basal GH 5 µg/L (&lt;10 µU/I), OGTT GH &lt;2 µg/L (4 µU/I), normal IGF-I levels</td>
</tr>
</tbody>
</table>

IGF-I levels but about 70% of patients exhibited normal IGF-I levels when treated during longer follow-up. Radiotherapy does not normalize GH secretory patterns, probably accounting for persistently elevated IGF-I levels in the presence of apparently controlled GH levels. Thus, during the initial years after irradiation, most patients are still exposed to unacceptably high levels of circulating GH and IGF-I. Promising stereotactic pituitary tumor ablation by the gamma knife has been reported. Long-term outcomes and potential long-term side effects are not yet available.

Side Effects

After 10 years about half of all patients receiving radiotherapy have signs of pituitary trophic hormone disruption, and the prevalence increases annually thereafter, requiring gonadal steroids, thyroid hormone, or cortisone replacement. Side effects of conventional radiation, including hair loss, cranial nerve palsies, tumor necrosis with hemorrhage, and rarely loss of vision or pituitary apoplexy, have been documented in up to 2% of patients. Lethargy, impaired memory, and personality changes may also occur. The incidence and extent of local complications have been markedly diminished by use of highly reproducible simulators, precise rotational isocentric arc capability, and doses less than 5000 cGy. Proton beam therapy (Bragg peak) is contraindicated in patients with suprasellar tumor extension because of unacceptably optic tract exposure to the radiation field. The rare development of second brain tumors in these patients has been reported at a cumulative risk frequency of 1.9% over 20 years.

Radiation therapy effectively shrinks over 95% of GH cell adenomas and lowers GH levels over 20 years in more than 90% of patients. In fact, GHD may result from radiation because of the side effects, radiation therapy should be employed as an adjuvant for patients not controlled by surgery or medical management or for those who refuse these therapies.

Dopamine Agonists

Because dopamine attenuates GH secretion in about one third of patients with acromegaly, D2 receptor agonists, including bromocriptine and cabergoline, have been used as either primary or adjuvant therapy for acromegaly. Bromocriptine may lower GH at a dose of 20 mg/day or more, which is higher than required to suppress PRL in patients harboring prolactinomas. Approximately 15% of patients worldwide have been reported to have suppressed GH levels below 5 µg/L when taking the medication (7.5 to 80 mg/day), and IGF-I became normal in 10% of patients. The drug causes minimal tumor shrinkage, but most patients experience subjective clinical improvement and report reduced perspiration, decreased soft tissue swelling, and improved fatigue and headache despite persistently elevated serum GH or IGF-I levels, or both. Side effects of bromocriptine are more marked, especially as high doses are required. These include gastrointestinal upset, transient nausea and vomiting, headache, transient postural hypotension with dizziness, nasal stuffiness, and rarely cold-induced peripheral vasospasm (see earlier).

Cabergoline, a long-acting dopamine agonist, is highly effective in suppressing PRL hypersecretion and shrinking prolactinomas. The drug has been reported to suppress GH to less than 2 µg/L and to normalize IGF-I in up to a third of patients with acromegaly. Side effects include gastrointestinal symptoms, dizziness, headache, and mood disorders. Patients with hyperprolactinemia and minimal GH elevation may benefit most from dopamine agonist treatment.

Somatostatin Release-Inhibiting Factor Receptor Ligands

Of the five SRIF receptor subtypes, SSTR2 and SSTR5 are preferentially expressed on somatotroph and thyrotroph cell surfaces and mediate suppression of GH and TSH secretion. Several SRIF ligands have been employed as approved or investigational drugs for acromegaly (Fig. 8-52). For over 15 years, these analogues have proved safe and effective for controlling acromegaly.

Octreotide (po-Phe-Cys-Phe-o-Trp-Lys-Thr-Cys-Thr-OH), an octapeptide SRIF analogue, binds predominantly to SSTR2 and SSTR5 and inhibits GH secretion with a potency 45 times greater than that of native SRIF, whereas its potency for inhibiting insulin release is only 1.3-fold that of SRIF. The in vivo half-life of the analogue is prolonged (up to 2 hours) because of its relative resistance to enzymatic degradation. The rebound GH hypersecretion seen after SRIF infusion does not occur after octreotide injection. These properties are highly advantageous for long-term use in acromegaly. A single subcutaneous dose (50 or 100 µg) suppresses GH secretion for up to 5 hours. In patients harboring microadenomas, integrated GH and IGF-I levels almost invariably became normal, but the response in larger tumors was less pronounced. In a double-blind, placebo-controlled trial, octreotide (8-hourly injections) significantly attenuated GH and IGF-I levels overall in over 90% of patients. A combination of octreotide and bromocriptine or cabergoline may provide added efficacy. In an open-label study of 151 patients responsive to octreotide, the analogue suppressed serum GH levels to less than 2.5 µg/L in about 70% of patients. Interestingly, tumor shrinkage was observed after up to 2 years in 12 of 15 patients receiving octreotide LAR as primary treatment, whereas no shrinkage was observed in 4 of 9 patients after surgery. Fifty-nine patients undergoing pituitary surgery were randomly assigned to treatment with or without octreotide, and 22 who received preoperative octreotide for 3 to 6 months demonstrated improved postoperative biochemical control and reduced hospital length of stay.

Effects of SRIF Receptor Ligands on Pituitary Adenomas

Invariably, tumor growth does not occur while patients receive depot preparations of SRIF analogues, and adenomas shrink by 20% to 80% in about one third of patients receiving these analogues. Interestingly, tumor shrinkage was observed after up to 2 years in 12 of 15 patients receiving octreotide LAR as primary treatment, whereas no shrinkage was observed in 4 of 9 patients after surgery. Fifty-nine patients undergoing pituitary surgery were randomly assigned to treatment with or without octreotide, and 22 who received preoperative octreotide for 3 to 6 months demonstrated improved postoperative biochemical control and reduced hospital length of stay. Over 70% of patients experience improved general well-being, and soft tissue swelling dissipates within several days of treatment, reflecting a specific central analgesic effect. Asymptomatic patients experience a significant decrease in acromegaly, usually resolves within minutes of injection,
of blood pressure, heart rate, and left ventricular wall thickness. In patients with cardiac failure, octreotide reversibly reduces systemic arterial resistance, oxygen consumption, and fluid volume and restores functional activity. In 30 patients, an improved left ventricular ejection fraction with unchanged diastolic filling was associated with octreotide-induced GH suppression to less than 2.5 µg/L. Persistently elevated GH levels after a year were associated with increased systolic blood pressure. Control of IGF-I and GH levels is associated with improved left ventricular ejection function; in patients in whom the levels were not controlled, cardiac performance worsened. Joint function and crepitus improve, ultrasonography shows evidence of bone or cartilage repair, and after several months sleep apnea improves.

Side Effects

SRIF receptor ligands are generally safe and well tolerated. Gastrointestinal side effects predominate and include transient loose stools, nausea, mild malabsorption, and flatulence, reported in about one third of patients. Hypoglycemia and hyperglycemia are not commonly encountered, and insulin requirements in diabetic patients with acromegaly are dramatically reduced within hours of receiving octreotide, concomitant with GH lowering. The drug attenuates gallbladder contractility, delays emptying, and leads to reversible sludge formation evidenced by ultrasonography in up to 25% of patients. Frank cholecystitis is rarely reported in these patients. The incidence of gallbladder sludge or stones is geographically variable, with higher rates reported in China, Australia, and the United Kingdom. In the United States, up to 30% of patients have demonstrable evidence of echogenic gallbladder deposits within the first 18 months of treatment. Thereafter, further sludge formation is not usually encountered. Octreotide may interact with several drugs including cyclosporine, enhancing transplant rejection risk. SRIF receptor ligand dose adjustments should be carefully titrated in patients requiring insulin or oral hypoglycemic agents, calcium channel blockers, and α-blockers. Asymptomatic sinus bradycardia has also been recognized.

Growth Hormone Receptor Antagonist

GH action through the surface membrane GH receptor is mediated by ligand-induced GH receptor dimerization and subsequent receptor signaling. The postreceptor GH signal is not elicited if the receptor fails to dimerize. Pegasysom, a GH receptor antagonist, blocks receptor dimerization and subsequent IGF-I generation. The pegylated molecule also binds to the GH receptor dimer and interacts with GHBP. Daily injections (20 mg) of this pegylated GH mutant molecule normalized IGF-I levels in over 90% of patients and dose dependently improved fatigue, decreased soft tissue swelling assessed by ring size, and diminished perspiration. The drug, not yet approved in the United States, may be particularly useful in patients resistant to somatostatin receptor ligand therapy as it effectively normalizes IGF-I levels in these patients. Long-term side effects are not yet known, but the action of the medication increases levels of GH, which are bioinactive because of receptor blockade. Long-term surveillance should also include monitoring of liver function and pituitary adenoma size.

Choice of Therapy

Tight control of GH secretion should be achieved because adverse mortality rates correlate strongly with GH levels. Each treatment modality has respective advantages and disadvantages that should be weighed in order to individualize patients' care. Selective surgical excision of a well-defined pituitary microadenoma is recommended for most patients. Remission rates are unacceptably low for patients with macroadenomas and locally invasive tumors. Attempted medical debulking of the sellar mass prior to surgery would be desirable, although controlled prospective studies are required to confirm the validity of this approach to improve surgical morbidity and possibly enhance subsequent postoperative outcomes, especially for patients with surgically inaccessible tumor tissue and cavernous sinus invasion.

Postoperatively, patients who do not achieve control can be treated with bromocriptine; although the efficacy of this drug is low, it is relatively inexpensive and free of major side effects. An SRIF analogue should be administered, GH and IGF-I measured after 2 hours, and Sandostatin LAR (10, 20, or 30 mg) initiated if patients are shown to be responsive. More frequent dosing rather than increases in total drug dose may be more efficacious and beneficial, and some may benefit by addition of bromocriptine or cabergoline with octreotide.

Gallbladder

Table 8–29 — Treatment Options for Acromegaly

<table>
<thead>
<tr>
<th>Surgery</th>
<th>Somatostatin Analogue</th>
<th>Radiotherapy</th>
<th>Dopamine Agonists (High Dose)</th>
<th>Growth Hormone-Receptor Antagonist</th>
</tr>
</thead>
<tbody>
<tr>
<td>Efficacy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80% of microadenomas: GH controlled</td>
<td>GH controlled in 65% of patients</td>
<td>GH &lt;5 µg/L in 90% of patients in 18 yr</td>
<td>GH &lt;5 µg/L in 15%</td>
<td>Elevated bioactive GH</td>
</tr>
<tr>
<td>&lt;50% of microadenomas: GH controlled</td>
<td>Normal IGF-I in 70%</td>
<td>Normal IGF-I in &gt;10 yr 54%</td>
<td>Normal IGF-I in 10%</td>
<td>Normal IGF-I in &gt;90%</td>
</tr>
<tr>
<td>IGF-I normalized in 50%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Advantages</td>
<td>Rapid onset</td>
<td>No hypopituitarism</td>
<td>Permanent</td>
<td>Oral administration</td>
</tr>
<tr>
<td></td>
<td>Rapid onset</td>
<td>One-time cost</td>
<td>Low cost</td>
<td>No hypopituitarism</td>
</tr>
</tbody>
</table>
Maybe permanent control | Sustained long-term efficacy | Good compliance by patients | No hypopituitarism | Sustained efficacy

**Disadvantages**

- New hypopituitarism (10%)
  - Cost of drug and monitoring
  - Ineffective and slow on-set
  - Relatively ineffective
  - Long-term safety unknown
- Diabetes insipidus (23%)
  - Asymptomatic gallstones (25%)
  - Hypopituitarism (70%)
  - Adverse events (30%)
- Local complications (6%)
  - Injections required
  - Visual and CNS dysfunction (2%)
  - High dose required
  - Not yet approved (2002)
- Cranial nerve or CNS damage (1%)
  - Cost of interim medical therapy
- Tumor persistence

**Adapted from Melmed S, Jackson I, Kleinberg D, Klibanski A. Current treatment guidelines for acromegaly. J Clin Endocrinol Metab 1998; 83:2646-2652.**

Ultrasoundography should be performed only in symptomatic patients, and those with demonstrable sludge or gallstones may require prophylactic anticholelithogenic agents or laparoscopic cholecystectomy if symptoms develop.

Primary therapy with SRIF receptor ligands may be offered to patients who refuse surgery or in whom the risks of surgery or anesthesia are unacceptable. Invasive macroadenomas invariably hypersecrete GH postoperatively and require somatostatin analogue treatment. In patients whose pituitary lesion does not compress vital structures, primary medical management may therefore be an appropriate therapeutic option. Radiation should be administered to patients who are resistant to or cannot tolerate the medication, prefer not to receive long-term injections, or cannot afford the medication. After radiation, medications are required for several years until GH levels are effectively controlled.

Tumors that recur despite medical therapy or radiation may rarely require reoperation. Although tight GH control is critical, these patients also require counseling for anxiety engendered by disfigurements and interpretation of laboratory test results. Patients should be observed quarterly until biochemical control is achieved. Thereafter, hormone evaluation is performed semiannually, and for patients who are biochemically in remission and in whom no residual tumor tissue is present, MRI should be repeated every 1 to 2 years. Follow-up evaluation includes documenting and treating new skin tag and lipoma growth; nerve entrapments; jaw overbites; rheumatologic, dental, and cardiac evaluations; and metabolic assessment. Visual field perimetry and pituitary reserve testing should be repeated semiannually and pituitary MRI annually, especially in patients with residual tumor or those requiring hormone replacement or medical treatment. Mammography and colonoscopy should be performed as clinically indicated for patients older than 50 or those harboring polyps. Maximal and sustained long-term GH and IGF-I control should ameliorate the deleterious effects of these hormones by judicious use of available treatment modalities.
Adrenocorticotropic Hormone

Corticotroph Cells

Corticotroph cells constitute about 20% of functional anterior pituitary cells and are the earliest detectable human fetal pituitary cell type, appearing by the eighth week of gestation. Corticotrophs are clustered mainly in the central median pituitary wedge and are readily identified by immunostaining with ACTH or ß-lipotropin antibodies. They are large, irregular cells with ultrastructural features including prominent neurosecretory granules (150 to 400 nm), endoplasmic reticulum, and Golgi bodies (Fig. 8-56 (Figure Not Available) and Fig. 8-57 (Figure Not Available)). These cells produce the POMC gene products including ACTH 1 to 39, ß-lipotropin, and endorphins. Because of the rich carbohydrate moiety of these molecules, the cells are strongly positive for periodic acid-Schiff stain. In the presence of excess glucocorticoid, characteristic hyaline deposits are evident (Fig. 8-58) (Figure Not Available).

Adrenocorticotropic Hormone Biosynthesis

The 8-kb human POMC gene, located on chromosome 2p23, consists of three exons with two intervening introns (Fig. 8-59). The first exon encodes a leader sequence, the second encodes the signal initiation sequence and the N-terminal portion of the POMC peptides, and the third exon encodes most of the mature peptide sequences and the chemokine-like regions and ß-lipotropin.

POMC is the precursor of ACTH, which acts on the adrenal glands to induce synthesis and secretion of adrenal steroids. The primary translation product of POMC is a 266-aminoacid POMC pro-hormone molecule encoding corticotrophic, opioid, and melanotropic peptides. The peptide contains a leader sequence and multiple dibasic proteolytic cleavage sites for glycosylation, acetylation, and amidation. Products of this processing include ACTH 1 to 39 and ß-lipotropin, which in turn give rise to ß-lipotropin and ß-endorphin, also containing met-enkephalin. ACTH itself may also be cleaved to melanocyte-stimulating hormone (1 to 13) and corticotroplike intermediate lobe peptide (CLIP; 18 to 39). The neurointermediate pituitary lobe is not developed in humans and is not normally a source of circulating POMC-derived peptides.

Multiple signals in synergy induce ACTH gene expression. These include CRH, cytokines, arginine vasopressin, catecholamines, and VIP. Glucocorticoids inhibit POMC gene expression. The CRH type 1 receptor is predominantly expressed on the corticotroph, and receptor activation increases cAMP, protein kinase A, and CREB induction of CRHBP binding to the promoter leading to POMC transcription. CRH also activates an AP-1 site within the first exon by a mitogen-activated protein kinase-mediated pathway. In addition to mediating ACTH secretion, this receptor appears critical for fear and anxiety responses, possibly through a related ligand, uricorin. The type 2 CRH receptor is predominantly important for cardiovascular function.

Leukemia inhibitory factor (LIF), a proinflammatory cytokine also expressed in the pituitary and hypothalamus, signals through the JAK/STAT pathway, acts in synergy with CRH, and induces direct STAT3 binding to POMC. Glucocorticoid receptor activation leads to transcriptional suppression through two cooperative binding sites. The intracellular glucocorticoid receptor binds directly to S-regulatory elements to suppress POMC transcription. CRH action is also potentiated by vasopressin (acting through phospholipase C) and ß-adrenergic catecholamines by enhancing POMC mRNA levels, increasing ACTH secretion, or both. The net effects of these intracellular signals are to regulate POMC gene transcription, peptide synthesis, and ACTH secretion for mediating appropriate neuroendocrine responses.

Pro-opiomelanocortin Processing

Several post-translational POMC modification steps are required for ultimate polypeptide hormone secretion (Fig. 8-60). First, the N-terminal signal sequence is removed followed by glycosylation through an O linkage to Thr45 and N linkage to Asn65. Serine phosphorylation then occurs within the Golgi apparatus. After being transported to secretory vesicles, the constituent peptides are cleaved at dibasic amino acid residues and ACTH-related peptides stored in dense secretory granules for ultimate regulated release. Some POMC products also undergo C-terminal amidation mediated by peptidylglycine-aminating monoxygenase and peptidylhydroxyglycine-aminating lase and N-terminal acetylation. POMC proteolytic processing occurs at Lys-Arg or Arg-Arg residues. Prohormone convertase 1 (PC1) or prohormone convertase 2 (PC2), related to the subtilisin-kexin proteinases, exert tissue-specific cleavage.
activities at dibasic sites. PC1 is most abundant in the pituitary and hypothalamus, and PC2 is present in white blood cells. Nevertheless, most circulating ACTH is derived from the anterior pituitary or from neuroendocrine tumor ectopic production. Extrapituitary neuroendocrine tumors associated with ectopic ACTH secretion do not process the prohormone efficiently. As ACTH is synthesized in non-tumorous neuroendocrine cells, "ectopic" tumor hormone production may in fact reflect inappropriate ACTH processing. These patients also exhibit a higher ratio of circulating ACTH precursors as well as smaller peptides, including CLIP.

Adrenocorticotropic Hormone Secretion

The complex control of ACTH secretion patterns is critical for maintenance of adrenal cortex function and reflects the neuroendocrine control of stress homeostasis. Essential metabolic and endocrine functions require a sensitively controlled nonstress pattern of hypothalamic-pituitary-adrenal (HPA) axis function. This baseline pattern allows the axis to mount an appropriate stress response, with a well-buffered reserve capacity to counteract life-threatening insults.

The 4.5-kd ACTH polypeptide consists of 39 amino acids. The highly conserved 12 N-terminal amino acid residues are critical for adrenal gland steroid synthesis. Several variables characterize the central and peripheral control of ACTH secretion. The ACTH circadian rhythm is generated in the suprachiasmatic nucleus, which signals CRH release. The hormone is secreted with both circadian periodicity and ultradian pulsatility, and this centrally controlled pattern is influenced by peripheral corticosteroids. ACTH pulse amplitudes may vary by 40% over 24 hours. The circadian pattern of ACTH secretion typically begins at about 4 AM and peaks before 7 AM, with both ACTH and adrenal steroid levels reaching their nadir between 11 PM and 3 AM. Within this overall diurnal cycle, periodic ACTH secretory bursts occur at a frequency of 40 pulses per 24 hours, with changing pulse amplitudes throughout the day. Each pulse contains an average ACTH level of 24 ng/L.

Pulse amplitude changes, rather than frequency, appear to determine ACTH circadian rhythm. ACTH circadian rhythm is entrained by visual cues and the light-dark cycle and is centrally controlled by CRH and other factors. Although continuous CRH administration desensitizes the ACTH response,

Adrenocorticotropic Hormone Action

The primary action of ACTH is to maintain adrenal gland size, structure, and function; ACTH induces adrenal steroidogenesis by activating ACTH receptors situated on the adrenal cortex cell surface. ACTH signals through adenyl cyclase to regulate P450 enzyme transcription, cortisol aldosterone (10%), 17-hydroxyprogestosterone, and to a lesser extent adrenal androgen synthesis and secretion. ACTH stimulates mitochondrial cholesterol transport and regulates the rate-limiting side-chain cleavage of cholesterol to pregnenolone. Secretory cortisol pulses follow ACTH pulses within 5 to 10 minutes, with a linear dose dependence, especially evident after physiologic CRH stimulation. However, circulating cortisol levels reach a plateau when pharmacologic levels of ACTH are attained by cosyntropin (Cortronesyn; Synacthen) injection.

The adrenal cortisol response to ACTH is sensitive to the background ambient ACTH milieu. In states of chronic ACTH deficiency, adrenal reserve is compromised, whereas during ongoing ACTH hypersecretion the gland is primed so that a given ACTH bolus elicits a higher cortisol response. Both basal ACTH secretion and stimulated (e.g., by CRH) ACTH secretion are blunted by glucocorticoids. Conversely, low or absent circulating glucocorticoids (e.g., after adrenalectomy) result in exaggerated ACTH secretion and corticotroph cell hyperplasia. The HPA axis is inhibited by a low feedback inhibition whereby cortisol rapidly inhibits hypothalamic CRH and pituitary ACTH. These effects may also be delayed by 30 to 60 minutes by inhibition of ACTH release rather than synthesis, especially after a single glucocorticoid bolus. After chronic glucocorticoid exposure (>24 hours), HPA suppression may persist for days or longer. In a short feedback loop, pituitary ACTH inhibits hypothalamic CRH, and in an ultrashort loop it may also suppress the corticotroph itself.

Physiologic Adrenocorticotropic Hormone Regulation

Exercise enhances ACTH and -endorphin levels, especially if it is exhausting and of short duration. Exercising up to 90% of maximum oxygen capacity causes a significant elevation of ACTH, similar to levels observed during surgery or hypoglycemia. Levels may remain elevated for up to 6 minutes after exercise cessation, and lower intensity exercise does not evoke ACTH. Well-trained athletes exhibit hypercortisolism, possibly a result of decreased adrenal ACTH sensitivity. Other causes of elevated ACTH include acute hemorrhage, surgery, and emotional stress. Acute illness is associated with increased ACTH and cortisol levels with loss of diurnal secretory patterns.

Stress Response

Both exogenous and endogenous stress stimuli activate the HPA axis to produce sufficient glucocorticoid in an attempt to counteract the insult. The HPA stress response occurs in the context of a wide variety of peripheral and central adaptors to stress, including vasovagal and catecholamine activation and cytokine secretion and action. A tightly controlled immunoneuroendocrine interface regulates the ACTH response to peripheral stressors, which include pain, infection, inflammation, hypovolemia, trauma, psychologic stress, and hypoglycemia. These signals vary in their ability to generate ACTH secretion and to sensitize the ACTH response to glucocorticoids. In addition to CRH, peripheral and centrally released proinflammatory cytokines potentially induce POMC transcription and ACTH secretion. Sensitive intracellular signals within the corticotroph also serve to override the ACTH response to stress, thus preventing persistent and chronic hypercortisolism.

Cytokines such as interleukin-6 and LIF activate the HPA axis and enhance glucocorticoid production, thus protecting the organism against lethality by constraining the inflammatory response. Thus, mice with inactivated CRH or LIF genes mount an inadequate neuroendocrine response to stress, inflammation, or endotoxins. During stress, glucocorticoid inhibition of ACTH is also prevented by nuclear factor B activation, which interferes with pituitary glucocorticoid receptor function, thus further exaggerating enhanced ACTH secretion.

Integrated Regulation of Adrenocorticotropic Hormone Secretion

As with other anterior pituitary hormones, ACTH regulation is subserved by at least three tiers of control. First, the brain and hypothalamus release regulatory molecules (including CRH, vasopressin, and other peptides) that traverse the portal system and directly signal for corticotroph secretory and mitotic activity. Second, intrapituitary cytokines and growth factors act locally to regulate ACTH either in concert with hypothalamic factors or independently. These paracrine controls often overlap and are redundant, and they have been shown to induce sensitive intracellular molecules that limit the ACTH response and prevent chronic ACTH hypersecretion. Third, peripheral hormones, especially glucocorticoids, maintain potent feedback inhibition control of corticotroph secretion and replication.

Measurement of Adrenocorticotropic Hormone

Both RIA and IRMA employ antisera specifically directed against intact ACTH (1 to 39) or other POMC fragments. Generally, the IRMA is more sensitive, reproducible, and rapid. Most IRMAs have a sensitivity of less than 0.5 ng/L with precise variations of less than 10%. Intact ACTH or POMC precursor peptides are detectable, depending on the sequence specificity of the assay employed. Awareness of the peptide specificity may be especially critical when evaluating ectopic POMC products secreted by lung tumors. ACTH precursors are assessed by a specific IRMA employing unique monoclonal antibodies to ACTH, N-POMC, -lipotropin (-LPH), or -endorphin.

Extrapancreatic Tumor Adrenocorticotropic Hormone Synthesis

Tissue POMC is also expressed in the gonads, lung, gastrointestinal and adrenal medulary neuroendocrine cells, and white blood cells. Nevertheless, most circulating ACTH is derived from the anterior pituitary or from neuroendocrine tumor ectopic production. Extrapancreatic neuroendocrine tumors associated with ectopic ACTH secretion do not process the prohormone efficiently. As ACTH is synthesized in non-tumorous neuroendocrine cells, "ectopic" tumor hormone production may in fact reflect inappropriate ACTH processing. These patients also exhibit a higher ratio of circulating ACTH precursors as well as smaller peptides, including CLIP.
corticotosteroid-binding globulin (CBG) levels and stress may influence measured cortisol values. Random ACTH values do not provide an accurate assessment of HPA function unless concurrent cortisol levels are obtained. Thus, an integrated assessment of both hormone levels is required for interpreting the significance of an appropriately obtained ACTH value. Often, measurement of cortisol levels alone may provide a useful surrogate endpoint for ACTH action and HPA axis integrity. Plasma ACTH levels fluctuate broadly within the same individual and are highly sensitive to stress, time of collection, and gender. Men exhibit greater ACTH pulse frequency and amplitude and pregnant women have higher ambient ACTH levels, possibly related to placental CRH secretion.

Dynamic Testing for Adrenocorticotropic Hormone Reserve

Hypothalamic

Insulin hypoglycemia is a potent endogenous stressor that evokes ACTH secretion. Thus insulin (0.1 to 0.15 U/kg) is injected intravenously after an overnight fast to achieve symptomatic hypoglycemia and a blood glucose level less than 40 mg/dL. This test correlates well with other indices of ACTH reserve. A normal HPA response to this stressor evokes cortisol levels above 20 µg/dL. As hypoglycemia acts centrally, a normal response implies integrity of all three tiers of HPA axis control. Up to 20% of patients may require insulin up to 0.3 U/kg or more to achieve symptoms of glucopenia including sweating, hunger, palpitations, and tremors.

Vencos are sampled at -15, 0, 15, 30, 45, 60, 90, and 120 minutes for measurement of glucose, ACTH, and cortisol levels. GH can also be measured. After the test, oral glucose should be administered. Intraindividual variations in blood glucose levels attained at a given dose of insulin, fluctuations in central sensitivity to glucose, and activation of catecholamines may lead to difficulties in reproducibility. The test is contraindicated in subjects with a history of seizures, those with active coronary or cerebral ischemia, and in pregnancy. If pronounced adrenal insufficiency is likely, insulin injection may provoke an adrenal crisis because of inadequate adrenal reserve, and hydrocortisone (100 mg) should be available for urgent intravenous use if required.

Metyrapone blocks cortisol synthesis by inhibiting adrenal 11-hydroxylase. Thus, the drug releases the HPA axis from negative feedback by cortisol, normally resulting in an ACTH surge and elevated levels of 11-deoxycortisol (compound 5). A single oral dose (2 to 3 g) is given at midnight, and serum levels of ACTH, 11-deoxycortisol, and cortisol are measured at 8 am the next morning. The test is valid only in the presence of documented suppressed cortisol levels less than 10 µg/dL. In normal subjects, peak ACTH values higher than 200 ng/L are achieved. Side effects include nausea, gastrointestinal upset, and insomnia.

False-positive results may be obtained when phenytoin is being administered because the drug prevents adequate enzymatic blockade. This test should be performed under observation in hospital because acute adrenal insufficiency may ensue.

Pituitary Stimulation

Pituitary ACTH secretion may be evoked by injecting either CRH or arginine vasopressin. Ovine or human CRH (100 µg or 1 µg/kg) is administered intravenously, and cortisol and ACTH are measured at -5, -1, 0, 15, 30, 60, 90, and 120 minutes. Normally, maximal ACTH responses (twofold to fourfold above baseline) are evoked at 30 minutes, and cortisol levels peak (over 20 µg/dL) at 60 minutes or increase more than 10 µg/dL above baseline. Although CRH readily induces ACTH secretion and demonstrates ACTH deficiency or ACTH excess, the wide variation of responses observed has limited its utility.

A useful application of the CRH test is in making the diagnosis of Cushing's disease with or without dexamethasone pretreatment and in the context of petrosal venous sampling for diagnosing ACTH-secreting pituitary adenoma. CRH injection allows a sensitive and specific ACTH gradient to be established, which effectively distinguishes peripheral from pituitary sources of excessive ACTH secretion. Because of the suppressive impact of circulating glucocorticoids on pituitary CRH responsiveness, it may be difficult to distinguish a corticotroph adenoma from pseudo-Cushing's disease, as hypercortisolism is associated with both conditions. In these circumstances, combining this test with dexamethasone suppression may be useful.

In the combined dexamethasone-CRH test, dexamethasone is administered at 0.5 mg every 6 hours for 48 hours starting at noon and ending at 6 am then CRH is administered intravenously at 8 am. In normal subjects or those with pseudo-Cushing's disorder, cortisol levels do not rise and are less than 1.4 µg/dL. If cortisol levels elicit at 15 minutes exceed 4 µg/dL, an ACTH-secreting pituitary tumor is invariably present with 100% sensitivity and specificity. CRH responsiveness (at least a 35% cortisol rise) is usually retained in ACTH-secreting adenomas but is not apparent in more than 90% of ectopic ACTH-producing tumors, with which ACTH levels elicited at 15 minutes exceed 4 µg/dL. The test is valid only in the presence of documented suppressed cortisol levels less than 10 µg/dL. In normal subjects, peak ACTH values higher than 200 ng/L are achieved. Side effects include nausea, gastrointestinal upset, and insomnia.

False-positive results may be obtained when phenytoin is being administered because the drug prevents adequate enzymatic blockade. This test should be performed under observation in hospital because acute adrenal insufficiency may ensue.

Adrenal Stimulation

The acute response of the adrenal gland to a bolus ACTH injection reflects ambient ACTH concentrations to which the gland has been exposed. Thus, the cortisol response to an acute ACTH injection is blunted if the subject has had chronic pituitary ACTH hyposecretion with resultant adrenal atrophy and diminished cortisol reserve. Conversely, persistently elevated ACTH levels lead to adrenal hyperthrophy and augmented cortisol responses. The utility of this test in diagnosing a diminished pituitary ACTH reserve has been challenged because the commonly employed dose of Cortrosyn (ACTH 1 to 24, 250 µg) or Synacthen is high and may evoke a "normal" cortisol response in hypopituitary subjects. An unacceptably high false-negative rate (about 85%) has been determined in a large series, although peak cortisol levels at 30 minutes do, in fact, correlate well with peak responses to an ITT.

A normal cortisol response is greater than 20 µg/dL or a doubling of baseline values. Basal cortisol levels correlate inversely with the incremental response to ACTH. Low-dose stimulation with 1 µg of Synacthen evoked maximal serum cortisol levels at 30 minutes, and these correlated well with values observed after insulin or high-dose ACTH administration. A cutoff of greater than 500 nmol/L provides almost 100% sensitivity and a specificity of 80% to 100%. Failure to respond to low-dose ACTH should be corroborated by a standard dose insulin or ACTH test stimulation.

Text.

A 250-µg dose of Cortrosyn (ACTH 1 to 24) is injected intramuscularly or intravenously, and cortisol levels are measured before and 30 and 60 minutes after injection. Cortisol values higher than 20 µg/dL reflect a normal adrenal reserve response.

Interpretation.

Fluctuation of CBG levels may confound interpretation of cortisol values. Thus, ciprofloxacin and hyperthyroidism lower CBG and cortisol levels, whereas estrogens elevate CBG concentrations.

Secondary Adrenal Insufficiency

Ideally, nonstressed resting subjects should have venous blood withdrawn between 6 and 9 AM. Because ACTH is relatively unstable at room temperature and has a propensity to adhere to glass, plasma samples should be separated immediately in iced siliconized glass tubes containing ethylenediaminetetraacetic acid (EDTA) and stored below -20°C for transport. The plasma ACTH levels at 8 AM range from 5 to 25 ng/L as measured by IRMA. Episodic secretion and short plasma half-life result in wide and rapid fluctuation of plasma measurements. Cortisol values at 4 AM are about half of morning levels, and at 11 AM levels are usually less than 5 µg/dL. Altered corticosteroid-binding globulin (CBG) levels and stress may influence measured cortisol values.

Random ACTH values do not provide an accurate assessment of HPA function unless concurrent cortisol levels are obtained. Thus, an integrated assessment of both hormone levels is required for interpreting the significance of an appropriately obtained ACTH value. Often, measurement of cortisol levels alone may provide a useful surrogate endpoint for ACTH action and HPA axis integrity. Plasma ACTH levels fluctuate broadly within the same individual and are highly sensitive to stress, time of collection, and gender. Men exhibit greater ACTH pulse frequency and amplitude and pregnant women have higher ambient ACTH levels, possibly related to placental CRH secretion.
ACTH deficiency is usually reflective of already profound pituitary insufficiency with disordered GH, gonadotropin, and TSH reserve. Rarely, isolated ACTH deficiency is manifest later in life, is more common in males, but may occur after childbirth associated with autoimmune thyroiditis and diabetes mellitus. Two families with recessive mutations in the corticotroph-specific TPIT gene have been described with congenital adrenal insufficiency and ACTH deficiency. Insufficient ACTH secretion leads to attenuated adrenal corticosteroid production, with relative mineralocorticoid preservation.

Patients present with slowly progressive weight and appetite loss, anorexia, and generalized fatigue. Because adrenal mineralocorticoid is largely unpaired, salt wasting, volume contraction, and hyperkalemia, commonly encountered features in Addison's disease, are not present. Furthermore, the hyperpigmentation usually associated with exuberant ACTH-related peptide secretion in the presence of adrenal damage does not occur. Morning serum cortisol levels less than 3 µg/dL suggest ACTH deficiency, and basal morning cortisol levels greater than 18 µg/dL usually indicate a normal ACTH reserve. Patients with ACTH deficiency have low to normal serum cortisol levels and low to normal plasma ACTH levels. Blunted responses to provocative tests such as insulin-induced hypoglycemia or metyrapone are required to document a partial deficiency.

**Treatment.**

Hydrocortisone is used to replace deficient glucocorticoid hormone directly. The normal secretory rate of cortisol is about 20 mg/day, which is the recommended total daily dose for correcting hypoadrenalism and maintaining blood pressure. The plasma circulating half-life of cortisol is less than 2 hours, and twice-daily dosing regimens may result in very low cortisol levels in the late afternoon with impaired quality of life. Hydrocortisone dosing three times daily for a total daily requirement of 20 mg (10 mg in the morning, 5 mg at noon, and 5 mg in the evening) is most effective for starting replacement. Although excessive dosing leads to iatrogenic Cushing's syndrome, doses should be increased during stress or prior to operative procedures. Cortisone acetate is metabolized to cortisol and has a slower onset of action and longer biologic activity than hydrocortisone. Other synthetic glucocorticoids, including prednisolone and dexamethasone, are less useful because they are difficult to monitor biochemically. Even modest cortisol overreplacement may result in bone mineral loss.

Mineralocorticoid replacement is rarely required. Central diabetes insipidus may rarely be unmasked after initial glucocorticoid replacement.

**Adrenocorticotropic HormoneSecreting Tumors**

The evaluation and management of Cushing's disease are described fully in Chapter 14. Briefly, the diagnosis of an ACTH-secreting pituitary tumor is suggested by features of hypercortisolism, elevated 24-hour urinary free cortisol levels, and failure to suppress morning cortisol levels to less than 3 µg/dL after 1 mg of dexamethasone administered at 11 PM. In healthy subjects, glucocorticoid feedback suppresses CRH and ACTH, attenuating cortisol secretion.

Surgical resection of an ACTH-secreting adenoma is the treatment of choice. Because these tumors are usually small, sometimes less than 2 mm in diameter, they may be localized incorrectly, or not at all, by venous sampling for ACTH (see earlier) and sensitive MRI. Therefore, these tumors pose a significant challenge even for the experienced surgeon. Furthermore, the disorder is also characterized by venous hypertension leading to turgid venous sinuses requiring control by the anesthetist.

Bilateral petrosal venous sampling for ACTH levels and cavernous sinus venography should ideally be performed before surgery. However, if sellar venous sinus drainage is predominantly unilateral, left-right ACTH gradients may not reliably lateralize the lesion. Cavernous sinus venography may also outline a filling defect representing the tumor. If an ACTH gradient is indeed detected with normal venous drainage patterns, hemihypophysectomy may be curative in 80% of such patients with clearly defined biochemical features of ACTH-dependent Cushing's disease. Metliculous surgical exploration of both anterior and posterior lobes is required for these tiny tumors, which are often off-white and speckled by petechiae and may be inadvertently suctioned. Unfortunately, even carefully performed preoperative lateralization is not infallible and the so-called normal side should also be carefully explored.

**Assessment of Surgical Outcome**

Transsphenoidal adenoma resection is the preferred treatment for these adenomas. After selective adenomectomy of a clearly identifiable adenoma, remission was achieved in 75% of 295 patients. However, partial hypophysectomy performed in 31 patients in whom an adenoma could not be identified resulted in biochemical remission in only 10 patients. On the third postoperative day, 1 mg of dexamethasone can be given at 10 PM and cortisol levels measured the following morning, prior to initiating hydrocortisone therapy. If the immediate postoperative cortisol level is less than 3 µg/dL, a 95% 5-year remission rate can be expected. In 21 of 27 patients tested prior to glucocorticoid administration, postoperative cortisol levels below 10 mg/dL or less than those obtained with preoperative midnight sampling were predictive of remission.

**Silent Corticotroph Adenoma**

These basophilic tumors are generally nonfunctional and yet exhibit POMC, -lipotropin, and -endorphin immunoreactivity. ACTH secretion is apparently unaltered, with no associated clinical or biochemical features of hypercortisolism, although these tumors are morphologically indistinguishable from adenomas associated with Cushing's disease. They may represent up to 7% of all surgically removed adenomas and are usually hemorrhagic and invariably macroadenomas. Unlike Cushing's disease, they have a 2:1 male preponderance, often occur with mass effects, and about one third have preoperative evidence of pituitary insufficiency. About half exhibit cavernous sinus or bone invasion, hemorrhage, necrosis, and cyst formation. These tumors often recur, and postoperative radiation and reoperation are required to eradicate tumor regrowth or residual mass. Unless appropriate immunostaining is performed, many of these tumors remain undiagnosed and are classified as recurrent nonfunctioning macroadenomas.
Thyroid-Stimulating Hormone

Thyrotroph cells constitute about 5% of the functional anterior pituitary cells and are situated predominantly in the anteromedial areas of the gland. They are smaller than the other cell types and are irregularly shaped with flattened nuclei and relatively small secretory granules ranging from 120 to 150 µm (Fig. 8-61 (Figure Not Available) and Fig. 8-62 (Figure Not Available)).

Figure 8-61 (Figure Not Available) Normal thyrotrophs have angular cell bodies with elongated processes. (From Asa SL. In Tumors of the Pituitary Gland. Atlas of Tumor Pathology. Washington, DC, Armed Forces Institute of Pathology, 1997, p 19.)

Figure 8-62 (Figure Not Available) Electron micrograph of normal thyrotrophs showing angular cell bodies with elongated processes. (From Asa SL. In Tumors of the Pituitary Gland. Atlas of Tumor Pathology. Washington, DC, Armed Forces Institute of Pathology, 1997, p 19.)

Thyrotroph cells constitute about 5% of the functional anterior pituitary cells and are situated predominantly in the anteromedial areas of the gland. They are smaller than the other cell types and are irregularly shaped with flattened nuclei and relatively small secretory granules ranging from 120 to 150 µm (Fig. 8-61 (Figure Not Available) and Fig. 8-62 (Figure Not Available)).

Thyroid-Stimulating Hormone Biosynthesis

TSH is a glycoprotein hormone that is a heterodimer of two noncovalently linked and subunits. The subunit is common to TSH, LH, FSH, and hCG, but the subunit is unique and confers specificity of action. The subunit is the earliest hormone gene expressed embryonically; activation of the subunit gene occurs later under the influence of GATA-2 and Pit-1. The 13.5-kb subunit gene is located on chromosome 6 and comprises four exons and three introns. Although the subunit gene is expressed in thyrotroph, gonadotroph, and placental cells, its regulation is uniquely cell-specific. The downstream promoter region (-200 and below) is required for placental expression, intermediate sequences are required for gonadotroph expression, and upstream promoter elements are required for thyrotroph-specific expression. The subunit transcript is inhibited by T3 at regions close to the transcriptional initiation site, in concert with other nuclear corepressors. The 4.9-kb TSH subunit gene located on chromosome 1 comprises three exons and two introns. Pit-1 binds directly to the gene promoter to confer tissue-specific expression. TSH gene transcription is suppressed by the thyroid hormone receptor acting directly on exon 1.

This potent suppression is evident within 30 minutes of T3 exposure and is a critical determinant of TSH synthesis and ultimate secretion. Transcription of both TSH subunit genes is induced by TRH, and depletion of CAMP by dopamine leads to suppressed gene transcription. Intrapituitary TSH is stored in secretory granules, and the mature hormone (28 kd) is released into the venous circulation primarily in response to hypothalamic TRH. The predicted structural model of the TSH molecule is that of a cystine knot growth factor. The tertiary TSH structure comprises three hairpin loops separated by central disulfide bonds, with the longer loop straddling one side. Production of the mature heterodimeric TSH molecule requires complex cotranslational glycosylation and folding of nascent and subunits. After subunit translation and signal peptide cleavage, glycosylation occurs at asparagine 23 on the subunit and at two asparagine residues, 52 and 78, on the subunit. Appropriate glycosylation is required for accurate molecular folding and subsequent combination of and subunits within the rough endoplasmic reticulum and Golgi apparatus. Both TRH and T3 regulate TSH glycosylation, albeit in opposite directions. TRH administration or T3 deprivation, such as occurs in hypothyroidism or T3 resistance, enhances oligosaccharide addition to the TSH molecule.

Thyroid-Stimulating Hormone Secretion

The TSH production rate is normally 100 to 400 mU/day, with a calculated circulating half-life of about 50 minutes. Secretion rates are enhanced up to 15-fold in hypothyroid subjects and are suppressed in states of hyperthyroidism. The degree of TSH glycosylation determines the metabolic clearance rate as well as bioactivity, and in hypothyroidism the molecule appears highly sialylated. Immunoactive fetal pituitary TSH is detectable by 12 weeks. Immediately after full-term birth, there is a brisk rise in TSH, which remains elevated for up to 5 days before stabilizing at adult levels.

Although TSH secretion is pulsatile, the low pulse amplitudes and long TSH half-life result in modest circulating variances. Secretory pulses every 2 to 3 hours are interspersed with periods of tonic, nonpulsatile TSH secretion. Circadian TSH secretion peaks between 11 PM and 5 AM, mainly because of increased pulse amplitude, which does not appear to be sleep-entrained. Pulsatile and circadian TSH secretory patterns are largely determined by ambient thyroid hormone levels, TRH release, dopamine, and cortisol. Primary hypothyroidism is associated with enhanced TSH pulse amplitudes occurring throughout the day, and nocturnal TSH surges are abrogated in patients with critical illness.

Thyrotropin-Releasing Hormone and Thyroid Hormone Regulation

Because feedback control of TSH secretion by peripheral thyroid hormones is so sensitive, most thyrotroph disorders can be diagnosed by measuring basal TSH and thyroid hormone levels. However, evoked dynamic TSH measurements may be required to assess fully the integrity of the hypothalamic-pituitary-thyroid axis. TRH (200 to 500 µg) is administered intravenously, and TSH levels are measured at -15, 0, 15, 30, 60, and 120 minutes.

In euthyroid subjects, peak TSH levels (up to 22-fold higher than basal) are observed after 30 minutes. Because feedback suppression of TSH by elevated thyroid hormone levels overrides positive hypothalamic signals, hyperthyroid subjects have undetectable basal TSH levels that do not respond to TRH. In subjects with primary thyroid failure, the TSH response is exuberant, but in those with secondary thyroid failure related to pituitary disease TSH levels do not change in response to TRH.

Sustained TRH infusions for up to 4 hours result in biphasic TSH increases, reflecting early release of preformed TSH, followed later by newly synthesized hormone. Further prolonged TRH infusions elevate thyroid hormone levels, which subsequently suppress pituitary TSH synthesis and release. Within hours of T3 administration, basal TSH levels are suppressed and TRH-evoked TSH levels are attenuated. Thyroid hormones suppress tonic TSH secretion and pulse amplitude but do not appear to regulate TSH pulse frequency. T3 also suppresses hypothalamic TRH synthesis and decreases pituitary TRH receptor number, thus further limiting TSH biosynthesis.

Other Factors

SRIF inhibits TSH pulse amplitude, blocks the nocturnal TSH surge directly at the pituitary level, and may also suppress TRH release and possibly TRH receptor abundance. Although SRIF analogues are used to treat TSH-secreting pituitary adenomas (see later), long-term SRIF treatment for acromegaly does not lead to hypothyroidism in adult subjects although T3 levels may be lowered within the normal range. Dopamine inhibits TSH subunit gene expression, and dopamine infusions suppress TSH pulse amplitude by 70% and abrogate the nocturnal TSH surge. Prolonged use of dopamine agonists, however, does not result in hypothyroidism.

Glucocorticoids suppress TSH secretion, and in patients with adrenal failure without autoimmune thyroid damage, TSH levels may be elevated. Sex steroids and cytokines alter TSH secretion in animal models, but their contribution to human TSH physiology is unclear. Nonsteroidal anti-inflammatory agents, especially...
meclomenamate and fenclofenac, decrease serum TSH levels, albeit still within the normal range. The mechanism may involve displacement of thyroid hormone ligands from their binding proteins or direct inhibition of pituitary TSH.

Thyroid-Stimulating Hormone Action

TSH acts on the thyroid gland to induce thyroid hormone synthesis and release and to maintain trophic thyroid cell integrity. The TSH G protein-coupled (GPC) receptor is located on the thyrocyte plasma membrane and is encoded by a gene on chromosome 11q31. Its regulation is comprehensively described in Chapter 10.

Thyroid-Stimulating Hormone Assays

The challenge for a clinically compelling robust TSH assay is to differentiate circulating TSH levels in euthyroid subjects from those in both hyperthyroid and hypothyroid patients. The development of immunoradiometric TSH assays has provided high specificity with little or no cross-reactivity with other glycoprotein hormones. These assays detect quantifiable TSH levels in euthyroid control subjects with no overlap with the low values associated with hyperthyroidism. The most sensitive commercially available third-generation assays have a functional sensitivity of 0.01 to 0.02 mIU/L, and newer fourth-generation assays should have greatly enhanced sensitivity (0.001 to 0.002 mIU/L). Levels of free subunit (normal range 0.1 to 1.6 μg/L) are elevated in patients harboring TSH-secreting or nonfunctional pituitary adenomas, choriocarcinomas, and several malignancies.

TSH deficiency results in childhood mental or growth retardation, or both, and hypothyroidism in adults is associated with a broad spectrum of clinical features including hypothemia, fluid retention, voice and skin changes, and ultimately frank myxedema and death. Pituitary damage may result in functional TSH deficiency, often without a clearly demonstrable reduction in serum TSH levels. Although impractical to measure, nocturnal TSH pulse amplitudes may be attenuated in patients with pituitary dysfunction. TSH deficiency should be diagnosed by measuring free T levels because TSH measurements are not helpful in diagnosing central hypothyroidism. In fact, only about one third of patients with secondary hypothyroidism have abnormally low basal TSH levels. TSH deficiency is thus associated with low T levels concomitant with low, normal, or even minimally elevated TSH levels. This biochemical profile may also be encountered in critically ill patients with low TSH and T levels without evidence of pituitary disease.

Treatment

L-Thyroxine is used for replacement therapy, and dosing variables are similar to those required for treating primary hypothyroidism. Hypothyroid features are effectively ameliorated by T (0.05 to 0.25 mg/day). The molecule is converted peripherally into the active T and has a 7-day half-life with stable blood levels. The dose of levothyroxine in hypopituitary patients is titrated to achieve midnormal clinically euthyroid serum free T levels because serum TSH levels are low or undetectable in patients with damaged pituitary function.

Measurement of TSH levels is not useful in determining thyroid hormone replacement because the damaged thyrotroph is unlikely to reflect appropriate feedback suppression. Many women with pituitary failure also receive estrogen replacement, and measuring free T levels is required because of increased TBG levels. TSH overdosing may also lead to osteopenia and cardiac arrhythmias. Some patients may have associated ACTH deficiency, and thyroid hormone replacement should not be initiated until adrenal reserve has been evaluated and, if necessary, treated. Thyroid hormone replacement may also accelerate cortisol metabolism or requirements, or both, and may therefore exacerbate primary hypoadrenalism or precipitate adrenal crisis in patients with perturbed adrenal function.

Thyrotropin-Secreting Tumors

TSH-producing pituitary tumors are rare. Most older series indicate that they represent less than 1% of pituitary tumors. From 1979 to 1992, Mindermann and Wilson analyzed tumor type by immunohistochemistry and found that the overall prevalence of TSH-secreting tumors was 19 in

2225 (0.85%). Between 1989 and 1991, they found a prevalence of 2.8%. It is not clear whether the incidence of this tumor type is increasing or whether tumors are now more readily recognized. Enhanced recognition may be a result of the development of high-sensitivity TSH assays that distinguish between normal TSH levels that are, in fact, inappropriately elevated for thyroid hormone levels in some patients with TSH-producing tumors and frankly low ones. TSH-secreting tumors can also cosecrete other hormones including GH, PRL, and rarely ACTH.

Pathology

These tumors are invasive, but for the most part benign, and distant metastases are extremely rare. The secretory pattern is determined by a panel of antibodies to TSH, subunit, GH, PRL, and ACTH. TSH-secreting tumors exhibit positive immunostaining for subunit and TSH in 20% to 75% of cells and for Pit-1.

Presentation

Patients with TSH-secreting tumors present with symptoms related to tumor size (e.g., visual field abnormalities, cranial nerve palsies or headache) or to hormone overproduction. Signs and symptoms of hyperthyroidism, including palpitations, arrhythmias, weight loss, tremor, and nervousness, or a goiter are common. A case of periodic paralysis has been reported. Serum TSH is often but not invariably elevated, and the combination of abnormally high thyroid hormone and TSH within the normal range points to a TSH-producing pituitary tumor. A relatively long period of hyperthyroidism, initially thought to represent Graves’ disease and treated accordingly, often predates the realization that the hyperthyroidism is a result of a TSH-secreting pituitary tumor. Alternatively, thyroid hormone insensitivity can exist with similar laboratory profiles.

TSH-secreting tumors are usually large. A review of six reports indicates that 80% of TSH-secreting tumors are macroadenomas and 12% microadenomas. Over 60% are also locally invasive. From an analysis of 10 reports on a total of 153 patients, we estimate that TSH is frankly elevated in 58% of patients, with the remainder having normal although inappropriately elevated levels. Patients previously treated with radioactive iodine for presumed Graves’ disease present with significantly higher TSH levels than patients not previously radioablated (mean of 56 and 9 mIU/L, respectively). An ecotropic TSH-producing tumor has also been reported. Serum TSH is high in the majority of patients, as is the glycoprotein hormone subunit. Approximately two thirds of patients with TSH-producing pituitary tumors have a goiter with elevated radioactive iodine uptakes. Signs or symptoms of acromegaly or hyperprolactinemia may also be present in controls.

Evaluation

The T, T, TSH (by high-sensitivity assay), and subunit should be measured. The combination of high T, T, and subunit; high or inappropriately normal TSH; and a pituitary tumor strongly confirms the diagnosis of a TSH-producing pituitary adenoma. TRH stimulation distinguishes between TSH overproduction by a TSH-secreting tumor and thyroid hormone insensitivity. With TSH-secreting tumors, the TSH response to TRH is blunted. In contrast, TSH usually rises in response to TRH in thyroid hormone insensitivity and in normal subjects. Concomitant measurement of subunit at each point during the TRH test is helpful because the molar ratio of subunit to TRH is high (>1) in almost 85% of patients with TSH-secreting tumors. Ratios greater than 1 can also be seen in normal subjects.

A T suppression test is helpful in that complete inhibition of TSH does not occur in patients with TSH-secreting tumors. This test can also differentiate subclinical hyperthyroidism in a patient treated with radioactive iodine for hyperthyroidism in the past but found to have an incidental pituitary tumor. TSH elevation may also result from inadequate thyroid hormone replacement. Pituitary MRI should be performed and IGF-I and PRL levels determined to exclude acromegaly or hyperprolactinemia. The presence of other pituitary hormones in immunostained histologic sections does not necessarily imply that their levels are elevated.

The degree of hyperthyroidism should be assessed to determine whether control of these signs and symptoms should be undertaken prior to further evaluation or treatment of the pituitary tumor. One report characterized the hyperthyroidism in this condition as being severe in 14 of 25 patients and having been present in most patients for years before the diagnosis was made. Perioperative deaths in patients with TSH-secreting tumors have been reported, which might be attributed to

251
poorly controlled hyperthyroidism.

Management

Surgery

Surgery is recommended as the first-line treatment, but surgical cures occur in no more than 40% of patients (Table 8-30). However, the rarity of this tumor type has precluded large controlled studies. Fourteen of 22 patients had cavernous or sphenoid sinus invasion and tumors were fibrous and unusually hard. Eight patients were considered cured after surgery. In another study, surgery resulted in normalization of T₄ in 15 and of parameters of cure in 7 of 17 patients. Over half of the patients, when assessed by MRI 6 months after surgery, exhibited evidence of residual tumor.

Radiation Therapy

There are no large series reporting treatment of TSH-secreting tumors with radiotherapy alone. Radiation has mostly been employed as adjunctive therapy to surgery, especially when the latter was not curative.

Somatostatin Analogues

Octreotide, used as either primary or adjunctive treatment, normalized T₄ and T₃ and reduced TSH levels by half in 25 patients treated with octreotide (100 to 500 µg/day) for up to 61 months, 84% had controlled thyroid function. However, tachyphylaxis developed in five patients, and three escaped the effect of the drug. Overall, tumor shrinkage occurs in about a third of patients. In 18 patients with TSH-secreting adenomas, lanreotide (30 mg every 10 or 14 days), significantly decreased TSH levels from 2.72 to 1.89 mU/L, decreased T₄ levels, but did not shrink tumors. Responsiveness to octreotide LAR (up to 30 mg monthly) appeared similar to that observed for the subcutaneous preparation in seven patients.

Unless vision is threatened, patients should be evaluated to determine whether the clinical signs of hyperthyroidism warrant immediate treatment (Fig. 8-63). Propranolol, radioactive iodine thyroid ablation, thyroidectomy, antithyroid medications including methimazole (Tapazole) and propylthiouracil, and somatostatin analogues are employed. Both radioactive iodine and antithyroid medications are targeted to the thyroid gland rather than the pituitary seat of the disorder. Both also

<table>
<thead>
<tr>
<th>Source</th>
<th>Number of Patients</th>
<th>Micro-adenomas</th>
<th>Macro-adenomas</th>
<th>Extrasellar Extension</th>
<th>Visual Field Deficit</th>
<th>Cured by Surgery</th>
<th>Immunostain</th>
<th>Radiation</th>
<th>Cured by Radiation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grisoli, 1987 (776b)</td>
<td>6</td>
<td>0/6</td>
<td>6/6</td>
<td>4/6</td>
<td>3/6</td>
<td>2/5</td>
<td>4/4 TSH</td>
<td>3/6</td>
<td>1/3</td>
</tr>
<tr>
<td>Gesundheit, 1989 (776a)</td>
<td>9</td>
<td>2/9</td>
<td>7/9</td>
<td>3/5*</td>
<td>2/7 TSH</td>
<td>5/7 TSH</td>
<td>3/8</td>
<td>0/3</td>
<td></td>
</tr>
<tr>
<td>McCutcheon, 1990 (779)</td>
<td>8</td>
<td>1/8</td>
<td>7/8</td>
<td>6/7</td>
<td>4/8</td>
<td>4/8</td>
<td>6/7 TSH</td>
<td>3/8</td>
<td>0/3</td>
</tr>
<tr>
<td>Beckers, 1991 (788)</td>
<td>7</td>
<td>1/7</td>
<td>6/7</td>
<td>3/4</td>
<td>2/7 pure TSH</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chanson, 1992 (789a)</td>
<td>37</td>
<td>2/37</td>
<td>35/37</td>
<td></td>
<td></td>
<td>8/37</td>
<td></td>
<td>0/8</td>
<td></td>
</tr>
<tr>
<td>Chanson, 1993 (789)</td>
<td>52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9/52</td>
<td></td>
<td>0/9</td>
<td></td>
</tr>
<tr>
<td>Mindermann, 1993 (777)</td>
<td>19</td>
<td>0/19</td>
<td>19/19</td>
<td>12/19</td>
<td>6/19</td>
<td>NA</td>
<td>6/14 pure TSH</td>
<td>9/19 E</td>
<td>NA</td>
</tr>
</tbody>
</table>

ACTH, adrenocorticotropic hormone; GH, growth hormone; PRL, prolactin; TSH, thyrotropin.
inhibit the remaining negative feedback of T₃ on TSH and lead to increased tumor TSH production. 

Surgery and somatostatin analogues simultaneously treat hyperthyroidism and tumor TSH hypersecretion. Propranolol is important for inhibiting peripheral hyperthyroid manifestations. Somatostatin analogues lower TSH, subunit, and T₄ and are recommended as first-line drugs in the initial control of hyperthyroidism related to TSH-secreting tumors because they act more rapidly than other therapeutic approaches and tumor shrinkage occurs in up to 40% of patients. If these drugs are ineffective or only partially effective, other therapeutic modalities should be used. Thus, no single treatment is expected to cure patients with TSH-secreting adenomas. Surgery is curative in only a minority of patients, and although tumor bulk removal may normalize thyroid function when invasive tumor tissue persists, patients continue to have abnormal TSH responses to TRH and require somatostatin analogue therapy.
PITUITARY FAILURE

Impaired synthesis of one or more anterior pituitary hormones may result from heritable genetic factors, acquired anatomic insults, inflammation, or vascular damage. Because of its close anatomic contiguity, impaired hypothalamic hormone synthesis or secretion may also occur as a component of the pituitary gland insult, especially after external radiation. Furthermore, distinct hypothalamic lesions may result in diminished pituitary hormone secretion by abrogating hypothalamic hypophysiotropic signals.

Developmental and Genetic Causes of Pituitary Failure

Developmental Pituitary Dysfunction

Congenital pituitary gland absence (aplasia), partial hypoplasia, or ectopic tissue rudiments are rarely encountered. Pituitary development follows midline cell migration from Rathke’s pouch, and impaired midline anomalies including failed forebrain cleavage and anterior commissure and corpus callosum defects lead to structural pituitary anomalies. Craniofacial developmental anomalies including anencephaly result in cleft lip and palate, basal encephalocele, hypertelorism, and optic nerve hypoplasia with varying degrees of pituitary dysplasia and aplasia. If these infants survive, appropriate lifelong pituitary hormone replacement is required. Children with mild forms of midline anomalies are also more susceptible to GH deficiencies.

With sensitive MRI techniques for pituitary visualization, several anatomic features characteristic of hypopituitarism are now apparent. Evidence for acquired pituitary gland damage or destruction is often clearly visible on MRI, and patients presenting with hypopituitarism of undetermined etiology may exhibit decreased gland volume, partial or complete empty sella, disturbed sella turcica architecture, absent or transected pituitary stalk, and an absent or ectopic posterior pituitary bright intensity signal. An absent infundibulum noted on MRI is associated with pituitary hormone deficits, and about 25% of patients with GHD of unclear cause show imaging evidence of mild stalk defects, reflecting a midline developmental anomaly. Congenital basal encephalocele may result in the pituitary herniating through the sphenoid sinus roof, resulting in pituitary failure and diabetes insipidus.

Heritable Disorders of Pituitary Failure

Mutations of transcription factors that determine anterior pituitary development may lead to pituitary deficiency syndromes. Patients heretofore diagnosed with idiopathic isolated or polyhormonal pituitary failure may, in fact, harbor such a mutation. As the genetic control of pituitary development has been clarified, increasing numbers of mutant genes have become apparent.

PROP1

PROP1 (Online Mendelian Inheritance in Man [OMIM] 601538) gene expression is required for subsequent Pit-1 activation. The gene, located on chromosome 5q, encodes a 223-amino-acid protein expressed in cells secreting GH, PRL, and TSH. The Ames dwarf mouse harbors a missense PROP1 mutation (Ser83Pro) and exhibits a hypoplastic pituitary gland with combined GH, PRL, and TSH deficiency. This mutation abrogates Pit-1 activation and results in failed development of Pit-1-dependent cell lineages. At least eight human mutations have been associated with GH, PRL, TSH, and gonadotropin deficiencies. The most commonly encountered mutation is a 2-bp deletion at position 296 (301-302delAG), resulting in early translation termination and a nonfunctional protein.
Hormone mutation

<table>
<thead>
<tr>
<th>Gene</th>
<th>Hormones</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pit-1 (POU1F1)</td>
<td>PRL, GH, TSH</td>
</tr>
<tr>
<td>HESX1</td>
<td>GH, PRL, TSH, LH, FSH, ACTH</td>
</tr>
<tr>
<td>LHX3</td>
<td>GH, PRL, TSH, LH, FSH</td>
</tr>
<tr>
<td>DAX1</td>
<td>Adrenal, LH, FSH</td>
</tr>
</tbody>
</table>

ACTH, adrenocorticotropic hormone; FSH, follicle-stimulating hormone; GH, growth hormone. Functional defects include missense or frameshifts leading to truncated or deleted protein, DNA binding abnormality, inactivated protein, or impaired coactivation. *POU1F1* mutations result in varying phenotype of early growth failure with or without hypothyroidism. *PROP1* mutations may be fully manifest only in adulthood. *HESX1* is critical for corpus development and associated with structural brain defects.

protein product. The clinical spectrum of combined pituitary hormone deficiency associated with *PROP1* mutations varies with both the type of mutation and the age of the patient. 798

Human *PROP1* mutations are associated with deficiencies in Pit-1-dependent lineages (GH, PRL, and TSH) as well as impaired FSH, LH, and ACTH reserve function. 517 Because the development and mature function of the latter cell types are not Pit-1-dependent, it appears that additional critical developmental factors are disrupted in these patients, leading to the clinical phenotype. Over 50 patients with *PROP1* mutations leading to combined pituitary hormone deficiency have been described since the original report in 1996, 799 and this disorder appears to be the most common heritable cause of combined pituitary hormone deficiency.

**Molecular Analysis**

Modes of inheritance of *PROP1* mutations usually reflect autosomal recessive patterns. Thus, patients are usually homozygous for either deletion or missense frameshift mutations leading to truncated *PROP1* protein products devoid of functional activity. *PROP1* has been identified at a GA repeat in exon 2. Combination of a GA or AG deletion in this repeat results in a coding frameshift and premature termination at codon 109. Nonafflicted siblings are heterozygous or bear a normal *PROP1* sequence on both alleles.

**TABLE 8-32** — Etiology of Acquired Pituitary Insufficiency

<table>
<thead>
<tr>
<th>Category</th>
<th>Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Traumatic</strong></td>
<td>Surgical resection, Radiation damage, Head trauma</td>
</tr>
<tr>
<td><strong>Infiltrative or inflammatory</strong></td>
<td>Primary hypophysitis, Lymphocytic, Granulomatous, Xanthomatous, Secondary hypophysitis, Sarcoidosis, Histioctytosis X, Infections, Wegener’s granulomatosis, Takayasu’s disease, Hemochromatosis</td>
</tr>
<tr>
<td><strong>Infections</strong></td>
<td>Tuberculosis, Pneumocystis carinii, Fungal (histoplasmosis, aspergillosis), Parasites (toxoplasmosis), Viral (cytomegalovirus)</td>
</tr>
<tr>
<td><strong>Vascular</strong></td>
<td>Pregnancy-related, Aneurysm, Apoplexy, Diabetes, Hypotension, Arteritis, Sickle cell disease</td>
</tr>
<tr>
<td><strong>Neoplastic</strong></td>
<td>Pituitary adenoma, Parasellar mass, Rathke’s cyst, Dermoid cyst</td>
</tr>
</tbody>
</table>
Meningioma
Germinoma
Ependymoma
Glioma
Craniopharyngioma
Hypothalamic hamartoma, gangliocytoma
Pituitary metastatic deposits
Hematologic malignancy
Leukemia
Lymphoma

**Functional**
-- Nutritional
-- Caloric restriction
-- Malnutrition
-- Excessive exercise
-- Critical illness
-- Acute illness
-- Chronic renal failure
-- Chronic liver failure

**Hormonal**
-- Hyperprolactinemia
-- Hypothyroidism
-- After treatment of Cushing's disease

**Drugs**
-- Anabolic steroids
-- Glucocorticoid excess
-- Gonadotropin-releasing hormone agonists
-- Estrogen
-- Dopamine
-- Somatostatin analogue
-- Thyroid hormone excess

**Causes of acquired growth hormone deficiency in 1034 hypopituitary adult patients:**

<table>
<thead>
<tr>
<th>Cause</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pituitary tumor</td>
<td>53.9</td>
</tr>
<tr>
<td>Craniopharyngioma</td>
<td>12.3</td>
</tr>
<tr>
<td>Idiopathic</td>
<td>10.2</td>
</tr>
<tr>
<td>Central nervous system tumor</td>
<td>4.4</td>
</tr>
<tr>
<td>Empty sella syndrome</td>
<td>4.2</td>
</tr>
<tr>
<td>Sheehan's syndrome</td>
<td>3.1</td>
</tr>
<tr>
<td>Head trauma</td>
<td>2.4</td>
</tr>
<tr>
<td>Hypophysitis</td>
<td>1.6</td>
</tr>
<tr>
<td>Surgery other than for pituitary treatment</td>
<td>1.5</td>
</tr>
<tr>
<td>Granulomatous diseases</td>
<td>1.3</td>
</tr>
<tr>
<td>Irradiation other than for pituitary treatment</td>
<td>1.1</td>
</tr>
<tr>
<td>Central nervous system malformation</td>
<td>1.0</td>
</tr>
<tr>
<td>Perinatal trauma or infection</td>
<td>0.5</td>
</tr>
<tr>
<td>Other</td>
<td>2.5</td>
</tr>
</tbody>
</table>


**Clinical Features**

The frequency of PRO1 gene mutations in patients with combined pituitary hormone deficiency is high, the mutations occurring in approximately 50% of affected subjects. However, in families with multiple affected subjects, PRO1 mutations account for virtually all affected individuals. Patients harboring PRO1 mutations exhibit a predominantly hypogonadal phenotype. Puberty is often delayed or absent, with markedly attenuated LH and FSH responses to GnRH stimulation. Some patients enter puberty spontaneously and develop subsequent...
features of central hypogonadism, akin to an acquired presentation. Although the pituitary gland is small or normal in size, patients have been described with grossly hypoplastic anterior pituitary glands with cystic changes and development of a secondary empty sella. Slowing of linear growth usually becomes apparent after the age of 3 years, and these patients usually do not enter puberty. Height SDs may be severely impaired and may range to -10 with eunuchoidal proportions and reduced upper body/lower body ratios. Affected adults are short and have infanteile external genitilia. The onset of clinically evident pituitary failure is usually characterized by GHD (80%) and thyroid failure (TSH deficiency, 20%), followed by hypogonadism and later subclinical or overt adrenal insufficiency.

Features of PROP1 excess have been described in a mouse model. These animals have hypothyroidism, hypogonadism, and persistent Rathke's cleft cysts and, after 1 year, develop pituitary adenomas.

**Evaluation**

Combined hypothalamic hormone stimulation (GnRH, TRH, CRH, and GHRH) or insulin-evoked hypoglycemia reveals blunted responses consistent with varying degrees of pituitary hormone deficiencies. Serum IGF-I and IGFBP3 levels are usually low, and peripheral thyroid hormone levels are low or at the lower limits of normal ranges. In the presence of low or absent TSH responses, these findings are consistent with secondary hypothyroidism. Most older patients also exhibit blunted cortisol responses to CRH or ACTH or insulin stimulation.

**Pituitary Size**

Heterogeneous changes in pituitary size may reflect a combination of adaptive signals for the Pit-1 lineage, compensatory cystic expansion of nonaffected pituitary cell types with subsequent autolaminction, and the absence of other unknown factors required for mature pituitary function.

**POU1F1**

The POU1F1 gene (Pit-1) is located on chromosome 3p11 (OMIM 173110) and encodes a 290-amino-acid protein. The N-terminal POU-specific domain activates gene transcription, and two DNA-binding domains recognize a TATNCAT consensus sequence present in the GH, PRL, and TSH genes and the GHRH receptor gene. The Pit-1 nuclear protein activates transcription of the GH, PRL, and TSH genes and the GHRH receptor gene and also interacts with coactivators including thyroid hormone, estrogen, and retinoic acid receptors as well as other transcription factors including CREB, P-Lim, Ptx-1, HESX-1, and Zn-15. Pit-1 autoregulates its own expression and is therefore critical for maintaining appropriate Pit-1 expression. Because of the absolute requirement of Pit-1 for GH, PRL, and TSH cell development and specific gene expression, inactivating mutations of the gene result in a spectrum of pituitary hormone deficiencies. Two dwarf mouse strains harboring POU1F1 gene mutations. The Snell dwarf mouse harbors a tryptophan cystine missense mutation (Trp261Cys). The Jackson mouse, also a dwarf, harbors a truncated POU1F1 protein with defective DNA binding.

Several POU1F1 mutations have been described, each of which is associated with a characteristic clinical phenotype. Arg172Tyr mutants are associated with neonatal hypothyroidism and GH and PRL deficiency. Both sporadic patients and multiplex families with multiple pituitary hormone defects have been described, and at least 10 recessive and 3 dominant Pit-1 mutations have been identified so far. Recessive mutations result in a varied spectrum of loss of DNA binding or transcriptional activation of TSH, GH, and PRL. Impaired retinoic acid activation of the Pit-1 distal enhancer has been described for a Pit-1 Lys261Glu mutation. This mutant protein also behaves as a dominant negative inhibitor of Pit-1 activation. A sporadic mutation (Arg271Try) results in a protein that binds to DNA but dominantly inhibits Pit-1 transcription with severe combined GH, PRL, and TSH deficiencies.

**Hesx1**

Hesx1 (Rpx) is an early transcriptional marker of the primitive pituitary, with expression restricted to Rathke's pouch. Coincidentally with the appearance of specific pituitary cell types, Hesx1 expression declines and is extinguished in the mature anterior pituitary.

The gene is located on chromosome 3p212, encodes a 1214-amino-acid protein, and competes with PROP1 protein for DNA binding. The heterogeneous syndrome of septo-optic dysplasia (hypoplastic optic nerves, absent corpus callosum and septum pellucidum, and hypopituitarism) is associated with a homozygous Arg53Cys homeodomain mutation. Although the mutant molecule exhibits reduced DNA binding, no specific hormonal target gene is yet apparent, and hypopituitarism may be secondary to the

**LHX3**

Missense and deletion mutations of LHX3 are associated with hypopituitarism except for intact ACTH reserve. These patients also exhibit defective neck rotation because of a rigid cervical spine. PX2 Rieger's syndrome (anterior eye, teeth, and umbilical maldevelopment) may be associated with GHD and haploinsufficiency of the RIEG (POX2) homeobox gene.

The frequency of heritable combined pituitary hormone deficiencies is rare. Nevertheless, within this cohort of patients, PROP1 mutations appear to be the most prevalent, accounting for well over 50% of retrospective reports and over 90% of patients with more than one affected sibling. Pit-1 mutations are less commonly encountered. Patients with a family history of pituitary dysfunction and those who exhibit blunted hormonal responses to TRH, GHRH, or GnRH stimulation should be subjected to molecular screening for PROP1 or Pit-1 defects. The pronounced clinical phenotype of Hesx-1 mutations determines the need for further molecular analysis.

**Lawrence-Moon-Biedl Syndrome**

This autosomal recessive disorder is characterized by hypogonadotrophic hypogonadism, mental retardation, obesity, retinitis pigmentosa, hexadactyly, brachydactyly, or syndactyly. By age 30, most patients are blind. Although most patients have evidence of GnRH deficiency, about 25% of affected males may have primary testicular failure.

**Prader-Willi Syndrome**

These patients have marked hyperphagia and obesity with retarded mental development, muscle hypotonia, and diabetes mellitus. Related conditions include
micrognathia, absent auricular cartilage, and acromesia. The condition has been ascribed to deletion or translocation of chromosome 15. In hypogonadal patients, bilateral cryptorchidism and absent scrotal folds are accompanied by evidence for attenuated GnRH secretion. LH and FSH levels have been restored in some patients with chronic GnRH treatment. Defective oxytocin and vasopressin synthesis has also been reported.

Kallmann's Syndrome

Kallmann's syndrome consists of defective GnRH synthesis with olfactory nerve agenesis or hypoplasia and variable anosmia. Associated developmental disorders include optic atrophy, color blindness, eighth-nerve deafness, cleft palate, renal agenesis, cryptorchidism, and movement disorders. This X-linked recessive disorder has been ascribed to a defective KAL gene located on chromosome Xp22.3. The KAL protein mediates hypothalamic migration of GnRH cells from the primitive olfactory placode, and its absence leads to defective GnRH synthesis and anosmia. Both autosomal recessive and dominant forms of the disorder have been described, indicating the involvement of additional genetic factors in the pathogenesis of the disorder.

Clinical Features

These patients are exposed to low or absent sex steroids from birth. Consequently, females are tall and present with primary amenorrhea and absent secondary sexual development and males have delayed puberty and micropenis.

Laboratory

Absent GnRH secretory pulses result in characteristically low LH and FSH levels in the presence of very low concentrations of estradiol or testosterone. Because the nonprimed normal pituitary may not respond initially to GnRH stimulation (25 to 100 µg intravenously), this test is of little value in distinguishing the hypothalamic defect. In some patients, repetitive GnRH priming may elicit normal pituitary LH and FSH responses, indicating a hypothalamic defect in GnRH secretion.

The differential diagnosis of congenital hypogonadotropic hypogonadism includes Kallmann's syndrome (KAL gene mutation), congenital adrenal hypoplasia (DAX1 mutation), GnRH receptor mutations, leptin and leptin receptor mutations, PROPT gene mutations, and mutations of the LH or FSH molecules themselves. These conditions are characterized by absent or low GnRH-mediated LH secretory patterns in the presence of a structurally normal pituitary gland. The cause of hypogonadotropic hypogonadism still remains elusive in over 80% of patients (see earlier). In the absence of a structural pituitary defect, genetic evaluation of these patients should be undertaken.
Acquired Pituitary Failure

Causes

In the absence of demonstrable hypothalamic-pituitary anatomic damage and after excluding genetic and syndromic causes of pituitary insufficiencies, acquired, often transient, causes of pituitary failure should be considered (Table 8-32). Causes of pituitary insufficiency including pituitary tumors, parasellar masses, hypophysectomy, aneurysms, and pituitary apoplexy have already been discussed. Hypophysitis, diabetes insipidus, diabetes mellitus, and vocal cord paralysis are all relatively common causes of pituitary hormone insufficiency. Following external radiotherapy for pituitary tumors in adults, hypopituitarism is a frequent complication. Growth hormone (GH) deficiency is the most common endocrine disorder seen in irradiated patients, but other anterior pituitary hormones are also affected. In about 50% of irradiated patients, symptoms and signs of pituitary hormone insufficiency develop over a period ranging from a few weeks to several decades. Seventy-five percent of patients with post-traumatic pituitary failure are men younger than 40 years who were involved in a motor vehicle accident within a year of diagnosis. Virtually all patients with subsequent pituitary failure have a history of loss of consciousness after the insult, and half of all such patients have documented skull fracture. One third of these patients have demonstrable signs of hypothalamic or posterior pituitary dysfunction caused by birth trauma, cranial hemorrhage, or death at birth. Although hypopituitarism after head trauma is usually manifest within a year after the insult, some patients may have overtly manifest signs of pituitary failure only after several decades. Seventy-five percent of patients with post-traumatic pituitary failure are men younger than 40 years who were involved in a motor vehicle accident within a year of diagnosis. Virtually all patients with subsequent pituitary failure have a history of loss of consciousness after trauma, and half of all such patients have documented skull fracture. One third of these patients have demonstrable signs of hypothalamic or posterior pituitary hemorrhage or anterior lobe infarction on MRI. Diabetes insipidus is the most common endocrine disorder, encountered in about 30% of these patients. Gonadotropin deficiency, amenorrhea, and hyperprolactinemia occur in children receiving 27 to 32 Gy or 35 Gy of cranial irradiation for a brain tumor in relation to time from irradiation (dxt). This illustrates that the speed at which individual pituitary hormone deficits develop is dose-dependent; the higher the radiation dose, the earlier GH deficiency occurs. The acquired immunedeficiency syndrome (AIDS) is associated with suppressed pituitary function independent of other associated infections. Drugs such as estrogens, which suppress FSH and LH, and GnRH analogues used for treating prostate cancer inhibit gonadotropin action. In addition to pituitary apoplexy, other vascular accidents such as aneurysms, strokes, cavernous sinus thrombosis, and arteritis can cause pituitary hormone insufficiency. Isolated pituitary hormone deficiencies may also occur as a manifestation of vascular abnormalities including arteritis.

Head Trauma

The pituitary gland may be partially or totally damaged by birth trauma, cranial hemorrhage, fetal asphyxia, or breech delivery. Head trauma may lead to direct pituitary damage by a sella turcica fracture, pituitary stalk section, trauma-induced vasospasm, or ischemic infarction after blunt trauma. The most common traumatic cause of compromised pituitary function in the adult is iatrogenic neurosurgical trauma. Adherent or inadvertent pituitary manipulation or damage during surgery leads to transient or permanent diabetes insipidus and varying degrees of anterior pituitary dysfunction.

Although hypopituitarism after head trauma is usually manifest within a year after the insult, some patients may have overtly manifest signs of pituitary failure only after several decades. Seventy-five percent of patients with post-traumatic pituitary failure are men younger than 40 years who were involved in a motor vehicle accident within a year of diagnosis. Virtually all patients with subsequent pituitary failure have a history of loss of consciousness after trauma, and half of all such patients have documented skull fracture. One third of these patients have demonstrable signs of hypothalamic or posterior pituitary hemorrhage or anterior lobe infarction on MRI. Diabetes insipidus is the most common endocrine disorder, encountered in about 30% of these patients. Gonadotropin deficiency, amenorrhea, and hyperprolactinemia may occur in the months following trauma or even years later. Pituitary testing performed within the first 48 hours of hospital admission shows that about 75% of patients have evidence of hypopituitarism, and the degree of pituitary failure correlates with the severity of head trauma.

Radiation

Pituitary radiation, usually indicated as therapy for pituitary adenoma, directly causes atrophy of the gland in addition to the damaging impact of radiation on hypothalamic synthesis of hypothalamic hormones. Pituitary function in children and adolescents is particularly sensitive to head and neck therapeutic radiation. Radiation dose exposure, time interval after completion of radiotherapy, and distance of the pituitary or hypothalamus from the central energy field correlate with the development of pituitary hormone deficits (Fig. 8-64 and Fig. 8-65).

After a median dose of 5000 rads directed at the skull base, nasopharynx, or cranium, up to 75% of patients experience pituitary insufficiency within 10 years. Later manifestations of pituitary failure usually reflect hypothalamic damage rather than atrophy of irradiated pituitary cells. Although the degree of hormone loss after radiation is variable, the pattern of loss usually occurs sequentially with GH before FSH and LH followed by ACTH and TSH. Thus, evidence for secondary thyroid or adrenal failure usually implies that the GH and gonadotropin axes are also compromised. Previously irradiated patients should therefore undergo lifelong periodic anterior pituitary hormone testing. Ideally, rigorous long-term screening should unmask incipient pituitary failure before the onset of morbidity.

Empty Sella Syndrome

Damage to the sellar diaphragm may lead to arachnoid hernaition into the sellar space. An empty sella may develop as a consequence of a primary congenital weakness of the diaphragm in patients in whom no secondary cause is evident. Up to 50% of patients with primary empty sella have associated benign intracranial hypertension. A secondary empty sella may develop after infarction of a pituitary adenoma or surgical or radiation-induced damage to the sellar diaphragm. MRI usually exhibits demonstrable pituitary tissue compressed against the sellar floor with lateral stalk deviation. Although an empty sella is usually an incidental finding, if more than 90% of pituitary tissue is compressed or atrophied, pituitary failure occurs. About 10% of patients may have small adenomas secreting GH or PRL within...
the rim of compressed pituitary tissue.
Clinical Features of Hypopituitarism

The spectrum of clinical features of pituitary insufficiency depends on several factors. In acquired pituitary insufficiency, the clinical spectrum depends on the degree of hormone deficiency, the number of hormones impaired, and the rapidity of onset. In congenital forms, the earlier the age of onset, the greater the severity of thyroid, gonadal, adrenal, growth, or water disturbances. Heritable genetic disorders invariably exhibit the most severe phenotypic changes, although later changes may also occur in these disorders, as seen with PROPT mutations.

The resilience of the individual pituitary cell lineages in the presence of compressive, inflammatory, vascular, radiation, and invasive insults also differs. The lactotroph cell is often hyperfunctional as a result of decreased tonic inhibitory signals. PRL deficiency is thus exceedingly rare except for complete pituitary destruction or genetic syndromes. The order of diminished trophic hormone reserve function with pituitary compression usually is GH prior to the other trophic hormones. The corticotroph and thyrotroph cells appear particularly resistant to hypothalamic or pituitary destruction and are usually the last to lose function. The qualitative phenotypic manifestations of pituitary failure are determined by which specific trophic hormones are lost (see the preceding descriptions of individual hormone deficiencies) (Table 8-34).

Adrenocorticotropic Hormone

Clinical symptoms and signs of ACTH deficiency are most profound and life-threatening. With acute pituitary failure, such as may occur with pituitary apoplexy, patients with ACTH deficiency may present with hypotension, shock, hypoglycemia, nausea and vomiting, extreme fatigue and asthenia, and dilutional hyponatremia. Serum potassium is normal because these patients are deficient in glucocorticoids but usually not mineralocorticoids.

When acute ACTH deficiency is suspected clinically, treatment with steroids should not be withheld. Serum cortisol and ACTH should be determined before glucocorticoid administration and would be expected to be low. In an acute setting, such as sudden apoplexy, responses to Synacthen stimulation may be normal and misleading because the blunted response of cortisol usually seen in secondary adrenal insufficiency is due to loss of glucocorticoid-producing cells, which requires at least several weeks after the onset of ACTH deficiency. When ACTH deficiency occurs gradually, features are more insidious and include weight loss, asthenia, weakness, fatigue, nausea, and dilutional hyponatremia. This may be due to corticotroph dysfunction arising as a result of an enlarging pituitary tumor, delayed effects of radiation, damage related to pituitary tumor surgery, or removal of parasellar masses. In these cases, tests for ACTH reserve are likely to be blunted. If unsuspected or untreated, this form of adrenal insufficiency may also lead to death. Caution must be exercised in performing an ITT or metyrapone test because the latter can cause worsening of adrenal insufficiency and the former seizures and nausea. These tests should be done only in a hospital setting under supervision with available intravenous cortisone and glucose.

Steroid replacement doses should be appropriate to the clinical situation. Under conditions of major stress, such as pituitary apoplexy or pituitary surgery, maximal cortisone requirements range from 200 to 300 mg daily. Because the adrenal glands are incapable of increasing cortisol production during stress in the presence of deficient ACTH, an initial intravenous dose of 100 mg of hydrocortisone (Solu-Cortef) in acute adrenal insufficiency or an intravenous infusion of the same dose during pituitary surgery is followed by 50 mg intravenously every 6 hours for the first day. Similar doses are employed for other forms of major stress in patients with established secondary adrenal insufficiency.

Clinical judgment must be used in determining how long patients should be exposed to supraphysiologic doses of steroids. In our view, doses should be lowered to maintenance as soon as clinically feasible without endangering the patient. We employ replacement doses of 10 to 20 mg of hydrocortisone daily, usually 10 mg in the morning and 5 mg in the evening. An additional 5 mg is administered in the afternoon if needed clinically. Because surgical decompression may lead to recovery from hormone deficiencies, a clinical decision should be made postoperatively about whether to wean patients from hormonal replacement therapy and retest them hormone by hormone. Great care must be taken when secondary adrenal insufficiency has been previously documented.

Thyroid-Stimulating Hormone

Because the half-life of serum T₄ is 6.8 days, hypothyroidism may not become apparent for several weeks in patients with acute pituitary insufficiency; therefore, thyroid function should be tested. However, serum T₄ is not elevated in secondary hypothyroidism and it cannot be used to assess the adequacy of thyroid hormone replacement, nor is a TRH test helpful. Severity of symptoms of hypothyroidism depends on the degree of hypothyroidism and length of time it has been present. Even in the absence of symptoms, T₄ should be administered if thyroid function studies are consistent with hypothyroidism. Glucocorticoids should be replaced before thyroid hormone because thyroid hormone in hypothyroid individuals increases the requirement for glucocorticoids in stressful situations.

Gonadotropins

Sexual dysfunction related to gonadotropin deficiency is far more common than hypothyroidism or hypoadrenalism in patients with pituitary disease. Its presence is established by the constellation of abnormal menses or amenorrhea with no elevated LH and FSH levels in women and sexual dysfunction in men with low testosterone and normal or low gonadotropin levels. Because even mild hyperprolactinemia may cause sexual dysfunction, it should be determined whether PRL is causing hypogonadism. Treatment with dopamine agonists may result in normal sexual function without the need to replace sex steroids. A GnRH stimulation test rarely differentiates causes of gonadotropin deficiency and is usually not indicated.

Sex steroid replacement in deficient patients has important effects on body composition in addition to normalizing sexual function. Testosterone replacement may not be as effective in normalizing sexual function in men with long-standing secondary hypogonadism and loss of libido as it is in men whose sexual dysfunction is recent. Osteoporosis is common in women deficient in estrogen and men deficient in testosterone, and replacement improves bone density. Testosterone reduces abdominal and visceral fat and improves muscle mass in testosterone-deficient men. Therefore, sex hormone replacement is important even though sexual function may not be normalized or is not desired.

Growth Hormone

GHD is discussed comprehensively earlier and is invariably present when two or more other trophic hormones are deficient.

Prolactin

PRL deficiency is extremely rare because it occurs only when the anterior pituitary is completely destroyed, as in patients after apoplexy, or in patients with congenital causes of PRL deficiency. When present, PRL deficiency prevents lactation. In fact, PRL is often elevated in most forms of pituitary insufficiency. For example, many patients with preoperative...
<table>
<thead>
<tr>
<th>Name of Test</th>
<th>Method</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>TSH-Secreting Tumors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TSHecretion Rate</td>
<td>Frankly elevated in 58%</td>
<td></td>
</tr>
<tr>
<td>Serum T&lt;sub&gt;4&lt;/sub&gt;</td>
<td>Elevated in most patients</td>
<td></td>
</tr>
<tr>
<td>Free T&lt;sub&gt;4&lt;/sub&gt;</td>
<td>Elevated in most patients</td>
<td></td>
</tr>
<tr>
<td>T&lt;sub&gt;3&lt;/sub&gt;</td>
<td>Elevated in most patients</td>
<td></td>
</tr>
<tr>
<td>subunit</td>
<td>May be elevated</td>
<td></td>
</tr>
<tr>
<td>T&lt;sub&gt;3&lt;/sub&gt; suppression test</td>
<td>RAI uptake before and after T&lt;sub&gt;3&lt;/sub&gt; 25 µg tid for 8 days</td>
<td>Incomplete suppression</td>
</tr>
<tr>
<td>TRH stimulation</td>
<td>200500 µg IV over 1 min TSH and subunit at 0, 30, 60, and 90 min</td>
<td>TSH response is blunted</td>
</tr>
<tr>
<td>Pituitary MRI</td>
<td>Mostly macroadenomas, some microadenomas</td>
<td></td>
</tr>
<tr>
<td>Thyroid ultrasound</td>
<td>Goiter present in majority</td>
<td></td>
</tr>
<tr>
<td>IGF-I</td>
<td>Can be elevated if GH co-secreted</td>
<td></td>
</tr>
<tr>
<td>GH</td>
<td>Can be co-secreted</td>
<td></td>
</tr>
<tr>
<td>PRL</td>
<td>Can be co-secreted</td>
<td></td>
</tr>
<tr>
<td>ACTH</td>
<td>Can be co-secreted</td>
<td></td>
</tr>
<tr>
<td><strong>Acromegaly</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serum GH</td>
<td>Random measurement is not helpful because fluctuations are too wide.</td>
<td></td>
</tr>
<tr>
<td>Serum IGF-I</td>
<td>Elevated for age and sex-matched controls</td>
<td></td>
</tr>
<tr>
<td>IGFBP-3</td>
<td>Elevated in most acromegaly patients; not as reliable as IGF-I</td>
<td></td>
</tr>
<tr>
<td>OGTT</td>
<td>75 g glucose solution hGH at 0, 30, 60, and 120 min</td>
<td>&lt;1 ng/mL is probably within normal limits</td>
</tr>
<tr>
<td>TRH test</td>
<td>200500 µg IV over 1 min hGH at 0, 30, 60 min</td>
<td>A minority of acromegaly patients have inappropriate GH elevation</td>
</tr>
<tr>
<td>Thyroid ultrasound</td>
<td>Often reveals goiter</td>
<td></td>
</tr>
<tr>
<td>Sleep studies</td>
<td>Consistent with obstructive sleep apnea</td>
<td></td>
</tr>
<tr>
<td>MRI of pituitary</td>
<td>Most have macroadenomas, a minority have microadenomas</td>
<td></td>
</tr>
<tr>
<td>EMG of wrists</td>
<td>Sometimes consistent with carpal tunnel syndrome</td>
<td></td>
</tr>
<tr>
<td><strong>ACTH Overproduction</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24-hour urinary free cortisol</td>
<td>HPLC</td>
<td>Elevated</td>
</tr>
<tr>
<td>8 AM and 4 PM serum cortisol</td>
<td>RIA</td>
<td>Sometimes intermittent elevation</td>
</tr>
<tr>
<td>Overnight 1 mg dexamethasone suppression test</td>
<td>1 mg dexamethasone at 11 pm the night before blood test</td>
<td>Should be &lt;5 µg/dL</td>
</tr>
<tr>
<td>2 mg and 8 mg dexamethasone suppression test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRH stimulation</td>
<td>100 µg IV draw ACTH and cortisol at 0, 15, and 30 min</td>
<td>ACTH: 34% increase c/w Cushing's syndrome</td>
</tr>
<tr>
<td>CRH stimulation after 2 days of dexamethasone</td>
<td>100 µg IV or IM after 0.5 mg q6h</td>
<td>In Cushing's disease, cortisol &gt;4 µg/dL</td>
</tr>
<tr>
<td>Bilateral inferior petrosal sinus sampling</td>
<td>CRH measure ACTH from left and right petrosal sinuses and peripheral blood before and 3, 5, and 10 min after ovine CRH 1 µg/kg IV</td>
<td>Gradient post-CRH above 2.43.2 indicative of central Cushing's disease</td>
</tr>
<tr>
<td><strong>Gonadotropin-Secreting Tumors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FSH</td>
<td>RIA</td>
<td>Usually normal, sometimes elevated in which case LH is normal</td>
</tr>
<tr>
<td>LH</td>
<td>RIA</td>
<td>Usually normal, rarely elevated with normal FSH</td>
</tr>
<tr>
<td>-subunit</td>
<td>May be elevated</td>
<td></td>
</tr>
<tr>
<td>TRH</td>
<td>200500 µg TRH IV over 1 min</td>
<td>LH, FSH, or subunit may be increased</td>
</tr>
<tr>
<td><strong>ACTH Deficiency</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITT</td>
<td>0.10-15 µL/kg IV</td>
<td>Peak cortisol response &gt;20 µg/dL or increase by 10 µg/dL</td>
</tr>
<tr>
<td>Metyrapone test</td>
<td>Oral administration of 30 mg/kg at 11 pm</td>
<td>Peak 11-DOC 7 µg/dL</td>
</tr>
<tr>
<td>CRH stimulation (limited utility)</td>
<td>100 µg IV</td>
<td>Peak ACTH 24-fold increase</td>
</tr>
<tr>
<td>Cortrosyn stimulation</td>
<td>250 µg IV or IM</td>
<td>Peak cortisol 20 µg/dL</td>
</tr>
<tr>
<td>Cortrosyn stimulation (low dose)</td>
<td>1 µg IV cortisol at 0, 30, and 60 min</td>
<td>&gt;20 µg is normal</td>
</tr>
<tr>
<td>8 AM cortisol</td>
<td></td>
<td>&gt;18 µg/dL goes firmly against ACTH deficiency</td>
</tr>
</tbody>
</table>
Urinary free cortisol 24-hour RIA (only)

<table>
<thead>
<tr>
<th>TSH Deficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serum T&lt;sub&gt;4&lt;/sub&gt;</td>
</tr>
<tr>
<td>Free T&lt;sub&gt;4&lt;/sub&gt;</td>
</tr>
<tr>
<td>Serum T&lt;sub&gt;3&lt;/sub&gt;</td>
</tr>
<tr>
<td>Serum TSH</td>
</tr>
<tr>
<td>RAI uptake</td>
</tr>
<tr>
<td>TRH test</td>
</tr>
</tbody>
</table>

**Growth Hormone Deficiency**

<table>
<thead>
<tr>
<th>Mean Peak in GHD Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITT</td>
</tr>
<tr>
<td>Arginine-GHRH</td>
</tr>
<tr>
<td>Arginine-GHRH</td>
</tr>
<tr>
<td>Hexarelin GHRH</td>
</tr>
<tr>
<td>L-Dopa</td>
</tr>
<tr>
<td>Arginine IV over 30 min</td>
</tr>
<tr>
<td>IGF-I</td>
</tr>
<tr>
<td>Lipid panel</td>
</tr>
</tbody>
</table>

**Gonadotropin Deficiency**

<table>
<thead>
<tr>
<th>LH and FSH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal or low in secondary hypogonadism</td>
</tr>
<tr>
<td>Testosterone</td>
</tr>
<tr>
<td>GnRH test</td>
</tr>
<tr>
<td>Clomiphene 50100 mg PO bid for 5 days</td>
</tr>
<tr>
<td>Clomiphene</td>
</tr>
<tr>
<td>Clomiphene</td>
</tr>
<tr>
<td>Clomiphene</td>
</tr>
</tbody>
</table>

ACTH, adrenocorticotropic hormone; CRH, corticotropin-releasing hormone; FSH, follicle-stimulating hormone; GH, growth hormone; GHD, growth hormone deficiency; GHRH, GH-releasing hormone; GnRH, gonadotropin-releasing hormone; PRL, prolactin; T<sub>3</sub>, triiodothyronine; T<sub>4</sub>, thyroxine; TRH, thyrotropin-releasing hormone; TSH, thyrotropin. PRL hypersecretion is diagnosed by single serum measurement.

*Not a good screening test for secondary or tertiary hypothyroidism.*

Hyperprolactinemia related to tumor pressure on stalk structures continue to have hyperprolactinemia even after tumors are surgically debulked. Likewise, hyperprolactinemia occurs in 50% of patients treated with whole-brain radiation, the most common endocrine disturbance.

**Posterior Pituitary**

Diabetes insipidus occurs frequently after pituitary surgery. Hyponatremia may also develop as the second of three phases of postoperative diabetes insipidus or may develop without evidence of diabetes insipidus after surgery. This subject is comprehensively covered in Chapter 9.
Screening for Pituitary Failure

The onset of hypopituitarism may be extremely slow, and subclinical pituitary failure is often not apparent to the patient or physician. Screening for pituitary dysfunction should be undertaken in patients with hypothalamic or pituitary mass lesions, developmental craniofacial abnormalities, inflammatory disorders, brain granulomatous disease, prior head or neck radiation, head trauma, prior skull base surgery, and newly discovered empty sella and in those who previously experienced pregnancy-associated hemorrhage or blood pressure changes.

Because hypopituitarism may develop insidiously and is often not readily clinically apparent, screening of appropriate patients is important to prevent long-term morbidity. Therefore, all patients harboring hypothalamic or pituitary masses should be screened for hypopituitarism. PRL should be measured because many patients with hypopituitarism also present with secondary hyperprolactinemia. Up to two thirds of patients harboring pituitary macroadenomas, craniopharyngiomas, and other parasellar lesions have compromised pituitary reserve function. Less commonly, patients with intrasellar aneurysms, pituitary metastases, parasellar meningiomas, optic gliomas, and hypothalamic astrocytomas also have pituitary failure. Although about a third of patients with hypopituitarism undergoing pituitary surgery recover function after decompression, about 25% of patients experience further loss of pituitary function after surgery and therefore should be screened annually. Treatment of pituitary failure was described fully earlier.
References


264


References


6. Lamberts SW, MacLeod RM. Regulation of prolactin secretion at the level of the lactotroph. Physiol Rev 1990; 70:2739318.


238. Ruan W, Kleinberg DL. Insulin-like growth factor I is essential for terminal end bud formation and ductal morphogenesis during mammary development. Endocrinology 1999; 140:50755081.


246. Richert MMA, Wood TL. The insulin-like growth factors (IGF) and the IGF type I receptor during postnatal growth of the murine mammary gland: sites of messenger ribonucleic acid expression and potential functions. Endocrinology 1999; 140:45461.


References


346. Milenkovic L, D'Angelo O, Kelly PA, Weiner RI. Inhibition of gonadotropin hormone-releasing hormone release by prolactin from GT1 neuronal cell lines through prolactin receptors. Proc Natl Acad Sci USA 1994; 91:12441247.


43. Landolt AM, Osterwalder V. Pervascular fibrosis in prolactinomas: is it increased by bromocriptine? J Clin Endocrinol Metab 1984; 58:11791183.
49. Andreotti AC, Planezello E, Persiani S, et al. Pharmacokinetics, pharmacodynamics, and tolerability of cabergoline, a prolactin-lowering drug, after administration of increasing oral doses (0.5, 1.0, and 1.5 milligrams) in healthy male volunteers. J Clin Endocrinol Metab 1995; 80:841845.


Tannenbaum GS, Ling N. The interrelationship of growth hormone (GH)releasing factor and somatostatin in generation of the ultradian rhythm of GH secretion. J Clin Endocrinol Metab 1984;


Mayo KE, Miller T, DaAlmeida V, et al. Regulation of the pituitary somatotroph cell by GHRH and its receptor. Recent Prog Horm Res 2000;


References


Larsen: Williams Textbook of Endocrinology, 10th ed., Copyright © 2003 Elsevier


581. Barkan AL, Shenker Y, Grekin RJ, Vale WW. Acromegaly from ectopic growth hormonereleasing hormone secretion by a malignant carcinoid tumor: successful treatment with long-acting
582. Frohman LA, Szabo M, Berelowitz M, Stachura ME. Partial purification and characterization of a peptide with growth hormonereleasing activity from extrapituitary tumors in patients with
References

Larsen: Williams Textbook of Endocrinology, 10th ed., Copyright © 2003 Elsevier


References


753. Tagami T, Madison LD, Nagaya T, Jameson JL. Nuclear receptor corepressors activate rather than suppress basal transcription of genes that are negatively regulated by thyroid hormone. Mol Cell Biol 1997; 17:26422648.


Chapter 9 - Posterior Pituitary Gland

Alan G. Robinson
Joseph G. Verbalis

ANATOMY

Normal Anatomy

The posterior pituitary gland is neural tissue and consists only of the distal axons of the hypothalamic magnocellular neurons that make up the neurohypophysis. The perikarya (cell bodies) of these axons are located in the paired supraoptic nuclei and the paired paraventricular nuclei of the hypothalamus.

During embryogenesis, neuroepithelial cells of the lining of the third ventricle mature into magnocellular neurons while migrating laterally to and above the optic chiasm to form the supraoptic nuclei (SON) and to the walls of the third ventricle to form the paraventricular nuclei (PVN). The axon tracts in the hypothalamus are shown in Figure 9-1. In the posterior pituitary gland, the axon terminals of the magnocellular neurons contain neurosecretory granules, membrane-bound packets of hormones stored for subsequent release. The blood supply for the anterior pituitary is through the hypothalamic-pituitary portal system but the posterior pituitary blood supply is directly from the inferior hypophyseal arteries, which are branches of the posterior communicating and internal carotid arteries. The drainage is into the cavernous sinus and internal jugular vein.

The hormones of the posterior pituitary gland oxytocin and vasopressin are synthesized in individual hormone-specific magnocellular neurons. In addition, the magnocellular neurons that synthesize vasopressin and oxytocin, respectively, are clustered into subdivisions of the supraoptic and paraventricular nuclei. The synthesis of oxytocin and vasopressin in separate neurons and the organization of the magnocellular neurons into clusters of oxytocinergic and vasopressinergic cells are compatible with the idea that the secretion and function of each hormone are distinct and individually controlled. Virtually all of the oxytocinergic neurons and vasopressinergic neurons in the supraoptic nucleus project their axons to the posterior pituitary. The organization of the paraventricular nucleus, however, is much more complex and varies among species. In addition to at least three distinct magnocellular divisions consisting of oxytocinergic neurons and vasopressinergic neurons, parvicellular (smaller cells) divisions synthesize other peptides (e.g., corticotropin-releasing hormone, thyrotropin-releasing hormone, somatostatin) and opioids. The parvicellular neurons project to the median eminence, brain stem, and spinal cord, where they play a role in a variety of neuroendocrine autonomic functions. The supraoptic nucleus is a more discrete nucleus, but there are concentrations of oxytocin-containing cells in its dorsal portion and vasopressin-containing cells in its ventral portion.

Dorsally and laterally, the supraoptic nucleus is surrounded by a cell-poor, fiber-rich area (the perinuclear zone). This zone contains GABAergic (γ-aminobutyric acid)-secreting neurons that project outside the general area and to the supraoptic nucleus, where they are reported to have an inhibitory function. Another nucleus that contains many vasopressin but not oxytocin neurons is the suprachiasmatic nucleus located in the midline at the base of and anterior to the third ventricle. The suprachiasmatic nucleus controls circadian as well as seasonal rhythms.
Ectopic Posterior Pituitary

With the development of magnetic resonance imaging (MRI) scans of the brain, it was discovered that T1-weighted MR images showed a bright signal in the posterior pituitary. This new diagnostic imaging technology (see later) allowed the identification of a group of patients in whom there was abnormal anatomy of the posterior pituitary, and the bright spot was recognized in the base of the hypothalamus. These cases are referred to as ectopic posterior pituitary. Most of these cases are recognized in children with growth retardation and anterior pituitary deficiency rather than posterior pituitary deficiency. Although most of these patients do not have clinically apparent diabetes insipidus, when specifically and systematically tested for quantitative abnormalities of thirst or vasopressin secretion, subnormal responses have been noted.

The degree of anterior pituitary deficit depends on the persistence of a pituitary stalk and a retained portal vasculature from the hypothalamus to the anterior pituitary. Most authors believe that this condition is not a traumatic but a congenital abnormality with an undescended posterior pituitary that may be at any level along the pituitary stalk.
SYNTHESIS AND RELEASE OF NEUROHYPOPHYSIAL HORMONES

Vasopressin and oxytocin are nonapeptides consisting of a six-amino-acid ring with a cysteine-to-cysteine bridge and a three-amino-acid tail (Fig. 9-2). All mammals have arginine vasopressin and oxytocin (see Fig. 9-2) with the exception of the pig. In the pig, a lysine is substituted for arginine in position 8 of vasopressin, producing lysine vasopressin. Both genes are found on chromosome 20, although they are situated in a tail-to-tail position and transcribed in opposite directions. The cellular anatomy and biochemistry of synthesis are illustrated in Figure 9-3. For oxytocin, the peptide products are the nonapeptide and a neurophysin but with no glycopeptide. The neurophysin is distinct for each hormone but with high homology.

When a stimulus for secretion of vasopressin or oxytocin acts on the appropriate magnocellular cell body, an action potential is generated and propagates down the long axon to the posterior pituitary. The action potential causes an influx of calcium, which induces a movement of neurosecretory granules to fuse with the cell membrane and extrude the entire contents of the neurosecretory granule into the perivascular space and subsequently into the capillary system of the posterior pituitary. At the physiologic pH of plasma, there is no binding of hormones (vasopressin and oxytocin) to their respective neurophysins and each peptide circulates independently in the bloodstream.

The control of hormone synthesis resides at the level of transcription. Stimuli for secretion of vasopressin or oxytocin also stimulate transcription and increase the messenger ribonucleic acid (mRNA) content in the magnocellular neurons. This has been studied in most detail in rats, in which dehydration accelerates transcription and increases the levels of vasopressin (and oxytocin) mRNA and hypo-osmolality produces a decrease in the content of vasopressin mRNA. The transport of neurosecretory vesicles from the site of synthesis to the posterior pituitary along microtubule tracks is also regulated. When synthesis is turned off, transport stops, and when synthesis is increased transport is up-regulated. Thus, there is coordination of stimulated release of hormone, transport of hormone, and synthesis of new hormone. There is, however, asynchronous in the timing of these events. The asynchrony is demonstrated by changes in the content of vasopressin stored in the posterior pituitary. The absolute content varies considerably among species but is quite a remarkable store, generally equivalent to the amount of hormone required to sustain basal release for 30 to 50 days or maximum release for 5 to 10 days.

In animals, prolonged and intense stimulation of vasopressin release, such as dehydration or salt loading, produces a depletion of stored hormone in the posterior pituitary. There is then a gradual recovery of pituitary content back to baseline (or above) 7 to 14 days after animals are returned to normal water intake. This phenomenon has been modeled by Fitzsimmons and colleagues, who provided experimental evidence that a long half-life of the vasopressin message, approximately 2 days, is (from a minimalist point of view) a plausible explanation for the events. When a strong or sustained stimulus releases vasopressin, there is an immediate stimulus to the synthesis of new mRNA. However, because it requires several days for the peak level of mRNA to be reached, synthesis increases slowly. When the stimulus is removed, the elevated mRNA synthesizes hormone to replete the store in the posterior pituitary. The mRNA slowly declines to the previous baseline, and synthesis of vasopressin returns to a basal rate.

The magnocellular neurons specific for oxytocin and vasopressin have intrinsic individual characteristic electrical firing patterns. These patterns are modulated by the paracrine and autocrine action of hormone released by the dendrites into the extracellular space surrounding the magnocellular neurons in the supraoptic nucleus and paraventricular nuclei. Oxytocin neurons develop a pattern of high-amplitude bursting activity (hormone release) followed by long pauses, a pattern that may facilitate the pumping action of muscular contraction of myoepithelial cells in the breast. For vasopressin, weakly active neurons and highly active neurons are brought to a medium level of alternate phasic firing and resting that facilitates optimal secretion of vasopressin.

Structural plasticity also enhances secretion. When stimulated to secrete, the neurons retract dendrites to become more compact, which may allow more efficient propagation of inputs and decrease nonspecific synaptic inputs. Retraction of glia around the magnocellular perikarya increases juxtaposition of like neurons to enhance recruitment of neighboring neurons and to synchronize firing. At the level of the posterior pituitary, retraction of pericytes surrounding axon terminals removes an immediate barrier between the axons and the perivascular space and facilitates diffusion of peptides into capillaries.
PHYSIOLOGY OF SECRETION OF VASOPRESSIN AND THIRST

The physiologic regulation of vasopressin synthesis and secretion involves two systems: osmotic and pressure-volume. (Fig. 9-4). The functions of these two systems are so distinct that historically it was thought there were two homologous antidiuretic hormone and a vasopressor hormone. Hence, the two names that are used interchangeably for (8-arginine) vasopressin.

There are separate systems at the level of the receptors on the end organs of response. V1 receptors on blood vessels are distinct from V2 receptors on renal collecting duct epithelia. A third receptor, V3, is responsible for the nontraditional biologic action of vasopressin to stimulate adenocorticotropic hormone (ACTH) secretion from the anterior pituitary, and V2 receptors regulate the nontraditional action of vasopressin to stimulate factor VIII production.

Vasopressin is the main hormone involved in regulation of water in humans, and all mammals control water to regulate osmolality. On the other hand, the main hormones involved in pressure-volume in humans are renin, angiotensin, and aldosterone; and controlling serum sodium concentration (Na"^+) largely regulates pressure-volume. Therefore, the pathology of disorders of the neurohypophysis is expressed primarily as abnormalities of osmolality produced by abnormal excretion or retention of water. In the case of osmoreceptors, the magnicellular neurons are chronically under some mild input to stimulate release of vasopressin, and the regulation of vasopressin in response to osmolality is relatively uncomplicated, with small decreases in osmolality causing a parallel decrease in

Figure 9-3 Vasopressin synthesis in a magnicellular neuron. A. The vasopressin gene is located on the short arm of chromosome 20. In the nucleus, the gene is transcribed to heteronuclear ribonucleic acid (RNA). B. The introns are then excised, and the three exons are spliced to form mature RNA, which consists of exon A, exon B, and exon C. The mature RNA exits the nucleus to the cytoplasm. It is targeted to the endoplasmic reticulum, where it is attached to ribosomes. C and D, There is translation of the three exons into pre-provasopressin. Exon A is translated to the 19-amino-acid signal peptide (SP), the nonpeptide arginine-vasopressin (AVP), and the amino-terminal portion of the 93- to 95-amino-acid neurophysin (NP). Exon B encodes the highly conserved middle region of neurophysin. Exon C encodes the variable carboxyl terminal of neurophysin and a 39-amino-acid glycopeptide (GP). The pre-provasopressin is transferred across the endoplasmic reticulum, the glycopeptide is glycosylated, and the signal peptide is cleaved (D). E. Provasopressin enters the Golgi apparatus, where the entire provasopressin complex is packaged into neurosecretory granules. The neurosecretory granules attach to microtubules and are transported along the microtubules to the posterior pituitary, where the neurosecretory granules are stored. F. During transport, enzymes in the acidic granules cleave the prohormone to vasopressin (which is amidated), to neurophysin, and to the glycopeptide. Neurophysins form dimers and, subsequently, tetramers with one vasopressin attached to each neurophysin. There is an auxiliary fifth binding site for vasopressin, which spans the four neurophysin molecules of the tetramer. G. When there is an action potential signaling release, a neurosecretory granule fuses with the axon membrane and the vasopressin, neurophysin, and glycopeptide are secreted into the extracellular space and, hence, into plasma, where they circulate independently of each other. (© 2003, UCLA, AG Robinson.)

Figure 9-4 Comparison in humans of the release of vasopressin in response to percentage changes of osmolality (increase) and pressure or volume (decrease). Note: To increase plasma vasopressin, the change in osmolality is much more sensitive, responding to as little as a 1% increase in osmolality, whereas volume and pressure require greater than a 10% to 15% change to stimulate release of vasopressin. (Redrawn from Robertson GL, Bier T. Water metabolism. In Brenner BM, Rector FC Jr [eds]: The Kidney, vol 1, 3rd ed. Philadelphia, WB Saunders, 1986, p 385.)

vasopressin and small increases in osmolality a parallel increase in vasopressin.

The regulation of volume and blood pressure is complicated (see Thrasher[0]), and experimental models of vasopressin and baroreceptor regulation in animals often involve inhibiting or measuring other concurrent sympathetic inputs to the system in order to determine any direct effect of a stimulus on secretion of vasopressin (see Fig. 9-4). Other influences on secretion of vasopressin, such as the inhibiting influence of glucocorticoids and the potent stimulus of nausea and vomiting, are less important as physiologic regulators of vasopressin but may be important in the differential diagnosis of the syndrome of inappropriate secretion of antidiuretic hormone (SIADH), discussed later.

Volume and Pressure Regulation

High-pressure arterial baroreceptors are located in the carotid sinus and aortic arch and low-pressure volume receptors in the atria and pulmonary venous system. [2][1] The afferent signals from these receptors are carried from the chest to the brain stem through cranial nerves IX and X. Interruption of the vagal input by vagotomy or vagal cold block[2][3][4] in dogs and destruction of the A1 area of the medulla, which receives input from IX and X, [1][2][3][4] in rabbits leads to an increase in vasopressin secretion.

These and other data led to the concept that baroreceptors and volume receptors normally inhibit the magnicellular neurons and that decreases in this tonic inhibition result in release of vasopressin. Arterial and venous constriction induced by vasopressin action on V1 receptors causes the vessels around the existing plasma volume to contract and effectively increase plasma volume and reestablish the inhibition of secretion of vasopressin. Although the action of vasopressin at the kidney to retain water helps to replace volume, the major hormonal regulation to control volume actually involves the renin-angiotensin system, which stimulates sodium reabsorption in the kidney (see Chapter 15).

The concept of tonic inhibition of vasopressin by baroreceptors has been questioned,[1][2][3] but most agree that the volume receptor and baroreceptor responses leading to an increase in vasopressin in humans are much less sensitive than are the osmoreceptor responses (see Fig. 9-4). It has been interpreted that the lesser response occurs because changes in volume and central venous pressure have little effect to increase vasopressin in humans as long as arterial pressure is maintained by sympathetic reflexes.[3][4] When hypovolemia is sufficient to cause a decrease in blood pressure, there is a sudden and exponential increase in the level of vasopressin in plasma (see Fig. 9-4). [1][2][4][5]

It is also agreed that changes in volume or pressure that are insufficient to cause increases in vasopressin nonetheless modify the response of the vasopressin system to osmoregulation.[3][4] Increases in pressure and central volume decrease the secretion of vasopressin, [3][4] but, again, the response of the renin-angiotensin system to cause sodium excretion is much more sensitive to increases of pressure and volume than the response to decrease secretion of vasopressin.[3][4] Thus, both excitatory and inhibitory influences exist from the brain stem to the magnicellular neurons, with the dominant influence depending on the physiologic circumstances.
Osmonic Regulation

The primary receptors for sensing changes in osmolality are located in the brain. Most of the brain is within the blood-brain barrier, which is impermeable to polar solutes. Because the osmostat is insensitive to urea and glucose, which readily cross cellular membranes (but not the blood-brain barrier), the osmoreceptors must be outside the blood-brain barrier.

Studies of experimental brain lesions in animals strongly suggest that cells in the organum vasculosum of the lamina terminalis (OVLT) and in areas of the adjacent anterior hypothalamus near the anterior wall of the third cerebral ventricle are osmoreceptors. These organs are perfused by fenestrated capillaries and are thus outside the blood-brain barrier. Surgical destruction of the OVLT abolishes vasopressin secretion and thirst responses to hyperosmolality but not responses to other stimuli such as hypovolemia. 

Essentially the same conclusion was drawn from clinical observations of human subjects with brain damage that destroyed the region around the OVLT, who are often unable to maintain normal plasma osmolalities even under basal conditions. In contrast, destruction of the macrocellular neurons of the supraoptic nucleus and paraventricular nuclei eliminates dehydration-induced secretion of vasopressin but does not alter thirst, clearly indicating that osmotically stimulated thirst must be generated proximal to the macrocellular cells.

Extracellular fluid (ECF) osmolality (determined predominantly by sodium concentration) in normal subjects varies from 280 to 295 mOsm/kg but in any individual is maintained in a narrow range. The ability to maintain this narrow range depends on (1) the sensitive response of plasma vasopressin to changes in plasma osmolality, (2) the sensitive response of urine osmolality to changes in plasma vasopressin, and (3) the gain in the system by the response of urine volume to changes in plasma vasopressin (Fig. 9-5). Basal plasma vasopressin is in the range of 0.5 to 2 pg/μL. As little as a 1% increase or decrease in plasma osmolality causes a rapid increase in plasma vasopressin by release of vasopressin from the store of hormone in the posterior pituitary. For levels of vasopressin in plasma to decrease rapidly requires rapid metabolism of vasopressin, and this is also characteristic of the hormone, which circulates freely in plasma and has a half-life of approximately 15 minutes. Thus, small increases in osmolality produce a concentrated urine and small decreases in osmolality produce a water diuresis.

Figure 9-5 illustrates the linear relationship between plasma osmolality and plasma vasopressin that has been described in humans. The linear relationship exists for osmolalities well above the normal excursion of osmolalities, as demonstrated when the increase is induced by infusion of hypertonic saline or is observed during dehydration of patients with nephrogenic diabetes insipidus. Similarly, Figure 9-5 shows the presence of a sensitive and linear relationship between the level of vasopressin in plasma and the induced osmolality of the urine. In this case, however, although plasma vasopressin may increase out of the normal physiologic range, the urine osmolality levels off at approximately 1000 to 1200 mOsm/kg. This occurs because the maximum concentration reached by the fluid in the collecting duct is determined by the osmolality of the inner medulla.

Figure 9-5 also shows the relationship of plasma vasopressin to urine volume. This relationship is calculated on the basis of the urine volume necessary to excrete a fixed quantity of osmoles (800 mOsm) at the urine osmolality produced by the change in plasma vasopressin. These graphs demonstrate the gain in the system when the changes in urine volume relative to plasma vasopressin are considered. With a decrease of plasma vasopressin, for example, from 5 to 1 pg/μL, urine volume is maintained at less than 4 L/day; however, urine volume increases dramatically to 18 to 20 L/day when plasma vasopressin is decreased further.

In the kidney, water is conserved by the combined functions of the loop of Henle and the collecting duct. The loop of Henle generates a high osmolality in the renal medulla through the countercurrent multiplier system. Vasopressin acts in the collecting duct to increase water permeability, thereby allowing osmotic equilibration between the urine and the hypertonic medullary interstitium. The net effect of this process is to extract water from the urine into the medullary interstitial blood vessels (vasa recta), resulting in increased urine concentration and decreased urine volume (antidiuresis). Vasopressin produces antidiuresis by its effects on the epithelial principal cells of the collecting tubule, which have vasopressin receptors of the V2 type.

The intracellular organelles responsible for water reabsorption across the collecting duct cells are called aquaporins, a widely expressed family of water channels that mediate rapid water transport across some cell membranes. Aquaporin-2 is regulated by vasopressin and mediates water transport across the apical plasma membrane of the principal cells of the collecting ducts. In contrast, aquaporin-3 and aquaporin-4 are expressed at high levels in the basolateral plasma membranes of principal cells and are responsible for the constitutively high water permeability of the basolateral plasma membrane.

Vasopressin binding to the V2 receptor increases intracellular cyclic adenosine monophosphate (cAMP) levels by activating adenylate cyclase. The cAMP induces a fusion of aquaporin-2 containing intracytoplasmic vesicles with the apical plasma membranes of the principal cells, a process that increases apical water permeability by markedly increasing the number of water-conducting pores in the apical plasma membrane. Dissociation of vasopressin from the V2 receptor allows intracellular cAMP levels to decrease, and the water channels are reinternalized into the intracytoplasmic vesicles, thereby terminating the increased water permeability.

The aquaporin-containing vesicles remain just below the apical membrane and can be quickly "shuttled" into and out of the membrane in response to changes in intracellular cAMP levels. This mechanism therefore allows minute-to-minute regulation of renal water excretion through changes in ambient levels of vasopressin in plasma. There is also long-term regulation of collect duct water permeability in response to prolonged high levels of circulating vasopressin. This response requires at least 24 hours to elicit and is not as rapidly reversible. The long-term effect is due to the ability of vasopressin to induce large increases in the abundance of aquaporin-2 and aquaporin-3 water channels in the collecting duct principal cells.
Thirst

Urine volume can be reduced to a minimum but not eliminated, and insensible water loss is a continuous process. To maintain water balance, one must consume water to replace the obligate urinary and insensible fluid losses, and this consumption is regulated by thirst. Thirst represents the body’s defense mechanism to increase water consumption in response to perceived deficits of body fluids.

Like vasopressin, thirst can be stimulated by increases in osmolality of the ECF or by decreases in intravascular volume. Furthermore, there is evidence that the receptors are similar, that is, osmoreceptors in the anterior hypothalamus and low-pressure or high-pressure, or both, baroreceptors (with a likely contribution from circulating angiotensin II during more severe degrees of intravascular hypovolemia and hypotension). Studies in animals have consistently reported thresholds for osmotically induced drinking ranging from 1% to 4% increases in plasma osmolality above basal levels, and analogous studies in humans using quantitative estimates of subjective symptoms of thirst have confirmed that increases in plasma osmolality of 2% to 3% are necessary to produce an unequivocal sensation described as thirst.

As with vasopressin, the threshold for producing thirst by hypovolemia is significantly higher. Studies in multiple species have shown that sustained decreases in plasma volume or blood pressure of at least 4% to 8%, and in some species 10% to 15%, are necessary to stimulate drinking consistently. In humans, it is difficult to demonstrate an effect of mild to moderate hypovolemia to stimulate thirst independently of osmotic changes occurring with dehydration. This blunted sensitivity to changes in ECF volume or blood pressure in humans probably represents an adaptation that occurred as a result of the erect posture of primates, which predisposes them to wider fluctuations in blood and atrial filling pressures as a result of orthostatic pooling of blood in the lower body. Stimulation of thirst (and secretion of vasopressin) by transient postural changes in blood pressure may lead to overdrinking and inappropriate antidiuresis in situations in which the ECF volume was actually normal but transiently maldistributed.

Although osmotic changes clearly are effective stimulants of thirst, it is not likely that changes in plasma osmolality are responsible for the major part of day-to-day fluid intakes. Most humans consume the bulk of their ingested water as a result of the relatively unregulated components of fluid intake, such as the consumption of beverages in association with food intake, for reasons of palatability or desired secondary effects (e.g., caffeine), or for social or habitual reasons (e.g., sodas or alcoholic beverages). As a result, both animals and humans generally ingest volumes in excess of what can be considered to be an actual need for fluid.

Consistent with this observation is the fact that, under most conditions, plasma osmolalities in humans remain within 1% to 2% of basal levels, and these relatively small changes in plasma osmolality are generally below the threshold levels that have been found to stimulate thirst. This suggests that despite the obvious vital importance of thirst during pathologic situations of hyperosmolality and hypovolemia, under normal physiologic conditions water balance in humans is accomplished more by free water excretion regulated by vasopressin than by water intake regulated by thirst. This also demonstrates why water intake must be consciously restricted in cases of persistent unregulated secretion of vasopressin.
Clinical Consequences of Osmotic and Volume Regulation

In most physiologic situations, there is concurrence and synergy between the effects of increased osmolality and decreased volume to stimulate release of vasopressin. For example, with dehydration, osmolality increases and volume decreases and each stimulates the release of vasopressin. Furthermore, there is good evidence that a decrease in volume shifts the plasma vasopressin/plasma osmolality response curve to the left, resulting in a greater release of vasopressin at a given osmolality. Similarly, excess of fluid produces a decrease in osmolality and an increase in volume, and both cause a decrease in vasopressin secretion.

The physiology underlying the relationships between plasma osmolality, plasma vasopressin, and especially urine volume determines some of the pathophysiology of decreased or increased secretion of vasopressin. In Figure 9-5, we see that a regular loss of vasopressin neurons, which may decrease the secretory capacity of the neurohypophysis from that able to produce a blood vasopressin level of 10 to 20 pg/mL down to a secretory capacity sufficient only to maintain a blood level of 5 pg/mL, may produce no change in the ability to attain a maximum urine osmolality. Below 5 pg/mL, there is a linear decrease in the ability to concentrate the urine maximally. However, from the volume curve we see that this results in only a modest increase in urine volume because of the logarithmic relationship of urine volume to urine osmolality and plasma vasopressin. Then, when the last few vasopressinergic neurons are lost and the maximum vasopressin level drops from 1 to 0.5 pg/mL, there might be a great increase in urine volume.

These responses might be viewed as protective, allowing water conservation, even with minimal ability to secrete vasopressin. This may be why even in idiopathic hypothalamic diabetes insipidus there is often a sudden onset of symptoms. The same physiologic considerations may explain why patients with diabetes insipidus that has persisted for a relatively long period of time (e.g., after surgery or head injury) may eventually be able to discontinue vasopressin treatment. The number of vasopressinergic neurons that need to recover to maintain an asymptomatic urine volume is small. The same pathophysiology is important in regard to SIADH. In this situation, however, one might consider the consequences of an inability to suppress vasopressin to less than 1 pg/mL. Note that the maximum urine volume with a standard osmolar load at 1 pg/mL can be as little as 2 L/day. If a patient's fluid intake is greater than that which can be excreted with the fixed level of vasopressin of 1 pg/mL, the extra fluid is retained and the sequence of events that causes hyponatremia in SIADH is initiated.

A synthesis of what is known about the regulation of thirst and secretion of vasopressin in humans contributes to our understanding of this simple but elegant system to maintain water balance. Under normal physiologic conditions, the sensitivity of the osmoregulatory system for secretion of vasopressin accounts for maintenance of plasma osmolality within narrow limits by adjusting renal water excretion to small changes in osmolality. Stimulated thirst does not represent a major regulatory mechanism under these conditions, and unregulated fluid ingestion and water from metabolized food supply water in excess of true need. Excess water is then excreted using osmoregulated secretion of vasopressin. However, when unregulated water intake does not supply body needs, even with plasma levels of vasopressin sufficient to produce maximal antidiuresis, plasma osmolality rises to levels that stimulate thirst and produce water intake proportional to the elevation of osmolality. Thirst thus represents a backup mechanism that is called into play when pituitary and renal mechanisms are insufficient to maintain plasma osmolality within a few percentage points of basal levels.

This arrangement has the advantage of freeing animals and humans from frequent episodes of thirst that would require a diversion of activities toward behavior oriented to seeking water when the water deficiency is sufficiently mild to be compensated for by renal water conservation, but it does stimulate water ingestion when water deficiency reaches a potentially harmful level. This system of differential effective thresholds for thirst and secretion of vasopressin therefore nicely complements the excess unregulated, or need-free, drinking in both humans and animals demonstrated in many studies.
Reset Osmostat during Pregnancy

During pregnancy, major shifts of fluid produce a decreased plasma osmolality of about 10 mmol/kg and an increase in plasma volume.\(^81\)\(^82\) This decrease in osmolality is a normal consequence of pregnancy and is probably the best example of a true resetting of the osmostat. A true resetting of the osmostat must regulate both increases and decreases of secretion of vasopressin at a lower than normal plasma osmolality. These criteria are met in pregnant women, and there is a resetting of the osmostat for thirst in parallel with the resetting of the osmostat for release of vasopressin.\(^83\)

Figure 9-6 shows the relationship of plasma vasopressin to plasma osmolality in normal and pregnant women. In pregnant women, as in nonpregnant women, the plasma vasopressin level is increased with as little as a 1% increase in plasma osmolality, but the entire curve with a normal slope is shifted to the left.\(^81\) The shift in osmotic threshold appears at about 5 to 8 weeks of gestation and persists throughout pregnancy, returning to normal by 2 weeks after delivery.\(^81\) Although the osmotic threshold is constant throughout pregnancy, there is a change in sensitivity at about 28 to 33 weeks of gestation. At this time, the threshold to release vasopressin occurs at the same plasma osmolality; however, the slope changes, so that for increases in osmolality, there is less vasopressin (see Fig. 9-6, slope C).

The physiology of the reset osmostat has been considered in relation to the expanded plasma volume. Total body water in pregnant women is also increased by 7 to 8 L as a result of profound vasodilatation.\(^84\) This volume is sensed as normal, and vasopressin responds normally to decreases and increases of volume.\(^81\)\(^84\)\(^85\)\(^86\) Both the changes in volume and the changes in regulation of osmolality have been reproduced by infusion of relaxin (a normal hormone of pregnancy that is a member of the insulin-like growth factor family) into virgin female and normal rats\(^87\)\(^88\) and reversed in pregnant rats by immunoneutralization of relaxin; therefore, relaxin is a proposed mediator of the effect.

In women, the placenta produces an enzyme, cysteine aminopeptidase, that is released into the plasma and is known as oxytocinase.\(^81\)\(^82\) This enzyme is as potent in degrading vasopressin as in degrading oxytocin and in making vasopressin biologically inactive. The activity of vasopressinase increases markedly around 20 weeks of gestation and increases further to 40 weeks, returning slowly to normal over a few weeks after delivery.\(^81\) The data are consistent with the supposition that the change in the sensitivity of the plasma vasopressin-plasma osmolality response in late pregnancy (see Fig. 9-6, line C) may be due to accelerated degradation of vasopressin.
Aging

Many physiologic processes are compromised in aging humans, and numerous studies have reported that elderly people are at risk for both hypernatremia and hyponatremia. In many older subjects, there is a decrease in glomerular filtration rate (GFR) and the collecting duct in the aged kidney may be less sensitive to vasopressin, limiting the ability to excrete free water. Many other abnormalities of fluid and electrolyte balance in elderly people are due to co-morbid conditions or to the numerous pharmacologic agents to which these patients are often exposed. Studies of responses to dehydration, osmolar stimulation, or volume stimulation in older people are complicated by the fact that by age 75 to 80 years the total body water level declines to 50% of the level in normal young adults.

Although some have reported altered morphology of the neurohypophyseal system in aged humans, there appears to be no change in the number of magnocellular neurons in normal elderly people or in patients with Alzheimer's disease. There is a greater range of normal levels of vasopressin and a less direct correlation of plasma vasopressin with plasma osmolality, but changes in levels of vasopressin in response to acute increases in serum sodium are either normal or increased. An increased response of vasopressin to changes in plasma osmolality has been ascribed to a decreased ability of vasopressin to stimulate levels of aquaporin-2 in the kidney. The decrease in renal sensitivity has also been interpreted as causing a chronic increase in secretion of vasopressin and a depletion of hormone stores in the posterior pituitary. This may be why elderly patients demonstrate a decreased incidence of visualization of the bright spot on T1-weighted MRI scans; in more than 70% of elderly subjects, the spot was not observed.

A number of studies have indicated that elderly people have decreased thirst with dehydration and less fluid intake to return their volume to normal during recovery from dehydration. At the other end of the spectrum, elderly patients have been found to excrete a water load less well than younger subjects and at least part of this is due to decreased suppression of vasopressin. Elderly people are also reported to have a decreased ability to shut off vasopressin in response to drinking and stimulation of oral-pharyngeal receptors.

In summary, there are age-related changes in body volumes and renal function that probably predispose elderly people to abnormalities in water and electrolyte balance. Diseases that are more common in elderly persons exacerbate this and, in addition to therapy for these diseases, affect water balance. Healthy older humans probably have at least a normal ability to secrete vasopressin but a decreased appreciation for thirst and a decreased ability to achieve either a maximum concentration of urine to retain water or a maximum dilution of urine to excrete water. Thus, it is necessary to pay attention to fluid balance problems in older people as undetected hypernatremia or hyponatremia can lead to increased morbidity and mortality.
DIABETES INSIPIDUS

Diabetes insipidus is a disorder in which there is a large volume of urine (diabetes) that is hypotonic, dilute, and tasteless (insipid). This is in contrast to the hypertonic and sweet urine of diabetes mellitus (honey). Four pathophysiologic mechanisms related to vasopressin produce large volumes of dilute urine and polydipsia, resulting from the following conditions:

1. Hypothalamic (central or neurohypophyseal) diabetes insipidus, with inability to secrete and usually to synthesize vasopressin in the neurohypophyseal system.
2. Nephrogenic diabetes insipidus, in which there is an inappropriate renal response to vasopressin.
3. Transient diabetes insipidus of pregnancy, produced by the accelerated metabolism of vasopressin.
4. Primary polydipsia, in which the initial pathophysiology involves the ingestion of fluid rather than the excretion of fluid.

Differential Diagnosis

To determine whether there is a large volume of urine, one can measure a 24-hour urine collection or the patient can keep a diary for 24 hours, recording the volume and the time of each voided urine. Simultaneously, it must be determined whether polyuria is due to an osmotic agent, such as glucose, or to intrinsic renal disease. Usually, routine laboratory studies and the clinical setting distinguish these diseases or disorders from consideration of diabetes insipidus. If the thirst mechanism is intact, most patients are ambulatory with normal serum sodium levels and no evidence of dehydration. All agree that the diagnosis of diabetes insipidus is confirmed by the presence of some dehydration to stimulate the normal release of vasopressin and then by the absence of the ability to concentrate the urine.

The test most commonly used clinically is a dehydration test in a controlled environment, followed by a response to administered vasopressin or to the analogue desmopressin. If the patient has mild polyuria, the test may begin in the evening with the majority of dehydration taking place overnight. If the patient gives a history of large volumes of urine during the night, it is best to perform the test during the day when the patient can be observed.

The patient is weighed at the beginning of test in attire that can be worn throughout the study and on a clinical quality scale that can be used for all repeated weighings. The patient voids, and the starting weight is recorded. A serum sodium level is obtained, and nothing is allowed by mouth (certainly no fluid) during the test. Each voided urine is then recorded and urine osmolality measured. The patient is weighed after each liter of urine is excreted.

When two consecutive measures of urine osmolality differ by no more than 10% and the patient has lost 2% of the body weight, plasma is drawn for Na⁺, osmolality, and vasopressin determinations. The patient is given 2 µg of desmopressin intravenously or intramuscularly (or 5 units of aqueous vasopressin subcutaneously), and urine output and osmolality are recorded hourly for an additional 2 hours. The test is discontinued if the patient loses more than 3% of the body weight or at any time that serum Na⁺ is elevated above the normal range. The duration of the test varies; patients with complete diabetes insipidus reach a maximum but very low urine osmolality within a few hours; patients with other disorders reach a maximum in up to 18 hours.

There is no difficulty in determining the diagnosis of severe hypothalamic or severe nephrogenic diabetes insipidus. In the former, urine has minimal concentration despite dehydration and there is a marked increase in urine osmolality in response to administered desmopressin, at least 50% but often a 200% to 400% increase. At the end of the test, these patients have undetectable vasopressin in plasma. In patients with nephrogenic diabetes insipidus, there is also little concentration of the urine despite achieving dehydration, but the urine osmolality also shows little or no response to the administered desmopressin. These patients are unequivocally distinguished from those with hypothalamic diabetes insipidus by high levels of vasopressin in plasma, often greater than 5 pg/ml, at the end of the dehydration phase.

The difficulty is in differentiating partial hypothalamic diabetes insipidus from primary polydipsia. In both disorders, the urine shows some concentration (often above plasma osmolality) with dehydration but the urine osmolality does not approach the 800 to 1000 mOsm/kg characteristic of normal subjects. In response to the administered desmopressin, patients with partial hypothalamic diabetes insipidus usually show a further concentration of the urine of at least 10%, whereas patients with primary polydipsia show no further increase. The reliability of the response to desmopressin is debated. Some patients with primary polydipsia may achieve a plateau level in urine osmolality before reaching their maximum attainable urine osmolality and hence respond to desmopressin. Alternatively, some patients with partial hypothalamic diabetes insipidus may, with severe dehydration, secrete sufficient vasopressin to achieve the maximum attainable urine osmolality and do not respond with a further increase to administered desmopressin.

Investigators who have a highly sensitive radioimmunoassay for vasopressin are able to distinguish between partial hypothalamic diabetes insipidus and primary polydipsia by the measurement of vasopressin at the end of the dehydration phase and further report that one of these disorders may be inappropriately diagnosed using the standard dehydration test. However, a longitudinal clinical study of patients with autoimmune hypothalamic diabetes insipidus reported good correlation between results of the dehydration test and measured vasopressin to diagnose partial diabetes insipidus occurring over time.

There is concern about making the diagnosis of partial diabetes insipidus in patients with primary polydipsia because patients given desmopressin may experience symptomatic hyponatremia as they continue to drink fluid despite desmopressin-induced water conservation. Therefore, when the diagnosis is in doubt, patients should have adequate follow-up to ensure that a good therapeutic response is obtained and that hyponatremia does not develop. This clinical follow-up and response have been considered a continuation of the diagnosis with the trial of desmopressin as a test agent. If a standard dose of desmopressin produces a decrease in polyuria, a decrease in thirst, and no reduction of sodium, the patient almost certainly has partial hypothalamic diabetes insipidus. If polydipsia does not improve and hyponatremia develops, the patient has some abnormality of thirst, and the diagnosis may be primary polydipsia.

The clinical presentation is often helpful in the differential diagnosis. In a patient with no previous history of polyuria or polydipsia who is found to have these symptoms immediately after surgery in the hypothalamic-pituitary area or after head trauma (especially with skull fracture and loss of consciousness), the diagnosis of hypothalamic diabetes insipidus is highly probable. Sometimes diuresis after surgery is the result of water retention during the procedure. Vasopressin is released during surgical procedures, and administered fluid may be retained. As the stress of surgery abates, the vasopressin level falls and administered fluid is excreted. If an attempt is made to match the urine output with further fluid infusion, persistent polyuria occurs and may be mistaken for diabetes insipidus.

Because these patients may be unconscious and may be unable to sense thirst, it is crucial that the diagnosis be established and that patients be treated appropriately to prevent severe dehydration. Because these patients do not sense thirst, it is easy to withhold fluids until there is a modest increase in sodium and then to measure urine osmolality and determine the response to administered desmopressin. If urine output decreases and the serum sodium level remains normal, the response was excretion of physiologically retained fluid. If the serum sodium begins to rise, a response to desmopressin should be determined and the diagnosis of diabetes insipidus established.

Patients with hypothalamic diabetes insipidus often experience a sudden onset of symptoms and persistent thirst throughout the day and night, whereas patients with renal disease experience a more gradual onset of disease and patients with primary polydipsia may have decreased thirst and urination during the night. Hypothalamic diabetes insipidus is associated more with a desire for cold liquids, probably because of dehydration. Patients with diabetes insipidus often have serum sodium levels in the high range of normal, whereas patients with primary polydipsia have serum sodium levels in the low range of normal. Blood urea nitrogen (BUN) concentration is often low in both hypothalamic diabetes insipidus and primary polydipsia because of the high renal clearance, but there is a difference in serum uric acid concentrations.
Serum uric acid is elevated in hypothalamic diabetes insipidus because of modest volume contraction and because vasopressin acts on V1 receptors in the kidney to increase urate clearance. Therefore, in patients with no vasopressin, the uric acid level is high; a value greater than 5 µg/dL has been reported to separate hypothalamic diabetes insipidus from primary polydipsia. Presumably, in patients with primary polydipsia, there is modest volume expansion and intermittent secretion of vasopressin to act on V1 receptors to clear serum urate. Urine volume greater than 18 L suggests primary polydipsia because the volume exceeds the amount of urine delivered to the collecting duct. In fact, most patients with hypothalamic diabetes insipidus have modest dehydration, have a decreased GFR, and excrete urine volumes in the range of 6 to 12 L/day.
Imaging of the Neurohypophysis

When MRI began to be used to evaluate the pituitary gland and hypothalamus, a bright spot in the sella was reported on T1-weighted images. The bright spot is due to stored hormone in neurosecretory granules in the posterior pituitary. The interest in the posterior pituitary bright spot as a diagnostic tool was heightened by reports that the bright spot was absent in patients with diabetes insipidus.

Many studies using small numbers of normal subjects have demonstrated this bright spot in all normal subjects; when larger numbers were evaluated, however, the bright spot was not seen in some normal subjects. In one study of 500 normal subjects, it was calculated that the bright spot would be present in 84% of normal youth. Although it had been suggested that the intensity of this bright spot might vary with the physiologic state of water balance in humans, this would occur only with a prolonged stimulus.

Absence of the bright spot is characteristic of diabetes insipidus, but some studies have reported the presence of a bright spot in patients with clinical evidence of diabetes insipidus. This may be of most interest in patients with familial hypothalamic diabetes insipidus (see later). In these cases, the posterior pituitary bright spot may be seen early in the disease (especially when the diabetes insipidus is partial) but usually disappears over time with increasing severity of the diabetes insipidus.

The role of stored oxytocin as a source of the pituitary bright spot has been largely ignored. Oxytocin is synthesized in the same nuclear groups and is transported and stored in the posterior pituitary in a manner similar to that with vasopressin. In humans, secretion of oxytocin is less responsive to changes in hydration. Therefore, it is possible that a persistent bright spot in patients with diabetes insipidus might be due to the pituitary content of oxytocin. Furthermore, oxytocinergic neurons are more resistant to destruction by trauma compared with vasopressinergic neurons in rats and humans.

The presence of a positive posterior pituitary bright spot has been variably reported in other polyuric disorders considered in the differential diagnosis of diabetes insipidus. The bright spot is usually seen in patients with primary polydipsia. This observation is consistent with studies in animals, in which even prolonged lack of secretion of vasopressin caused by hypotension did not produce a decreased content of hormone in the posterior pituitary. In nephrogenic diabetes insipidus, the bright spot has been reported to be absent in some patients but present in others. Because these patients have high levels of vasopressin in plasma and are chronically dehydrated, the posterior pituitary might be depleted of vasopressin and the bright spot might be absent.

Imaging of the hypothalamus is also an important diagnostic tool for diseases of the neurohypophysis. As noted earlier, the hormones of the neurohypophysis are synthesized in the paired paraventricular nuclei, located bilaterally in the walls of the third ventricle, and in the supraoptic nuclei, located at the extremes of the optic chiasm. When this anatomic information is coupled with the knowledge that 90% of the vasopressinergic neurons must be destroyed to produce symptomatic diabetes insipidus, it is apparent that for a mass lesion or a destructive lesion to produce diabetes insipidus, it either must destroy a large area of the hypothalamus or must be specifically located where the tracks converge in the base of the hypothalamus at the origin of the pituitary stalk. Furthermore, the hormones are synthesized in cell bodies quite distant from the site of release in the posterior lobe, and with section or damage of the axons at the level of the posterior lobe there is a reaccumulation of neurosecretory material and regeneration of a posterior lobe above the site of injury. Thus, tumors confined to the sella do not cause diabetes insipidus, and the area of interest is the discrete area immediately above the diaphragm sella at the base of the hypothalamus.

The pituitary stalk can also be readily identified by MRI and has been an additional tool in the differential diagnosis of diseases of the neurohypophysis. Enlargement of the stalk beyond 2 to 3 mm has been reported as pathologic. Metastatic tumors may be seen as enlargements of the pituitary stalk, probably related to seating of metastases in the long portal capillary system. Infiltrative diseases of the neurohypophysis, such as Langerhans cell histiocytosis, Wegener’s granulomatosis, and lymphocytic infundibulohypophysis, may enlarge the stalk. Sarcoidosis and tuberculosis are infiltrative lesions that can cause widening of the stalk. Even cases that remain idiopathic may involve enlargement of the stalk.

When the etiologic diagnosis of diabetes insipidus is in doubt and MRI reveals thickening of the stalk, especially with absence of the posterior pituitary bright spot, a search for systemic diseases is indicated. This search may result in a diagnosis of Langerhans cell histiocytosis, sarcoidosis, or tuberculosis. Further evaluation of cerebrospinal fluid (CSF) and plasma for secretion of human chorionic gonadotropin (hCG) and -fetoprotein may indicate suprasellar germinoma.

When a diagnosis is still in doubt, MRI should be repeated every 3 to 6 months, especially in children, in whom enlargement may indicate a germinoma. Decrease in size of the stalk with follow-up is more likely indicative of lymphocytic infundibulohypophysis or idiopathic diabetes insipidus (but many of these cases may be infundibulitis), although it may also occur with specific treatment of infiltrative diseases.

Finally, MRI may show formation of a new “posterior pituitary” after stalk transection. It has been known for many years that section of the neurohypophyseal stalk at a low level may produce transient diabetes insipidus with eventual return of function. Reaccumulation of neurosecretory material above the transection has been noted histologically. Indeed, the accumulation of neurosecretory products above the site of section was early evidence proving that neurosecretory material was transported along axons. Postoperative patients with diabetes insipidus or patients in whom the posterior pituitary was destroyed by compression of an adjacent anterior pituitary adenoma may “lose” the bright spot on MRI but may demonstrate reappearance of a bright spot at the level of the remaining stalk.
Clinical Syndromes of Hypothalamic Diabetes Insipidus

Hereditary Hypothalamic Diabetes Insipidus

Hereditary hypothalamic (central or neurohypophyseal) diabetes insipidus is characterized by the onset of classic diabetes insipidus, thirst, polydipsia, and polyuria in childhood, but during infancy those who carry the genetic defect may be asymptomatic. In contrast, in cases of familial nephrogenic diabetes insipidus, the defect is expressed as a polyuria disease at birth (see later). The relatively late onset of hereditary hypothalamic diabetes insipidus is also supported by MRI findings, which, although variable, have shown a positive bright spot suggesting vasopressin stores early in the disease but a loss of the bright spot (or a greatly diminished one) late in the disease.

More than 30 different families with autosomal dominant hypothalamic (neurohypophyseal) diabetes insipidus have been studied and the genetic defect identified. Only one defect has been described in the vasopressin gene itself, which was reported as an autosomal recessive with heterozygotic parents and late onset of the disease related to a biologically less active mutated vasopressin hormone. A number of families have been described with a genetic abnormality in the signal peptide of the pre-prohormone, most commonly at the extreme carboxyl terminus at the site of cleavage of the signal peptide from vasopressin (see Fig. 9-3). Disruption of the cleavage is thought to cause the disease. Except for the rare mutants involving the vasopressin hormone, all of the other defects have been in the neurophysin molecule. None have been reported for the glycopeptide.

Most authors have suggested that abnormalities of the folding of neurophysin might be toxic to the magnecellular neurons and that over time (consistent with the late onset of an autosomal dominant disease) may cause neuronal cell death. Of the few postmortem studies, some findings have been consistent with degeneration of magnecellular neurons but others have shown normal neurons with decreased expression of vasopressin or no hypothalamic abnormality.

The mutant deoxyribonucleic acid (DNA) of the vasopressin neurophysin precursor has been expressed in neurogenic cell lines, all showed abnormal trafficking and accumulation of mutant prohormone in the endoplasmic reticulum with difficulty with packaging into neurosecretory granules (see Fig. 9-3). The mechanism whereby this may lead to cell death has not been defined, but cell death may not be necessary to decrease available vasopressin.

Normally, proteins retained in the endoplasmic reticulum are selectively degraded, but if excess mutant is produced and the selective normal degradative process is overwhelmed, an alternative nonselective degradative system (autophagy) is activated. As more and more mutant precursor builds up in the endoplasmic reticulum, the normal wild type is trapped with the mutant protein and degraded by the activated nonspecific degradative system. By this mechanism, the amount of vasopressin that matures and is packaged is markedly reduced. This explanation is consistent with the cases in which little pathology is found in the magnecellular neurons and also with cases of diabetes insipidus in which some small amount of vasopressin can be detected.

Diabetes Insipidus Produced by Solid Tumors or Hematologic Malignancies

Some tumors such as craniopharyngioma and suprasellar germinoma or pinealoma characteristically occur in a suprasellar basal hypothalamic area and are regularly associated with diabetes insipidus. The latter are often diagnosed by accompanying precocious puberty or serum markers such as hCG or -fetoprotein in spinal fluid or plasma. It is not uncommon with pinealomas and suprasellar germinomas for diabetes insipidus to be the presenting complaint, although other evidence of hypopituitarism may be present. MRI may not demonstrate a mass in the suprasellar area for a few months.

Metastatic disease involving the pituitary is usually found in association with widespread metastatic disease and is reported at autopsy but is often not symptomatic during life. Metastases are twice as likely to involve the posterior pituitary as the anterior pituitary, which is thought to be due to a more direct arterial blood supply to the posterior pituitary. It is also possible that any potential metastases to the anterior pituitary lodge in the portal system and occur as hypothalamic tumors. In either case, with metastatic tumors, diabetes insipidus is more common than is deficiency of anterior pituitary hormones.

The diagnosis is usually made in a patient who is known to have primary cancer with metastases elsewhere. MRI of the brain usually demonstrates the pituitary metastasis, often with other metastases in the brain or skull. Occasionally, only micrometastases are found at autopsy and enlargement of the stalk may be the presenting finding. Most primary tumors in the hypothalamic-pituitary area that cause diabetes insipidus grow relatively slowly, and any tumor in this area that shows rapid growth in a short period of time should be considered a possible metastatic tumor. Carcinoma of the breast is the most common primary cancer in women, and carcinoma of the lung is the most common primary tumor in men. Other tumors that have been reported in the area include adenocarcinoma of the stomach, pancreas, uterus, thyroid, and bladder.

Diabetes insipidus has been reported with lymphomas in the hypothalamic-pituitary area. Usually, lymphoma is recognized elsewhere, but rarely it is a primary central nervous system (CNS) lymphoma. There may be an increased incidence of lymphoma with diabetes insipidus because of the increased incidence of lymphoproliferative disease with human immunodeficiency virus (HIV) and hepatitis C infection.

Diabetes insipidus is also associated with leukemia. The mechanism is thought to be infiltration of the hypothalamus, thrombosis, or infection. Although acute lymphocytic leukemia is as common as nonlymphocytic leukemia and is well known to involve the CNS, diabetes insipidus is distinctly more common with nonlymphocytic leukemia. As many as 75% of the cases of diabetes insipidus with leukemia involve nonlymphocytic leukemia. There is also a suggested association with monosomy 7, although a mechanism has not been defined. MRI results in leukemia may show infiltration or an infundibular mass but are often normal even when leukemic cells are found in CSF. In other cases, the CSF has no leukocytes and thrombosis of small vessels in the hypothalamus might be a more likely cause of the diabetes insipidus. Posterior pituitary deficiency may be associated with panhypopituitarism, and the diabetes insipidus may not be apparent because of coexisting adrenal insufficiency and hypothyroidism. Indeed, in some patients symptomatic diabetes insipidus occurs only when prednisone therapy is initiated as treatment for the leukemia.

Response of the Neurohypophyseal System to Surgery or Trauma

Although diabetes insipidus is well known to occur after hypothalamic-pituitary surgery, this diagnosis should be made with caution. Vasopressin is normally secreted in the stress of surgery, and fluid may be retained and then normally excreted after surgery (see "Differential Diagnosis"). The stress of surgery may also induce insulin resistance and may exacerbate diabetes mellitus, producing an osmotic diuresis resulting from glucose. The patterns of diabetes insipidus after surgery have been described in detail. As many as 50% to 60% of patients have some transient diabetes insipidus within 24 hours of pituitary surgery that usually resolves, especially with transphenoidal surgery in which the resection of a tumor is confined to the sella.

If there is a complete section of the stalk, patients may exhibit a pattern known as triphasic diabetes insipidus (Fig. 9-7):
The second (antidiuretic) phase, although originally described as a normal interphase, is not normal and is thought to be due to unregulated release of vasopressin from the store of hormone in the axons of the posterior pituitary as these axons degenerate. Because the release of vasopressin in this phase is unregulated, excess administration of fluids produces hyponatremia as in other forms of SIADH.

The third phase, the return of diabetes insipidus, occurs when all of the hormone has been released from the posterior pituitary. The course of diabetes insipidus may be permanent or may subsequently resolve to partial or clinically inapparent disease.

Magnicellular neurons are unique in that, after the axons are sectioned, the neurons survive and there is outgrowth of dendrites and regeneration of new axons. A factor contributing to the ability of magnicellular neurons to regenerate is the close association of these neurons with specialized glial cells. The glial cells in the area of magnicellular neurons and the median eminence synthesize and release growth factors that may stimulate nerve growth. The newly formed axons grow along fixed glial cells (tanycytes) that span the median eminence from the third ventricle to the external zone of the median eminence. Thus, the regenerating axons and spraying axons create neurosecretory processes in the CSF of the third ventricle as well as in the perivascular region of the external zone of the median eminence. The transected neurosecretory axons may promote capillary sprouting, but its importance in the return of function is uncertain. These newly formed capillaries may have tight interendothelial junctions similar to those elsewhere in the brain, whereas the capillaries that are already present in the external zone of the median eminence are fenestrated capillaries capable of receiving secreted peptides. Several studies have shown that oxytocin neurons survive better than vasopressin neurons after transection of the pituitary stalk. In studies in the rat, the activity of the magnicellular vasopressin neurons had a dramatic effect on recovery. After stalk compression in the rat, if synthesis of vasopressin was inhibited, fewer vasopressin neurons survived, and if vasopressin synthesis was stimulated, more vasopressin neurons survived. However, because the magnitude and duration of both the hyponatremia and the hypotension in the rat studies exceeded those seen in patients, the application of the findings to clinical medicine is uncertain.

An important observation is that the second phase of the triphasic response—uncontrolled release of vasopressin related to axon trauma—may occur without preceding or subsequent diabetes insipidus. This has been observed clinically and has been produced experimentally in the rat by unilateral lesion of the supraoptichypophyseal tract. The interpretation is that if the trauma is only to some of the axons coursing to the posterior pituitary, the remaining intact axons have sufficient vasopressin function to avoid clinically apparent diabetes insipidus characteristic of the first and third phases of the triphasic response. However, the store of hormone in the posterior pituitary is sufficiently large that regeneration and nerosis of even a fraction of these vasopressin neurons causes enough uncontrolled release of vasopressin to produce hyponatremia if excess fluid is administered. The hyponatremia becomes apparent because it is often symptomatic with new-onset headache, nausea, and emesis. When all the vasopressin from the damaged neurons has been secreted, the stimulus for water retention resolves and the retained water is excreted, resulting in recovery from the hyponatremia. Thus, the clinical picture is one of hyponatremia occurring about 7 days after pituitary surgery, persisting for a few days, and then returning to normal. This syndrome of transient hyponatremia has been referred to as an isolated second phase to emphasize the pathophysiologic etiology. In one series, isolated hyponatremia occurred in as many as 25% of patients after pituitary surgery. The hyponatremia was associated with lack of suppression of vasopressin, inability to excrete a water load, and inappropriate natriuresis and was observed in spite of normal levels of cortisol or glucocorticoid replacement. In larger series, including various sizes and etiologic mechanisms of tumors, isolated hyponatremia was reported in about 10%, with only 2% symptomatic.

The same patterns of diabetes insipidus that occur after surgery can be seen in patients after closed-head trauma. Seventy-five percent of these cases are due to factors that do not involve the vasopressin neurons[219] and there is a great preponderance of male patients, usually young men with a mean age in the 20s. More than 90% of patients experience coma and a high percentage have associated skull fracture. Computed tomography (CT) or MRI in a large group of patients with post-traumatic hypopituitarism including diabetes insipidus showed hemorrhage in the hypothalamus or posterior pituitary in 55% and stalk resection or infarction of the posterior pituitary in approximately 6%.

Several important clinical points should be emphasized concerning diabetes insipidus induced by head trauma:

1. These patients are virtually always unconscious and do not have the normal ability to sense thirst.
2. In this situation, large volumes of fluid may be given because of blood loss or other volume deficits; this fluid loss or stress may induce diabetes mellitus and an osmotic diuresis (see “Differential Diagnosis”).
3. There may be a greater risk if the second phase is unrecognized because hyponatremia may produce cerebral edema and worsen any edema related to trauma. Therefore, in administering desmopressin, the effect of one dose should be allowed to wane before another dose is administered in order to ensure that the patient has not entered the second phase.
4. There is a high incidence of anterior pituitary deficiency in association with diabetes insipidus induced by head trauma.

The possibility of cortisol deficiency should be considered immediately, as it may be life-threatening in these patients. It is also well known that anterior pituitary deficiency and, especially, decreased ACTH and adrenal function interfere with the ability to dilute the urine maximally. Cortisol deficiency should also be considered subsequently if diabetes insipidus appears to improve because of a decrease of water excretion in the absence of an administered antidiuretic agent.

Finally, in the long-term follow-up of these patients, the possibility of late development of anterior pituitary deficiency should be kept in mind as well as the possible return of sufficient vasopressin function that the patient no longer has symptomatic diabetes insipidus.

Granulomatous Diseases
Langerhans Cell Histiocytosis

The term Langerhans cell histiocytosis is now applied to a spectrum of diseases from the severe fulminant visceral Letterer-Siwe disease to the multifocal Hand-Schüller-Christian disease to benign eosinophilic granuloma. The etiology is unknown, but the condition is characterized by proliferation of monoclonal Langerhans cells. It may have an acute fulminant course or be marked by spontaneous remission and recurring disease.

In patients with Langerhans cell histiocytosis, diabetes insipidus occurs as a manifestation of CNS involvement and usually in association with other involvements of the head, including cranial bones, oral mucosa, or other areas of the brain. Diabetes insipidus is also more common when there is systemic disease, especially involving the lung, but occasionally diabetes insipidus may be the only systemic manifestation other than diffuse involvement of the skin and may even precede the diagnosis of Langerhans cell histiocytosis. A variety of disorders of water balance may be produced, including complete hypothalamic diabetes insipidus (the most common), partial hypothalamic diabetes insipidus, abnormalities of thirst, and the disorder of essential hypernatremia.

The reported frequency of diabetes insipidus with Langerhans cell histiocytosis depends on whether it is routinely sought by endocrine tests and MRI of the brain or noted only on the basis of the diagnosis in retrospective series. Higher incidences were reported in earlier papers, possibly because the new definition of Langerhans...
diabetes insipidus is due solely to oxytocinase. Instead, these patients must be evaluated to establish an etiologic diagnosis. Because of the accelerated metabolic clearance of vasopressin in pregnancy, symptomatic diabetes insipidus may also develop in patients with borderline

obvious cause of diabetes insipidus have been monitored and have shown regression of the thickened pituitary stalk and tumor-like appearance.

patients underwent operations because of a suspicion of a pituitary tumor. When hypophysal diabetes insipidus is found, MRI usually demonstrates absence of the pituitary bright spot and widening of the stalk. As with other disorders, the degree of involvement is better demonstrated with the administration of contrast agents but occasionally the MRI is completely normal and the pathology is thought to be a vasculitis that is not apparent on imaging studies.

A number of cases have shown complete resolution of diabetes insipidus with appropriate therapy and response of Wegener’s granulomatosis; others have demonstrated persistent diabetes insipidus in spite of response of peripheral manifestations of the disease and even decreased granulomatous lesions in the hypothalamic. Other less specifically defined granulomatous diseases of the periphery have also rarely been reported with hypophysal diabetes insipidus; however, when diabetes insipidus and abnormalities of the pituitary stalk and posterior pituitary are the only findings, these nonspecific inflammatory disorders may be indistinguishable from infundibulohypophysis.

Sarcoidosis

Although only about 5% of patients with sarcoidosis have symptoms of neurosarcoidosis, neurologic symptoms are the presenting complaint in about 50% of those patients, and 25% of the patients with neurosarcoidosis have hypophysal diabetes insipidus. MRI findings may mimic those with other causes of hypophysal diabetes insipidus and consist of a widened stalk with contrast and absence of a posterior pituitary bright spot. Thus, in cases considered as idiopathic diabetes insipidus or infundibuloneurohypophysis, sarcoidosis should be considered and possible systemic manifestations of the disease should be sought. The study would include at a minimum chest radiography, erythrocyte sedimentation rate (ESR), and serum and CSF levels of angiotensin-converting enzyme and serum calcium. Tuberculosis should also be considered and a PPD test done.

Because early treatment of neurosarcoidosis is recommended, it is desirable to make the diagnosis. Although hypophysal diabetes insipidus is the most common water balance problem in sarcoidosis, disordered regulation of thirst and nephrogenic diabetes insipidus related to the disease or hypercalcemia have also been noted. Whereas neurosarcoidosis may be treated with a variety of mechanisms leading to remission of the disease, diabetes insipidus, once established, is usually permanent.

Lymphocytic Infundibulohypophysitis

When obvious causes of diabetes insipidus are not present, most cases of diabetes insipidus are idiopathic. As in other endocrine systems in which loss of function is not associated with a specific etiology, the possibility of an autoimmune process has been considered. Vasopressin antibodies have been reported in serum in up to one third of patients with idiopathic diabetes insipidus and in two thirds of those with Langerhans cell histiocytosis but were absent in patients with tumors. Furthermore, a relatively high incidence of other autoimmune diseases have been reported. In one study, 878 patients with autoimmune endocrine diseases, but without hypophysal diabetes insipidus, were screened for vasopressin antibodies and 9 patients were found to have them. With careful testing, four of these patients were found to have partial diabetes insipidus and five were normal. After a 4-year follow-up, though, three of the normal subjects had experienced partial diabetes insipidus.

A rare but now well-recognized cause of autoimmune diabetes insipidus is lymphocytic infundibulohypophysitis. Lymphocytic infiltration of the anterior pituitary, lymphocytic hypophysitis, has been recognized as a cause of anterior pituitary deficiency for a number of years, but it was not until an autopsy called attention to a similar finding in the posterior pituitary of a patient with diabetes insipidus that this pathology was recognized for the neurohypophysis. Since that report, a number of cases have been described, including cases in the postpartum period, which is characteristic of lymphocytic hypophysitis. Since the advent of MRI, lymphocytic infundibulohypophysitis has been diagnosed on the basis of the appearance of a thickened stalk or enlargement of the posterior pituitary mimicking a pituitary tumor, or both. In these cases, the characteristic bright spot on T1-weighted MRI images is lost.

Enlargement of the stalk so resembled pituitary tumor that before lymphocytic infundibulohypophysitis was known to produce stalk enlargement, some of these patients underwent operations because of a suspicion of a pituitary tumor. Recently, a number of patients with suspected infundibulohypophysitis and no other obvious cause of diabetes insipidus have been monitored and have shown regression of the thickened stalk or tumor-like appearance. Treatment of these patients with prednisone may be associated with a decrease in size of the stalk, but a decrease may also occur spontaneously.

Some cases show coexistence of infundibulohypophysitis and adenohypophysitis. Autoimmune diseases may affect more than one endocrine organ without any specific link to explain the association, such as diabetes insipidus in association with systemic lupus erythematosus or Behçet’s disease.

Diabetes Insipidus in Pregnancy

In rare cases, women with normal regulation of vasopressin have symptoms of diabetes insipidus during pregnancy because of extremely elevated activity of oxytocine aminopeptidase (vasopressinase). This syndrome has been referred to as vasopressinase-resistant diabetes insipidus of pregnancy. Levels of vasopressinase are markedly elevated above levels in normal pregnancy in such patients, and the concurrence of preeclampsia, acute fatty liver, and coagulopathies has been noted. That the excess vasopressinase is due to the condition of the pregnancy (i.e., preeclampsia) is further shown by the report that subsequent pregnancies of these women are uncomplicated by diabetes insipids or acute fatty liver. It has been suggested that the products of vasopressinase degradation produced by vasopressinase might be biologically active in increasing blood pressure, but during testing they were not thought to be causative in the preeclampsia.
insipidus is specifically related to the pregnancy (e.g., as when it occurs with Sheehan's syndrome and infarction of the neurohypophysis). One might anticipate an increased incidence of lymphocytic infundibuloneurohypophysitis in pregnancy because other autoimmune diseases occur during or just after pregnancy. Although such cases have been reported, there does not appear to be a significant association of infundibulohypophysitis with pregnancy.

Desmopressin is clearly the treatment of choice for diabetes insipidus in pregnant women (see "Treatment of Diabetes Insipidus"). In general, labor and parturition proceed normally, and patients have no trouble with lactation. There is a possibility of chronic and severe dehydration when diabetes insipidus is unrecognized, and this may pose a threat in pregnant women. In one case of severe oligohydramnios found to be due to unrecognized diabetes insipidus, the patient responded promptly to therapy with desmopressin.

Essential Hypernatremia

One variant of diabetes insipidus is the syndrome of absent osmostat with intact baroreceptors. Because of the dysfunction of the osmostat, patients do not sense thirst and do not drink water. Unlike the situation with normal subjects, however, as the serum sodium level rises, there is no (or there is a markedly subnormal) release of vasopressin and a hypotonic polyuria continues. Even when patients are made euvoletic with the infusion of normal saline and the serum sodium concentration is allowed to rise to high levels, appropriate secretion of vasopressin is not present. Vasopressin is synthesized and stored, however, because maneuvers to stimulate baroreceptors cause secretion of vasopressin and concentration of the urine.

The proposed pathophysiologic mechanism is that the inadequate water intake and excess water excretion produce a degree of dehydration with hypernatremia. When dehydration is sufficient to stimulate the baroreceptors, vasopressin is released, urine is concentrated, and patients remain in a steady state of hypernatremia with modest dehydration. The increased concentration of sodium per se also causes sodium excretion to help maintain the new steady state.

Infectious Diseases

Infectious granulomatous diseases (tuberculosis, syphilis) are rare causes of diabetes insipidus in the United States. Occasional cases of tuberculosis are still reported, however, and it is important to consider the diagnosis, because a tuberculoma in the suprasellar area may mimic an endocrine tumor or sarcoidosis. It is desirable to make the diagnosis on the basis of peripheral manifestations or CSF findings rather than at surgery.

Diabetes Insipidus and Brain Death

Endocrine failure has been a topic of several clinical reports of brain death. Diabetes insipidus is virtually universal in animal models of brain death. It is extremely common in clinical series, with the incidence ranging from 30% to 80%. higher in series involving only adults than in series including children, and obviously dependent on the rigor with which the diagnosis is sought in patients with diuresis. Important clinical considerations are whether coexistent anterior pituitary deficiency exists, especially whether steroids should be administered, and whether the diabetes insipidus should be treated.

The subject of treatment is controversial. Although there is no firm evidence that treating the diabetes insipidus affects the quality of donor organs, there is also no evidence that treating diabetes insipidus results in any complications. It is important to bear in mind that not all brain death is accompanied by diabetes insipidus and that diabetes insipidus in a comatose patient does not necessarily connote brain death and even severely injured comatose patients may survive (see earlier discussion of diabetes insipidus after head injury). The treatment of diabetes insipidus with brain death is similar to that of acute head injury.
Primary Polydipsia

Primary polydipsia and subsequent polyuria must be differentiated from diabetes insipidus and may also contribute to SIADH. Primary polydipsia may be induced by an organic structural lesion in the hypothalamus identical to that in any of the disorders described as causes of diabetes insipidus and may be especially associated with sarcoidosis of the hypothalamus. It may also be produced by drugs that cause a dry mouth or by any peripheral disorder causing an elevation of renin or angiotensin, or both. If the pathologic etiology is not identifiable, the disorder may be habitual throughout a lifetime or, more commonly, may be associated with psychiatric syndromes.

Series involving patients with polydipsia in psychiatric hospitals have shown an incidence as high as 42% of patients with some form of polydipsia, and for more than 50% of these, the cause of the polydipsia remained unexplained. In some cases, the unexplained presence of polydipsia represented a resetting of the osmotic threshold for thirst independent of the osmotic threshold for vasopressin and these patients responded to administered desmopressin producing SIADH and hyponatremia. Usually, the patient’s condition was refractory to attempts to restrict fluid. Propranolol has been used with some success, presumably because of its ability to inhibit the renin-angiotensin system.
Clinical Syndromes of Nephrogenic Diabetes Insipidus

Congenital Nephrogenic Diabetes Insipidus

Babies with nephrogenic diabetes insipidus present with vomiting, constipation, failure to thrive, fever, and polyuria. Symptoms usually occur during the first week of life, and on testing the patients are found to have hyponatremia and a low urine osmolality. Historically, the patients were reported to have mental retardation, complications of urinary tract dilation, and intracranial calcification. An association between intracranial calcification and mental retardation has not been established, and the calcification may be a consequence of hyponatremia rather than of congenital disease.

The diagnosis is established by high levels of vasopressin in the plasma in the presence of hypotonic polyuria and, subsequently, by the absence of a response to administered desmopressin. There are two causes: (1) mutation in the V2 receptor and (2) mutations in the aquaporin-2 water channels; however, the presentation of diabetes insipidus is independent of the genotype. In patients with the aquaporin defect, one can detect other V2 responses (e.g., stimulation of factor VIII secretion).

As genetic testing becomes more widely available, it may be unnecessary to perform extensive clinical testing to establish the diagnosis; instead, treatment may be initiated when the syndrome is suspected and direct genetic testing of the child and of the parents performed.

Nephrogenic Diabetes Insipidus Caused by Mutation of the V2 Receptor

More than 90% of cases of congenital nephrogenic diabetes insipidus are X-linked disorders occurring in males and are caused by one of more than 132 different V2 receptor mutations. Most mutations occur in the part of the receptor that is highly conserved among species or that is conserved among similar receptors (e.g., homology with V1 vasopressin receptors or oxytocin receptors). Mutations at other locations on the gene may cause mild abnormalities of urinary concentration, but mild defects in concentrating ability may not cause polyuria and may escape clinical detection.

Most reported receptor abnormalities produce disruption of protein folding of the V2 receptor, and the receptor is trapped in the endoplasmic reticulum, destroyed in the cell, or improperly inserted into the cell membrane. In a few cases, the receptors reached the cell membrane and were inserted; the defect was due to inhibition of binding of vasopressin to the receptor or to improper signaling of G proteins and subsequent generation of cAMP. When the V2 receptor is inserted in the cell membrane, clotting and expression of the mutant gene in cell culture may show that the receptor responds to high concentrations of vasopressin or desmopressin, yet only a few of these are responsive to physiologic levels of vasopressin to express congenital nephrogenic diabetes insipidus as a partial disorder.

In clinical series, approximately 10% of the V2 receptor defects causing congenital nephrogenic diabetes insipidus are thought to be de novo. This high incidence of de novo cases, coupled with the large number of mutations that have been identified, hinders the clinical use of genetic identification because it is necessary to sequence the entire open reading frame of the receptor gene rather than short sequences of DNA. This degree of sophistication exceeds the level of expertise that is found in most clinical laboratories.

Although most female carriers of the X-linked V2 receptor defect have no clinical disease, some females were reported to have symptomatic nephrogenic diabetes insipidus. Carriers may have a decreased maximum urine osmolality in response to the plasma level of vasopressin that they obtain but are asymptomatic because of normal urine volume. Female carriers respond to desmopressin with an increase in factor VIII, although the response may be only a fraction of the normal response. Some girls have as severe a defect as boys, and this is thought to be due to inactivation of the normal X chromosome.

In some cases, there may be some heterogeneity of X inactivation in various tissues, so that there is symptomatic nephrogenic diabetes insipidus but a normal response to stimulate factor VIII.

Nephrogenic Diabetes Insipidus Caused by Mutation of Aquaporin-2

When the proband is a girl, it is likely that the defect is a mutation of the aquaporin-2 water channel gene on chromosome 12, q1213 producing an autosomal recessive disease. This should be especially considered when consanguinity is known in the family and the family history shows disease expressed in men and women. The phenotype of nephrogenic diabetes insipidus is identical to the receptor defect when the abnormality is in the aquaporin-2 protein. More than 20 different mutations of the aquaporin-2 protein have been described. The patients may be heterozygous for two different recessive mutations or homozygous for the same abnormality from both parents.

The biologic mechanism responsible for the defect in the aquaporin proteins may be related to the site of the mutation. In some mutations, aquaporins are not properly processed in the endoplasmic reticulum or released into the cytoplasm. In other mutations, aquaporins are processed but are not appropriately inserted in the cell membrane. Mutations in the C-terminal portion of the aquaporin-2 protein produce an autosomal dominant nephrogenic diabetes insipidus. A mutant aquaporin-2 protein produces oligomers with the wild-type protein in the Golgi apparatus, and no functional aquaporin-2 proteins are produced. This form is similar to the autosomal dominant form of congenital hypothalamic diabetes insipidus described earlier.

Acquired Nephrogenic Diabetes Insipidus

The ability to produce a concentrated urine depends on maintaining hyperosmolality of the inner medulla of the kidney. The following conditions are necessary to produce and maintain hyperosmolality of the inner medulla:

1. An intact kidney architecture, with an intact tubular structure of the descending limb and ascending limb of the loop of Henle (essential to the development of the countercurrent multiplier) and a normal anatomy of the collecting duct to pass back through the inner medulla.
2. Active sodium transport in the thick ascending limb and functional aquaporins, to allow water transport across membranes.
3. An anatomically intact vascular structure so that the hyperosmolality of the inner medulla is not washed away by normal blood flow.

Acquired nephrogenic diabetes insipidus without the capacity to produce a concentrated urine may be caused by numerous chronic renal diseases that distort the architecture of the kidney (e.g., polycystic kidney disease, renal infacts with neovascularization, such as produced by sickle cell anemia; and infiltrative disease of the kidney). Washout of the medullary gradient may be produced by excessive water intake of primary polydipsia or by forced sodium and water loss related to administration of a diuretic.

A low-protein diet may be associated with reduced medullary urea concentration, causing inability to produce maximum concentration of the urine. Aquaporin is decreased in chronic renal failure, and other more specific disorders of acquired nephrogenic diabetes insipidus are related to abnormalities of aquaporins.

Studies have shown that the polyuria associated with potassium deficiency develops in parallel with decreased expression of renal aquaporin-2. Repletion of potassium reestablished the normal urinary concentrating mechanism and normalized the expression of aquaporin-2. Hypercalcemia was also associated with down-regulation of aquaporin-2. A low-protein diet diminished the ability to concentrate the urine primarily by decreased delivery of urea to the inner medulla, thus decreasing the hypertonicity of the medulla, but rats on a low-protein diet also down-regulated aquaporin-2, which may be an additional component of the decreased ability to concentrate the urine. Bilateral urinary tract obstruction caused inability to produce a maximum concentration of the urine and rat models demonstrated a down-regulation of aquaporin-2 that persisted for several days after release of the obstruction. In addition, the washout of accumulated waste...
products contributed to postobstructive polyuria.

Drug-Induced Nephrogenic Diabetes Insipidus

Administration of lithium to treat psychiatric disorders is the most common cause of drug-induced nephrogenic diabetes insipidus and illustrates the mechanisms. Lithium administration of lithium to treat psychiatric disorders is the most common cause of drug-induced nephrogenic diabetes insipidus and illustrates the mechanisms. Lithium is known to interfere with the production of cAMP, and it has been demonstrated in studies of animals that lithium produces a dramatic reduction in aquaporin-2 levels. The defect is both of aquaporin-2 on the luminal surface of the collecting duct and of aquaporin-3 on the basal lateral membrane, producing a severe defect in aquaporin-2 levels. The defect is both of aquaporin-2 on the luminal surface of the collecting duct and of aquaporin-3 on the basal lateral membrane, producing a severe defect in aquaporin-2 levels. There is as much as a 95% decrease in aquaporin-2 content, and even the 5% of aquaporin-2 that persists is not normally transported to the collecting duct membrane. Sodium transporters either show no change or are up-regulated in an effort to compensate for the water loss. The defect of aquaporins is slow to correct in both experimental animals and humans, may be permanent, and may be associated with glomerular or tubulointerstitial nephropathy. Lithium may also cause nephrogenic diabetes insipidus by inducing hypercalcemia. Other drugs that are known to induce renal concentrating defects or disorders such as nephrotic syndrome may be associated with abnormalities of aquaporin synthesis. Demeclocycline is another drug known to cause nephrogenic diabetes insipidus and is used clinically to treat SIADH (discussed later). See Bendz and Aurell for a list of drugs also known to induce diabetes insipidus.

Deficient Vasopressin

There may be an element of nephrogenic diabetes insipidus associated with severe central diabetes insipidus related to decreased aquaporin-2. Indeed, even physiologic suppression of vasopressin by chronic administration of water may produce down-regulation of aquaporin-2 in the renal collecting duct. As discussed earlier in this chapter, when a dehydration test is followed by administration of desmopressin to differentiate various causes of diabetes insipidus, the ability to concentrate the urine maximally is impaired both for central diabetes insipidus and for primary polydipsia. Although this has long been attributed to washout of the medullary concentration gradient, part of the decreased response to vasopressin is due to the down-regulation of aquaporin-2. In addition, it takes some time to restore normal expression of aquaporin-2, and this may contribute to the long time it takes patients with primary polydipsia and central diabetes insipidus to achieve maximum concentration of urine after water restriction (or treatment) is initiated.
Treatment

Patients with diabetes insipidus and inadequate thirst can rapidly become dehydrated and may experience severe hypernatremia with devastating effects on the CNS. Hypertonic encephalopathy with obtundation, coma, and seizures may be produced by brain shrinkage. A decreased volume of brain in the skull may lead to subarachnoid hemorrhage, intracerebral bleeding, or petechial hemorrhage. Fortunately, these problems associated with severe hypernatremia are not observed in patients with diabetes insipidus who are ambulatory and have an intact thirst mechanism. Furthermore, they are not a part of the syndromes of hypodipsia (e.g., essential hypernatremia) in ambulatory patients, probably because of the chronic and slower onset of hypernatremia in these cases. Thus, although hypernatremia is a commonly observed clinical

| TABLE 9-1 – Therapeutic Agents for Treatment of Diabetes Insipidus |

<table>
<thead>
<tr>
<th>Water-Retaining Agents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arginine vasopressin</td>
</tr>
<tr>
<td>1-(3-Mercaptopropionic acid)-d-arginine vasopressin</td>
</tr>
<tr>
<td>Chlorpropamide</td>
</tr>
<tr>
<td>Carbamazepine</td>
</tr>
<tr>
<td>Clonidine</td>
</tr>
<tr>
<td>Indomethacin</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Natriuretic Agents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thiazide diuretics</td>
</tr>
<tr>
<td>Amiloride</td>
</tr>
<tr>
<td>Indapamid</td>
</tr>
</tbody>
</table>

*Not recommended.*

event in hospitalized patients. In most cases the dehydration is due to increased insensible water loss or gastrointestinal loss in elderly or young patients at admission (e.g., with fever) rather than diabetes insipidus. Hypernatremic encephalopathy is a risk in patients with diabetes insipidus only when patients cannot respond to thirst because of either age or level of consciousness. Patients with diabetes insipidus should carry a medical card and wear a medical alert tag indicating that they have this disorder; because diabetes insipidus is rare, the name of a physician who is familiar with the disorder and whom one can notify in an emergency should be included.

Most patients with diabetes insipidus have intact thirst and drink sufficient fluid to maintain relatively normal fluid balance. The absence of vasopressin per se does not produce pathology, and a major goal of therapy is to decrease the thirst and polyuria to an acceptable level. The therapeutic regimen should be easy for the patient to accommodate, and the timing and quantity of dosage should be individually prescribed. The safety of the prescribed agent and a regimen that avoids any detrimental effects of overtreatment are primary considerations because of the relatively benign course of diabetes insipidus and the adverse consequences of hyponatremia.

The therapeutic agents used to treat diabetes insipidus are shown in Table 9-1. Water is considered a therapeutic agent for this disease; when water is taken in sufficient quantity, there is no metabolic abnormality. As noted, therapy is designed to reduce the necessary water intake (and polyuria) to an acceptable level, but occasional lapses in pharmacologic therapy are not detrimental, may avoid overtreatment producing hyponatremia, and may allow recognition of any spontaneous recovery.

Chronic Severe Hypothalamic Diabetes Insipidus

In severe diabetes insipidus, there is virtually no vasopressin in the circulation and no ability to concentrate the urine with dehydration. An antidiuretic hormone is required, and the drug of choice is desmopressin. In this synthetic analogue, the substitution of d-arginine markedly reduced pressor activity, and removing the terminal amine increased the half-life. The two changes produced an agent nearly 2000 times more specific for antidiuresis than naturally occurring vasoressin. Most patients prefer the desmopressin tablets (0.1 and 0.2 mg), although many patients continue to be treated successfully with the desmopressin intranasal spray.

Because of the variability among patients, it is desirable to determine the duration of action of individual doses in each

---

patient. First the patient is allowed to be cleared from the effects of any previous medication, and for each voided urine the time can be recorded and the volume and osmolality measured. The patient is allowed to drink fluid ad libitum. A dose is administered; a decrease in urine volume is noted in 1 to 2 hours, and the total duration of action usually ranges from 6 to 18 hours. A satisfactory schedule can usually be determined with a modest dose, and the maximum dose needed is rarely above 0.2 mg orally or 20 µg (two sprays) given two or three times a day (usually two). Tablets allow considerable flexibility in dosage because they can be used either whole or split. For intranasally administered desmopressin, there is less flexibility with the metered spray, which is fixed at 10 µg in 100 µL. If more flexibility is necessary, the patient should be taught to use the nasal catheter (see Robinson and Verbalis for specific directions). When a dose is sufficient to elicit a stable therapeutic response, further increasing the dose (e.g., doubling it) produces only a moderate increase in duration of few hours. Tablets consistent with the half-life of desmopressin in plasma. The medication is expensive, and in many patients a smaller dose given more often is more cost-effective than a larger dose given less often. Rarely is it necessary to resort to parenterally administered desmopressin (2-mL vials, 1-mL ampules, or 10-mL vials of 4 µg/mL) for ambulatory patients. If an intercurrent illness or allergy is observed in patients with diabetes insipidus who are ambulatory and have an intact thirst mechanism. Furthermore, they are not a part of the syndromes of hypodipsia (e.g., essential hypernatremia) in ambulatory patients, probably because of the chronic and slower onset of hypernatremia in these cases. Thus, although hypernatremia is a commonly observed clinical

---

Hyponatremia is a rare complication of desmopressin therapy and occurs only if the patient is continually antidiuretic while maintaining a fluid intake sufficient to become volume-expanded and natriuretic. Thirst may be protective, and most patients with diabetes insipidus receiving standard therapy may not be continuously

---

The medication is expensive, and in many patients a smaller dose given more often is more cost-effective than a larger dose given less often. Rarely is it necessary to resort to parenterally administered desmopressin (2-mL vials, 1-mL ampules, or 10-mL vials of 4 µg/mL) for ambulatory patients. If an intercurrent illness or allergy is observed in patients with diabetes insipidus who are ambulatory and have an intact thirst mechanism. Furthermore, they are not a part of the syndromes of hypodipsia (e.g., essential hypernatremia) in ambulatory patients, probably because of the chronic and slower onset of hypernatremia in these cases. Thus, although hypernatremia is a commonly observed clinical

---

Hyponatremia is a rare complication of desmopressin therapy and occurs only if the patient is continually antidiuretic while maintaining a fluid intake sufficient to become volume-expanded and natriuretic. Thirst may be protective, and most patients with diabetes insipidus receiving standard therapy may not be continuously
maximally antidiuretic. Hyponatremia has been reported in patients with normal thirst and normal vasopressin function after taking desmopressin for treatment of von Willebrand’s disease and in children treated for primary enuresis. Hyponatremia is often indicated first by the onset of convulsions and coma. Hyponatremia in patients with diabetes insipidus who are receiving desmopressin may be avoided by ensuring, either with routine therapy or by occasionally delaying a dose of desmoppressin, that polyuria recurs once or twice a week, so that any excess fluid can be excreted.

Chronic Partial Hypothalamic Diabetes Insipidus

Patients with partial diabetes insipidus have some ability to secrete vasopressin and to concentrate the urine, but the function is inadequate to maintain normal water intake and urine output. These patients are more likely to respond to therapeutic agents such as chlorpropamide or thiazide diuretics because only a modest decrease in urine volume may make them asymptomatic. The major action of chlorpropamide is on the renal tubule to increase the hydro-osmotic action of vasopressin. The agent can also produce significant antidiuresis even in patients with severe hypothalamic diabetes insipidus.

The usual dose is 250 to 500 mg/day, with a response noted in 1 to 2 days and a maximum antidiuresis in 4 days. This is an off-label use of the drug. This agent should not be used in pregnancy and is not recommended for children, especially those with concurrent hypopituitarism, because of the possibility of severe hypoglycemia. These patients also respond to desmopressin; guidelines for initiating and maintaining therapy are essentially the same as those described earlier for patients with severe diabetes insipidus.

Chronic Diabetes Insipidus with Inadequate Thirst

The presence of diabetes insipidus in a patient with inadequate thirst is a difficult management problem because patients with lack of thirst can experience severe hyponatremia and, if given an antidiuretic agent and encouraged to drink, can become hypotensive. Thus, these patients are subject not only to wide swings in osmolality but also, most characteristically, to persistent hyponatremia. The spectrum of disorders includes that described as essential hyponatremia.

The first therapeutic agent to try is chlorpropamide; it has been found useful for treating diabetes insipidus and for increasing the thirst response. Chlorpropamide does not produce adequate control, the appropriate therapy is a fixed dose of desmopressin and a prescribed quantity of water. These patients usually require encouragement to drink. A constant antidiuresis is maintained by a rigid regimen of desmopressin and water intake prescribed for every 6 to 8 hours during a 24-hour period. Regular follow-up with measurement of serum sodium concentrations is essential to prevent development of water intoxication with hyponatremia or recurrent dehydroinfusion with hyponatremia.

Diabetes Insipidus in Pregnancy

Desmopressin is the only therapeutic agent recommended for treatment of diabetes insipidus during pregnancy. Desmopressin has 2% to 25% of the oxytocic activity of lysine vasopressin or arginine vasopressin and can be used with minimal stimulation of the oxytocin receptors in the uterus. This is the only antidiuretic agent recommended for fluid management during pregnancy. The agent can also produce significant antidiuresis even in patients with severe hypothalamic diabetes insipidus.

During delivery, these patients should maintain adequate oral intake and continued administration of desmopressin. Physicians should be cautious about overadministration of fluid parenterally during delivery because these patients are not able to excrete the fluid and may experience water intoxication and hyponatremia.

After delivery, oxytocinase decreases in plasma and, depending on the cause of the diabetes insipidus, the disorder may disappear or the patient may become asymptomatic with regard to fluid intake and urine volume.

Diabetes Insipidus after Hypothalamic or Pituitary Surgery and after Head Injury

Pituitary Surgery

The surgeon often knows how severely the posterior pituitary or stalk has been injured. The difficulty in making the diagnosis in this clinical setting has been discussed under “Differential Diagnosis.” Sometimes the duration of diabetes insipidus is transient, and the surgeon may prefer to treat only with fluid replacement parenterally or orally (if the patient is awake and able to respond to thirst).

To treat diabetes insipidus, desmopressin may be given at 0.5 to 2 μg subcutaneously, intramuscularly, or intravenously. The intravenous route may be preferable because (1) there is no question about absorption and (2) with the lack of pressor activity, desmopressin is safe. Urine output is reduced in 1 to 2 hours, and the duration of effect is 6 to 24 hours. If the patient is alert, thirst is a good guide to fluid replacement. Because diabetes insipidus may be transient and some patients may experience the triphasic pattern described previously, it is desirable to allow polyuria to return before administering subsequent doses of desmopressin.

Head Injury

The treatment of acute diabetes insipidus after blunt trauma to the head, usually from a motor vehicle accident, is similar to that in the postoperative situation, except that the patient with head injury is more likely to be comatose and unable to respond to thirst. Therefore, hypotonic polyuria is more likely to be associated with hyponatremia. Because a comatose patient must be given fluids parenterally, some clinicians prefer to use a continuous infusion of low-dose vasopressin. The vasopressin can either be added directly to the crystalloid solution that is being administered or infused separately to maintain a constant antidiuresis while fluid intake is adjusted appropriately to any persistent polyuria and to cover insensible water loss. Doses of 0.25 to 2.7 mL/kg per hour have been described. With this method, there is a potential to produce hyponatremia, and serum sodium levels must be checked regularly. Furthermore, with continuous replacement, one does not know whether normal function has resumed or whether a patient is entering the second phase of a triphasic pattern.

Nephrogenic Diabetes Insipidus

Adequate water intake should always be maintained; indeed, appropriate water intake may be lifesaving in congenital nephrogenic diabetes insipidus. By definition, these forms of diabetes insipidus do not respond to vasopressin or desmopressin, although there may be rare partial defects with some response to high doses of desmopressin.

Therapy is aimed at reducing symptomatic polyuria by reducing the volume of urine output. The volume is reduced primarily by causing an element of volume contraction. For congenital nephrogenic diabetes insipidus, volume contraction has been produced by a low-sodium diet and a thiazide diuretic. Thiazide diuretics act at the distal convoluted tubule to inhibit sodium chloride absorption, producing sodium (and water) diuresis. The antidiuretic effect has been attributed to ECF volume contraction. For congenital nephrogenic diabetes insipidus, volume contraction has been produced by a low-sodium diet and a thiazide diuretic. Thiazide diuretics act at the distal convoluted tubule to inhibit sodium chloride absorption, producing sodium (and water) diuresis. The antidiuretic effect has been attributed to ECF volume contraction. For congenital nephrogenic diabetes insipidus, volume contraction has been produced by a low-sodium diet and a thiazide diuretic. Thiazide diuretics act at the distal convoluted tubule to inhibit sodium chloride absorption, producing sodium (and water) diuresis. The antidiuretic effect has been attributed to ECF volume contraction.

Addition of amiloride therapy is now recommended because of the absence of duodenal ulcer and gastrointestinal hemorrhage that may be produced by nonsteroidal anti-inflammatory agents. If the condition is recognized early and treated vigorously,
complications such as intracerebral calcification, mental retardation, growth failure, and hydronephrosis are largely preventable. 

Drug-induced nephrogenic diabetes insipidus should be treated by stopping the offending agent if possible. Persistence of nephrogenic diabetes insipidus can be similarly treated by hydrochlorothiazide and amiloride. With the induced volume contraction, the patient should be closely monitored for the development of renal or other toxicity of the drug causing the diabetes insipidus. 

For example, volume contraction produced by thiazide diuretics, when used to treat lithium-induced nephrogenic diabetes insipidus, may decrease lithium excretion and predispose to lithium toxicity. 

Amiloride has the advantage of decreasing lithium entrance into cells in the distal tubule and may have a specific and preferable action for the treatment of lithium-induced nephrogenic diabetes insipidus. 

Indapamide is an antihypertensive diuretic agent with a structure similar to that of hydrochlorothiazides and chlorpropamide. It has been reported to reduce urine volume in patients with hypothalamic diabetes insipidus.

Diabetes Insipidus in Association with Other Therapeutic Decisions

Routine Surgical Procedures

In most routine surgical procedures, the patient is not unconscious long enough to require anything more than administration of the usual dose of desmopressin and careful monitoring of fluids during the surgery to avoid overhydration. If the patient has been taking desmopressin orally and is to be given nothing by mouth, a nasal or a parenteral dose can be administered before the procedure.

Panhypopituitarism

Because hypothyroidism and adrenal insufficiency act directly on the kidney to inhibit the ability to excrete water, any patient who has anterior pituitary deficiency in association with diabetes insipidus is at risk for hypernatremia if treatment for diabetes insipidus is continued but treatment with thyroid hormone and (more dramatically) hydrocortisone is stopped. For such patients, it is important to maintain treatment of all anterior and posterior pituitary deficiencies continuously as the balance of these replacements is essential.

Promoting a Saline Diuresis

In some clinical situations, such as chemotherapy or use of some contrast agents, diuresis is desirable to minimize renal toxicity. Continuing desmopressin while giving a large volume of normal saline induces a prompt natriuresis and hypernatremia. Withholding desmopressin and replacing with 5% dextrose in water may lead to hyperglycemia, and replacing with normal saline may lead to hyponatremia. It has been reported that continuous intravenous administration of low-dose vasopressin in a manner similar to that described earlier for comatose patients can be used. In this case, the dose of vasopressin is even lower (e.g., 0.08 to 0.1 mU/kg per hour) to allow a moderate and controlled diuresis. As with any situation in which vasopressin is given continuously, serum sodium levels must be checked regularly and the amount of fluids infused monitored carefully.

Hypertonic Encephalopathy

Conditions other than diabetes insipidus are the more common causes of hypernatremia with coma and hypertonic encephalopathy. These conditions often affect older patients with concurrent renal problems or who are receiving treatment with diuretics, in whom total body sodium may be decreased despite hypernatremia, or children with insensible loss or diarrhea.

and who are taking sodium-containing fluids. In patients with severe hypernatremia and hypertonic encephalopathy, overaggressive treatment of the hypernatremia may cause cerebral edema and may worsen the neurologic condition. Sodium is mainly an extracellular cation, and hypernatremia invariably leads to movement of water out of cells and to cellular dehydration.

In the brain, idiogenic osmoles are generated intracellularly, and the degree of cell shrinkage is less than would be expected on the basis of the degree of hypernatremia. The idiogenic osmoles are in three organic classes: (1) polyols, (2) trimethylamines, and (3) amino acids and their derivatives. The increase in these organic osmoles has occurred within 1 to 2 hours in experimental animals but may be somewhat slower in humans. The important clinical observation is that when fluid is replaced, these organic osmoles decrease much more slowly intracellularly than the decrease in osmolality of the ECF. This asynchrony increases the potential for cerebral edema and worsening of the neurologic condition with overzealous treatment of hypernatremia.

In most cases of diabetes insipidus that are seen immediately after surgery or that are identified promptly after head injury, the diagnosis is made within a few hours and therapy is instituted promptly. When the duration of the hypernatremia is not known, the degree of correction of hypernatremia should not exceed 0.5 mEq/L per hour to prevent cerebral edema and convulsions.

Organ Donors

As noted earlier, diabetes insipidus is commonly associated with brain death. Because the patient may be a candidate for organ donation, regulating fluid homeostasis is thought to be desirable for maintaining the health of the organs. Although this is controversial, some degree of treatment of diabetes insipidus is not unreasonable. This may be a situation in which continuous administration of vasopressin in a low dose is easier than maintaining antidiuresis with intermittent doses of desmopressin.
Severe hypo-osmolality (serum Na$^+$) is almost always a consequence of pathological conditions. Even relatively mild hypo-osmolality can quickly progress to more dangerous levels during the therapeutic management of other disorders. Overly rapid correction of hyponatremia can itself cause severe neurologic morbidity and mortality. If the calculated effective plasma osmolality is less than 275 mOsm/kg H$_2$O, the patient usually has true osmotic diuresis and underlying diseases that could reverse the hypo-osmolality. However, in most cases, the calculated effective plasma osmolality is normally based on the calculated serum sodium concentration and glucose elevation; traditionally, the correction factor has been 1.6 mEq/L for each 100 mg/dL increase in serum glucose concentration above normal levels, but later studies showed a more complex relation between serum glucose and serum Na$^+$. The calculated effective plasma osmolality is less than 275 mOsm/kg H$_2$O or if the corrected serum Na$^+$ is less than 135 mEq/L, significant hypo-osmolality exists, provided that large concentrations of unmeasured solutes or pseudohyponatremia secondary to hyperglycemia or hyperproteinemia is not present. To eliminate the latter possibilities, plasma osmolality should also be measured directly when the hyponatremia cannot be accounted for by elevated serum glucose levels. The absence of a discrepancy between the calculated and measured total plasma osmolalities (<10 mOsm/kg H$_2$O) confirms the absence of

**THE SYNDROME OF INAPPROPRIATE ANTIDIURETIC HORMONE SECRETION**

SIADH is produced when plasma levels of arginine vasopressin are elevated at times during which the physiologic secretion of vasopressin from the posterior pituitary would normally be suppressed. Because the clinical abnormality is in the osmotic pressure of body fluids, the hallmark of SIADH is hypo-osmolality. In 1957, this finding led to the identification of the first well-described cases of this disorder; it is thus necessary to summarize general issues concerning hypoosmolality and hyponatremia before we present details specific to SIADH.

**Hypo-osmolality and Hyponatremia**

**Incidence**

Hypo-osmolality is one of the most common disorders of fluid and electrolyte balance in hospitalized patients. The incidence and prevalence of hypo-osmolar disorders depend on the nature of the population of patients studied as well as the laboratory methods and criteria used to diagnose hyponatremia. Most investigators have used serum sodium as a screening test to determine the clinical incidence of hypo-osmolality. When hyponatremia was defined as a serum Na$^+$ less than 135 mEq/L, incidences as high as 15% to 30% were observed in studies of both acutely and chronically hospitalized patients. However, incidences decreased to the range of 1% to 4% when only patients with serum Na$^+$ under 130 to 131 mEq/L were included, which represents a more appropriate level at which to define the occurrence of clinically significant cases of this disorder. Even when these more stringent criteria have been used, incidences from 7% to 53% have been reported in institutionalized geriatric patients.

All studies to date have noted a high proportion of iatrogenic or hospital-acquired hyponatremia, which has accounted for as many as 40% to 75% of all patients studied. Therefore, although hyponatremia and hypo-osmolality are quite common, most cases are relatively mild and most are acquired during the course of hospitalization. Nonetheless, hyponatremia is important clinically because:

1. Severe hypo-osmolality (serum Na$^+$ < 120 mEq/L) is associated with substantial morbidity and mortality.
2. Even relatively mild hypo-osmolality can quickly progress to more dangerous levels during the therapeutic management of other disorders.
3. Overly rapid correction of hyponatremia can itself cause severe neurologic morbidity and mortality.
4. Mortality rates are much higher (threelfold to 80-fold higher) in patients with even asymptomatic degrees of hypoosmolality compared with normonatremic patients.

In many but not all cases, hypo-osmolality is more an indicator of the severity of underlying illnesses than an independent factor contributing to mortality.

**Osmolarity, Toxicity, and Serum Sodium**

As discussed previously, the osmolarity of body fluid is normally maintained within narrow limits by osmotically regulated vasopressin secretion and thirst. Although basal plasma osmolarity can vary appreciably among individuals, the range in the general population under conditions of normal hydration is between 280 and 295 mOsm/kg H$_2$O. Plasma osmolarity can be determined directly by measuring the freezing-point depression or the vapor pressure of plasma. Alternatively, it can be calculated indirectly from the concentrations of the three major solutes in plasma:

$$\text{plasma osmolality} = 2 \times \text{Na}^+ \text{ (mEq/L)} + \text{glucose (mg/dL)} / 18 + \text{BUN (mg/dL)} / 2.8$$

Both methods produce comparable results under most conditions. Although either method produces valid measures of total osmolality, this is not always equivalent to the effective osmolality, commonly referred to as the tonicity of the plasma. Only cell solutes such as Na$^+$ and Cl$^-$ that cannot permeate the cell membrane and remain relatively compartmentalized within the ECF space are effective solutes because they create osmotic gradients across cell membranes and regulate the osmotic movement of water between the intracellular fluid (ICF) compartment and the ECF compartment. Solutes that readily permeate cell membranes (e.g., urea, ethanol, methanol) are not effective solutes. Therefore, only the concentrations of effective solutes in plasma should be used to ascertain whether clinically significant hyperosmolality or hypo-osmolality is present.

Because sodium and its accompanying anions are the major effective plasma solutes, hyponatremia and hypo-osmolality are usually synonymous. However, there are two situations in which hyponatremia does not reflect true hypo-osmolality:

1. **Pseudohyponatremia**, produced by marked elevations of either lipids or proteins in plasma. When Na$^+$ is measured by flame photometry, the concentration of Na$^+$ per liter of plasma is artificially decreased because of the larger relative proportion of plasma volume that is occupied by the excess lipids or proteins. However, because the increased protein or lipid does not appreciably change the total number of solute particles in solution, the directly measured plasma osmolality is not significantly affected. Measurement of serum Na$^+$ by ion-specific electrodes, now commonly performed by most clinical laboratories, is less influenced by high concentrations of lipids or proteins than is measurement of serum Na$^+$ by flame photometry.

2. **Presence of high concentrations of effective solutes other than Na$^+$ in plasma**. The initial hyperosmolality produced by the additional solute causes an osmotic shift of water from ICF to ECF, which in turn produces a dilutional decrease in serum Na$^+$. When equilibrium between both fluid compartments is achieved, the total effective osmolality remains relatively unchanged. This situation most commonly occurs with hyperglycemia and is a frequent cause of hyponatremia in hospitalized patients, accounting for up to 10% to 20% of all cases.

Misdagnosis of true hypo-osmolality in such cases can be avoided by measuring plasma osmolality directly or by correcting the measured serum Na$^+$ for the glucose elevation; traditionally, the correction factor has been 1.6 mEq/L for each 100 mg/dL increase in serum glucose concentration above normal levels, but later studies showed a more complex relation between hyperglycemia and serum Na$^+$. The initial hyperosmolality produced by the additional solute causes an osmotic shift of water from ICF to ECF, which in turn produces a dilutional decrease in serum Na$^+$. When equilibrium between both fluid compartments is achieved, the total effective osmolality remains relatively unchanged. This situation most commonly occurs with hyperglycemia and is a frequent cause of hyponatremia in hospitalized patients, accounting for up to 10% to 20% of all cases. Therefore, only the concentrations of effective solutes in plasma should be used to ascertain whether clinically significant hyperosmolality or hypo-osmolality is present.

**Because of these potential confounders, determining whether true hypo-osmolality is present can sometimes be difficult. A straightforward and relatively simple approach is as follows:**

1. The effective plasma osmolality should be calculated from the measured serum Na$^+$ and glucose concentration (2 × [Na$^+$] + glucose/18); alternatively, the measured serum Na$^+$ can simply be corrected by 1.6 to 2.4 mEq/L for each 100 mg/dL increase in serum glucose concentration above normal levels (100 mg/dL).
2. If the calculated effective plasma osmolality is less than 275 mOsm/kg H$_2$O or if the corrected serum Na$^+$ is less than 135 mEq/L, significant hypo-osmolality exists, provided that large concentrations of unmeasured solutes or pseudohyponatremia secondary to hyperglycemia or hyperproteinemia is not present.
3. To eliminate the latter possibilities, plasma osmolality should also be measured directly when the hyponatremia cannot be accounted for by elevated serum glucose levels. The absence of a discrepancy between the calculated and measured total plasma osmalities (<10 mOsm/kg H$_2$O) confirms the absence of
significant amounts of unmeasured solutes; if a significant discrepancy between these measures is found (an osmolar gap), appropriate tests must be conducted to rule out pseudohyponatremia or to identify possible unmeasured plasma solutes.

**Pathogenesis of hypo-osmolality**

Because water moves freely between ICF and ECF, osmolality is always equivalent in both of these fluid compartments. Because the bulk of body solute consists of electrolytes, namely the exchangeable Na⁺ (Na⁺ₑ) in ECF and the exchangeable K⁺ (K⁺ₑ) in ICF along with their associated anions, total body osmolality is largely a function of these two components:

\[
\text{Total body osmolality} = (2 \times \text{Na}⁺ + 2 \times \text{K}⁺ + \text{nonelectrolyte solute})/\text{body water}
\]

According to this definition, the presence of plasma hypo-osmolality indicates a relative excess of water to solute in the ECF. The excess can be produced either by an excess of body water, resulting in dilution of remaining body solute, or by a depletion of body solute, either Na⁺ or K⁺, relative to body water. This classification is an oversimplification because most hypo-osmolar states involve significant components of both solute depletion and water retention. Nonetheless, it is conceptually useful for understanding the underpinnings of hypo-osmolality and as a framework for therapy of hypo-osmolar disorders.

**Solute Depletion**

Depletion of body solute can result from any significant losses of ECF. Body fluid losses by themselves rarely cause hypo-osmolality because excreted or secreted body fluids are usually isotonic or hypertonic relative to plasma and therefore tend to increase plasma osmolality. When hypo-osmolality accompanies ECF losses, it is the result of replacement of body fluid losses by more hypotonic solutions either by drinking or by infusion, thereby diluting the remaining body solutes. If the solute losses are marked, these patients show signs of volume depletion (e.g., Addisonian crisis). However, such patients often have a more deceptive clinical presentation because the volume deficits were partially replaced. Moreover, they may not manifest signs or symptoms of cellular dehydration because osmotic gradients draw water into the relatively hypertonic ICF.

Therefore, clinical evidence of hypovolemia strongly supports solute depletion as the cause of plasma hypo-osmolality, but absence of clinically evident hypovolemia never completely eliminates this as a possibility. Although ECF solute losses are responsible for most cases of depletion-induced hypo-osmolality, ICF solute loss can also cause hypo-osmolality as a result of osmotic water shifts from the ICF into the ECF. This mechanism probably contributes to some cases of diuretic-induced hypo-osmolality, in which depletion of total body K⁺ often occurs.

**Water Retention**

Despite the importance of solute depletion in some patients, most cases of clinically significant hypo-osmolality are caused by increases in total body water rather than by primary losses of extracellular solute. Such increases can occur because of either impaired renal free water excretion or excessive free water intake. However, the former accounts for most hypo-osmolar disorders because normal kidneys have sufficient diluting capacity to allow excretion of free water up to approximately 18 L/day. Intakes of this magnitude are occasionally seen in some psychiatric patients but not in most patients with SIADH, in whom fluid intakes average only 2 to 3 L/day. Consequently, dilutional hypo-osmolality is usually the result of an abnormality of renal free water excretion. The renal mechanisms responsible for impairments in free water excretion can be sub grouped according to whether the major impairment in free water excretion occurs in proximal or distal parts of the nephron, or both.

Any disorder that leads to a decrease in GFR causes increased reabsorption of both Na⁺ and water in the proximal tubule. As a result, the ability to excrete free water is limited because of decreased delivery of tubular fluid to the distal nephron. Disorders causing solute depletion through nonrenal mechanisms (e.g., gastrointestinal fluid losses) also produce this effect. Disorders that cause a decreased GFR in the absence of significant ECF fluid losses are, for the most part, edema forming states associated with decreased effective arterial blood volume (EABV) and secondary hyperaldosteronism.

Even though these conditions are characterized by increased proximal reabsorption of both Na⁺ and fluid, water retention also results from increased distal reabsorption caused by non-osmotic baroreceptor-mediated stimulated increases in plasma vasopressin levels. Distal nephron impairments in free water excretion are characterized by inability to dilute tubular fluid maximally. These disorders are usually associated with abnormalities in the secretion of vasopressin from the posterior pituitary. However, just as depletion-induced hypo-osmolar disorders usually include an important component of secondary impairments of free water excretion, most dilution-induced hypo-osmolar disorders involve significant degrees of secondary solute depletion. This is described later with SIADH.

Some dilutional disorders do not fit well into either category, specifically the hyponatremia that sometimes occurs in patients who ingest large volumes of beer with little food intake for prolonged periods, called beer potomania. Even though the volume of fluid ingested may not seem sufficiently excessive to overwhelm renal diluting mechanisms, free water excretion is limited by low urine solute excretion, thereby causing water retention and dilutional hyponatremia. A case in which hyponatremia occurred in an ovulotovaguetarian with a very low protein intake but no beer ingestion is consistent with this pathophysiology.

**Adaptation to Hyponatremia: Intracellular Fluid and Extracellular Fluid Volume Regulation**

Many past studies have indicated that the combined effects of water retention and urinary solute excretion cannot adequately explain the degree of plasma hypo-osmolality observed in patients. This observation led to the theory of cellular inactivation of solute. The theory suggested that as ECF osmolality falls, water moves into cells along osmotic gradients, causing the cells to swell. At some point during this volume expansion, the cells osmotically inactivate some of their intracellular solutes as a defense mechanism to prevent continued cellular swelling and detrimental effects on cell function and survival. This inactivation decreases the intracellular osmolality and water shifts back out of the ICF into the ECF, further worsening the dilution-induced hypo-osmolality. Despite the appeal of this theory, its validity has never been demonstrated conclusively in either human or animal studies.

An appealing alternative theory is that cell volume is maintained under hypo-osmolar conditions by extrusion of potassium. Whole-brain volume regulation through electrolyte losses was first described by Yannet and has long been recognized as the mechanism by which the brain is able to adapt to hyponatremia and limit brain edema to sublethal levels. After the recognition that low-molecular-weight organic compounds (called organic osmolytes) are a significant osmotic component of a wide variety of cell types, studies demonstrated accumulation of these compounds in response to hyperosmolality in both kidney and brain tissue (see hyperosmotic encephalopathy earlier). Conversely, the brain loses organic osmolytes in addition to electrolytes during volume regulation to hypo-osmolar conditions in experimental animals and human patients. These losses occur relatively quickly (within 24 to 48 hours in rats) and can account for as much as one third of the brain solute losses during hyponatremia. Such coordinate losses of both electrolytes (K⁺) and organic osmolytes from brain tissue allow effective regulation of brain volume during chronic hyponatremia. Consequently, it is now clear that cellular volume regulation in brain tissue occurs predominantly through depletion of a variety of intracellular solutes rather than intracellular osmotic inactivation.

Although contemporary studies of volume regulation during hyponatremia have focused on the brain, all cells regulate volume by cellular losses of both electrolytes and organic solutes to varying degrees. However, volume regulatory processes are not limited to cells. In most cases of hyponatremia induced by stimulated antidiuresis and water retention, natriuresis also regulates the volumes of the ECF and intravascular spaces.

Many experimental and clinical observations are consistent with ECF volume regulation through secondary solute losses, as described next.

First, the concentrations of most blood constituents other than Na⁺ and Cl⁻ are not decreased in patients with SIADH; this suggests that plasma volume is not nearly as expanded as would be predicted simply by the measured decreases in serum [Na⁺].

Second, an increased incidence of hypertension has never been observed in patients with SIADH; again, this is evidence against significant expansion of the arterial bed.
blood volume.

Third, results of animal studies in both dogs and rats have indicated that a significant component of chronic hyponatremia is attributable to secondary Na⁺ losses rather than water retention. The relative contributions of water retention and sodium loss vary with the duration and severity of the hyponatremia: water retention was found to be the major cause of decreased serum [Na⁺] in the first 24 hours of induced hyponatremia in rats, but Na⁺ depletion became the predominant etiologic factor after longer periods (7 to 14 days) of sustained hyponatremia, particularly at low (<115 mEq/L) serum Na⁺ levels.\(^\text{424}\)

Finally, multiple studies of body fluid compartment volumes in hyponatremic patients have not shown either plasma or ECF expansion. For example, a report of body fluid space measurements using isotope dilution techniques in hyponatremic and normonatremic patients with small cell lung carcinoma showed no differences between the two groups with regard to exchangeable sodium space, ECF volume determined by\(^\text{35}\) SO₄ distribution, or total body water.\(^\text{425}\)

Such results have generally been explained by the relative insensitivity of isotope dilution techniques for measurement of body fluid compartment spaces, but an equally plausible possibility is that body fluid compartments have regulated back toward normal through a combination of extracellular (predominantly electrolyte) and intracellular (electrolyte and organic osmolyte) solute losses.\(^\text{442}\) Figure 9-8 (Figure Not Available) schematically illustrates some of the volume regulatory processes that probably occur in response to water retention induced by inappropriate anti-diuresis. The degree to which solute losses versus water retention contribute to the resulting hyponatremia depends on many different factors, including:

1. The cause of the hyponatremia.

2. The rapidity of hyponatremia development.
3. The chronicity of the hyponatremia.
4. The volume of daily water loading and subsequent volume expansion.
5. Undoubtedly, some degree of individual variability as well.

Figure 9-8 (Figure Not Available) Schematic illustration of potential changes in whole-body fluid compartment volumes at various times during adaptation to hyponatremia. A, Under basal conditions the concentrations of effective solutes in the extracellular fluid ([S]\text{ECF}) and the intracellular fluid ([S]\text{ICF}) are in osmotic balance. B, During the first phase of water retention resulting from inappropriate antidiuresis, the excess water distributes across total body water, causing expansion of both ECF and ICF volumes (dashed lines) with equivalent dilutional decreases in [S]\text{ECF} and [S]\text{ICF}. C, In response to the volume expansion, compensatory volume regulatory decreases (VRD) occur to reduce the effective solute content of both the ECF (through pressure diuresis and natriuretic factors) and ICF (through increased electrolyte and osmolyte extraction mediated by stretch-activated channels and down-regulation of synthesis of osmolytes and osmolyte uptake transporters). D and E, If both processes go to completion, such as under conditions of fluid restriction, a final steady state can be reached in which ICF and ECF volumes have returned to normal levels but [S]\text{ECF} and [S]\text{ICF} remain low (B). In most cases, this final steady state is not reached and moderate degrees of ECF and ICF expansion persist but are significantly less than would be predicted from the decrease in body osmolality (D). Consequently, the degree to which hyponatremia is due to dilution from water retention versus solute depletion from volume regulatory processes can vary markedly depending on which phase of adaptation the patient is in and also on the relative rates at which the different compensatory processes occur (e.g., delayed ICF VRD can worsen hyponatremia related to shifts of intracellular water into the ECF as intracellular organic osmolytes are extruded and subsequently metabolized, probably accounting for some component of the hyponatremia unexplained by the combination of water retention and sodium excretion in previous clinical studies). (From Verbalis JG. Hyponatremia: epidemiology, pathophysiology, and therapy. Curr Opin Nephrol Hypertens 1993; 2:63652.)

Differential Diagnosis of Hyponatremia and Hypo-osmolality

Because of the multiplicity of disorders causing hypo-osmolality and the fact that many involve more than one pathologic mechanism, a definitive diagnosis is not always possible at the time of initial presentation. Nonetheless, an approach based on clinical parameters of ECF volume status and urinary sodium concentration generally allows sufficient categorization for appropriate decisions regarding initial therapy and further evaluation (Table 9-2).

Table: Differential Diagnosis of Hyponatremia

<table>
<thead>
<tr>
<th>Extracellular Fluid Volume</th>
<th>Urinary [Na⁺]</th>
<th>Presumptive Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Depletion (nonrenal): Gl, cutaneous, or blood ECF loss</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>Depletion (renal): diuretics, mineralocorticoid insufficiency (Addison’s disease), salt-losing nephropathy</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>Depletion (nonrenal): any cause + hypotonic fluid replacement</td>
<td></td>
</tr>
<tr>
<td>Dilation (proximal)</td>
<td>hyponatremia, early decreased effective arterial blood volume</td>
<td></td>
</tr>
<tr>
<td>Dilation (distal): SIADH + fluid restriction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>Dilation (distal): SIADH, glucocorticoid insufficiency</td>
<td></td>
</tr>
<tr>
<td>Depletion (renal): any cause + hypotonic fluid replacement (especially diuretic treatment)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dilation (proximal): decreased effective arterial blood volume (CHF, cirrhosis, nephrosis)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>Dilation (proximal): any cause + diuretics or improvement in underlying disease, renal failure</td>
<td></td>
</tr>
</tbody>
</table>

CHF, congestive heart failure; ECF, extracellular fluid; Gl, gastrointestinal; SIADH, syndrome of inappropriate antidiuretic hormone secretion.


*Urinary [Na⁺] values <30 mEq/L are generally considered to be low and values 30 mEq/L to be high, based on studies of responses of hyponatremic patients to infusions of isotonic saline.

Decompressed Extracellular Fluid Volume

Clinically detectable hypovolemia always signifies total body solute depletion. A low urinary [Na⁺] indicates a nonrenal cause and an appropriate renal response. A high urinary [Na⁺] indicates that renal causes of solute depletion are more likely. Therapy with thiazide diuretics is the most common cause of renal solute losses, particularly in elderly people, but mineralocorticoid deficiency as a result of adrenal insufficiency or mineralocorticoid resistance must be considered as well as (less commonly) renal solute losses related to salt-wasting nephropathy (e.g., polycyclic kidney disease), interstitial nephritis, or chemotherapy.

Reexpanded Extracellular Fluid Volume

Clinically detectable hypervolemia always signifies total body Na⁺ excess. In these patients, hypo-osmolality results from an even greater expansion of total body water caused by a marked reduction in the rate of water excretion (and sometimes an increased rate of water ingestion). The impairment in water excretion is secondary to a decreased effective arterial blood volume, which increases the reabsorption of glomerular filtrate not only in the proximal nephron but also in the distal and collecting tubules by stimulated secretion of vasopressin. These patients generally have a low urinary [Na⁺] because of secondary hyperaldosteronism. Under certain conditions, however, urinary [Na⁺] may be elevated if there is concurrent diuretic therapy or a solute diuresis (e.g., glucosuria in diabetic patients) or after successful treatment of the underlying disease (e.g., isotropic therapy in patients with congestive heart failure).
An additional disorder that can produce hypo-osmolality and hypervolemia is acute or chronic renal failure with fluid overload (although in early stages of renal failure polyuria resulting from vasopressin resistance is more likely). Urinary [Na⁺] in these cases is usually elevated, but it can be variable, depending on the stage of renal failure. It is important to remember that primary polydipsia is not accompanied by signs of hypervolemia because water ingestion alone, in the absence of Na⁺ retention, does not produce clinically apparent degrees of ECF volume expansion.

**Normal Extracellular Fluid Volume**

Many different hypo-osmolar disorders occur with euvolementia, and measurement of urinary [Na⁺] is an especially important first step. A high urinary [Na⁺] usually implies a distally mediated, dilution-induced hypo-osmolality such as that in SIADH. However, glucocorticoid deficiency can mimic SIADH so closely that these two disorders are often indistinguishable in terms of water balance. Hyponatremia resulting from diuretic use can also occur without clinically evident hypovolemia, and urinary [Na⁺] is usually elevated. A low urinary [Na⁺] suggests a depletion-induced hypo-osmolality resulting from ECF losses with subsequent volume replacement by water or other hypotonic fluids. The solute loss is often nonrenal, but an important exception is recent cessation of diuretic therapy because urinary [Na⁺] can decrease to low values within 12 to 24 hours after discontinuation of the drug.

The presence of low serum [K⁺] is an important clue to diuretic use. Low urinary [Na⁺] can also be seen in some cases of hypothyroidism, in the early stages of decreased effective arterial blood volume before the development of clinically apparent salt retention and fluid overload, or during the recovery phase of SIADH. Hence, a low urinary [Na⁺] is less meaningful diagnostically than a high value.
Clinical Syndrome of Inappropriate Antidiuretic Hormone Secretion

SIADH is the most common cause of euvoletic hypo-osmolality; it is also the single most common cause of hypo-osmolality of all etiologic mechanisms encountered in clinical practice, with prevalence rates from 20% to 40% among all hypo-osmolar patients. The clinical criteria necessary to diagnose SIADH remain basically as set forth by Bartter and Schwartz in 1967:

1. Decreased effective osmolality of the ECF (plasma osmolality < 275 mOsm/kg H₂O). Pseudohipernatraemia or hypoglycemia alone must be excluded.
2. Inappropriate urinary concentration (urine osmolality > 100 mOsm/kg H₂O with normal renal function) at some level of hypo-osmolality. This does not mean that urine osmolality is greater than plasma osmolality; rather, the urine is less than maximally dilute (i.e., urine osmolality > 100 mOsm/kg H₂O). Also, urine osmolality need not be elevated inappropriately at all levels of plasma osmolality because in the reset osmostat variant form of SIADH, vasopressin secretion can be suppressed with resultant maximal urinary dilution if plasma osmolality is decreased to sufficiently low levels.
3. Clinical euvoletic as defined by the absence of signs of hypovolemia (orthostasis, tachycardia, decreased skin turgor, dry mucous membranes) or hypervolemia (subcutaneous edema, ascites). Hypovolemia or hypervolemia strongly suggests different causes of hypo-osmolality. Patients with SIADH can become hypovolemic or hypervolemic for other reasons, but in such cases it is impossible to diagnose the underlying inappropriate antidiuresis until the patient is rendered euvoletic and is found to have persistent hypo-osmolality.
4. Elevated urinary sodium excretion with a normal salt and water intake. This criterion is included because of its utility in differentiating between hypo-osmolality caused by a decreased effective arterial blood volume, in which case renal Na⁺ conservation occurs, and distal dilution-induced disorders, in which urinary Na⁺ excretion is normal or increased secondary to ECF volume expansion. Patients with SIADH can have low urinary Na⁺ excretion if they subsequently become hypovolemic or solute-depleted, conditions that sometimes follow severe salt and water restriction. Consequently, high urinary Na⁺ excretion is the rule in most patients with SIADH, its presence does not guarantee this diagnosis, and its absence does not rule out the diagnosis.
5. Absence of other potential causes of euvoletic hypo-osmolality: hypothyroidism, hypercorticism (Addison's disease or pituitary ACTH insufficiency), and diuretic use.

Several other criteria support, but are not essential for, a diagnosis of SIADH. Because volume expansion and vasopressin acting on V₁ receptors in the kidney increase the clearance of uric acid, hypouricemia is found with SIADH. In hyponatraemic patients, values of uric acid have been less than 4 mg/dl (<0.24 mmol/L). A water-loading test is of value when there is uncertainty about the etiology of modest degrees of hypo-osmolality in euvoletic patients, but it does not add useful information if the plasma osmolality is already less than 275 mOsm/kg H₂O.

Inability to excrete a standard water load normally (with normal excretion defined as a cumulative urinary output of at least 90% of the administered water load within 4 hours to suppress urine osmolality to <100 mOsm/kg H₂O) confirms the presence of an underlying defect in free water excretion. However, water excretion is abnormal in almost all disorders that cause hypo-osmolality, whether dilutional or depletion-induced with secondary impairments in free water excretion. Two exceptions are primary hydropenia, in which hypo-osmolality can rarely be secondary to excessive water intake alone, and the reset osmostat variant of SIADH, in which normal excretion of a water load can occur when plasma osmolality falls below the new set-point for vasopressin secretion.

The water load test may also be used to assess water excretion after treatment of an underlying disorder thought to be causing SIADH. For example, after discontinuation of a drug associated with SIADH, a normal water load test can confirm the absence of persistent inappropriate antidiuresis. Despite its limitations as a diagnostic clinical test, water loading remains an extremely useful tool in clinical research for quantitating changes in free water excretion in response to physiologic or pharmacologic manipulations.

A second supportive criterion is an inappropriately elevated plasma vasopressin level in relation to plasma osmolality. With the development of sensitive vasopressin radioimmunoassays capable of detecting the small physiologic concentrations of this peptide that circulate in plasma, it had been hoped that measurement of plasma vasopressin levels might become the definitive test for diagnosis of SIADH. This has not occurred for several reasons:

First, although plasma vasopressin levels are elevated in most patients with this syndrome, the elevations generally remain within the normal physiologic range and are abnormal only in relation to plasma osmolality.

Second, 10% to 20% of patients with SIADH do not have measurably elevated plasma vasopressin levels, and the levels are at the limits of detection by radioimmunoassay.

Third, most disorders causing solute and volume depletion or decreased effective arterial blood volume are associated with elevations of plasma vasopressin levels secondary to nonosmotic hemodynamic stimuli.

Finally, the response to fluid restriction or volume expansion can be helpful in distinguishing between causes of hyponatraemia. Infusion of isotonic sodium chloride (NaCl) in patients with SIADH provokes a natriuresis with little correction of osmolality, whereas fluid restriction allows gradual achievement of solute and water balance through insensible free water losses. By contrast, isotonic saline is the treatment of choice in disorders of solute depletion; when volume deficits are corrected, the stimulus to continued secretion of vasopressin and free water retention is eliminated. The diagnostic value of this therapeutic response is limited somewhat by the fact that patients with proximal types of dilution-induced disorders may show a response similar to that found in patients with SIADH.

Etiology

Although the list of disorders associated with SIADH is long, they can be divided into four major etiologic groups: tumors, CNS disorders, drugs, and pulmonary disorders.

Tumors

The most common association of SIADH is with tumors. Although many different types of tumors have been associated with SIADH, bronchogenic carcinoma of the lung has been uniquely associated with SIADH since the first description of this disorder in 1957. In virtually all cases, the bronchogenic carcinomas causing this syndrome have been of the small cell (or oat cell) variety. Incidences of hyponatraemia as high as 11% of all patients with small cell carcinoma or 33% of those with more extensive disease have been reported.

The unusually high incidence of small cell carcinoma of the lung, together with the relatively favorable therapeutic response of this type of tumor, makes it imperative that all adult patients presenting with an otherwise unexplained SIADH be investigated thoroughly and aggressively for a possible lung tumor. The evaluation should include a chest CT scan or MRI study and bronchoscopy with cytologic analysis of bronchial washings even if the results of routine chest radiography are normal, because several studies have reported hypo-osmolality that predated radiographic abnormality by 3 to 12 months. Head and neck cancers are another group of malignancies associated with relatively higher incidences of SIADH and some of these tumors have been shown to synthesize vasopressin. A report from a large cancer hospital showed an incidence of hyponatraemia for all malignancies combined of 3.7%, with approximately one third of these related to SIADH.
A large number of CNS disorders are associated with SIADH; no common denominator links them. This is not surprising when one considers the neuroanatomy described earlier. Magnocellular

### TABLE 9-3 -- Common Causes of Syndrome of Inappropriate Antidiuretic Hormone Secretion

<table>
<thead>
<tr>
<th><strong>Tumors</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Pulmonary-medial (bronchogenic carcinoma; mesothelioma; thymoma)</td>
<td></td>
</tr>
<tr>
<td>2. Nonchest (duodenal carcinoma; pancreatic carcinoma; ureteral/prostate carcinoma; uterine carcinoma; nasopharyngeal carcinoma; leukemia)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Central Nervous System Disorders</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Mass lesions (tumors; brain abscesses; subdural hematoma)</td>
<td></td>
</tr>
<tr>
<td>2. Inflammatory diseases (encephalitis; meningitis; systemic lupus; acute intermittent porphyria; multiple sclerosis)</td>
<td></td>
</tr>
<tr>
<td>3. Degenerative-demyelinative diseases (Guillain-Barré; spinal cord lesions)</td>
<td></td>
</tr>
<tr>
<td>4. Miscellaneous (subarachnoid hemorrhage; head trauma; acute psychosis; delirium tremens; pituitary stalk section; transsphenoidal adenomectomy; hydrocephalus)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Drug Induced</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Stimulated AVP release (nicotine; phenothiazines; tricyclics)</td>
<td></td>
</tr>
<tr>
<td>2. Direct renal effects and/or potentiation of AVP antidiuretic effects (dDAVP; oxytocin; prostaglandin synthesis inhibitors)</td>
<td></td>
</tr>
<tr>
<td>3. Mixed or uncertain actions (angiotensin-converting enzyme inhibitors; carbamazepine and oxcarbazepine; chlorpropamide; clofibrate; clozapine; cyclophosphamide; 3,4-methylenedioxyamphetamine [Ecstasy]; omeprazole; serotonin reuptake inhibitors; vincristine)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Pulmonary Diseases</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Infections (tuberculosis; acute bacterial and viral pneumonia; aspergillosis; empyema)</td>
<td></td>
</tr>
<tr>
<td>2. Mechanical-ventilatory (acute respiratory failure; COPD; positive-pressure ventilation)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Other</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Acquired immunodeficiency syndrome (AIDS) and AIDS-related complex</td>
<td></td>
</tr>
<tr>
<td>2. Prolonged strenuous exercise (marathon; triathlon; ultramarathon; hot-weather hiking)</td>
<td></td>
</tr>
<tr>
<td>3. Senile atrophy</td>
<td></td>
</tr>
<tr>
<td>4. Idiopathic</td>
<td></td>
</tr>
</tbody>
</table>

- AVP, arginine vasopressin; COPD, chronic obstructive pulmonary disease; dDAVP, 1-deamino(8-arginine) vasopressin.

Vasopressin neurons receive excitatory inputs from osmoreceptive cells in the anterior hypothalamus but also a major innervation from brain stem cardiovascular regulatory and emetic centers. Although various components of these pathways have yet to be elucidated fully, many of them appear to have inhibitory as well as excitatory components. Consequently, any diffuse CNS disorder can potentially cause vasopressin hypersecretion either by nonspecifically exciting these pathways through irritative foci or by disrupting them and thereby decreasing the level of inhibition. The wide variety of CNS processes that can potentially cause SIADH stands in contrast to CNS causes of diabetes insipidus, which are limited to lesions of the suprasellar hypothalamus.

### Drugs

Drug-induced hyponatremia is a common cause of hypo-osmolality. Table 9-3 lists some of the agents that have been associated with SIADH, but new drugs are added continually. Pharmacologic agents may stimulate secretion of vasopressin, activate V2 receptor sites, or potentiate the antidiuretic effect of vasopressin. Not all of the drug effects are fully understood, however, and many appear to work through a combination of mechanisms.

A particularly interesting and clinically important class of agents is the selective serotonin reuptake inhibitors (SSRIs). In studies in rats, serotoninergic agents increased secretion of vasopressin but more directly oxytocin. In humans, SSRIs have generally not had significant effects on secretion of vasopressin. However, hyponatremia after SSRI administration has been reported almost exclusively in elderly persons, with rates as high as 22% to 25%, although in larger series the incidence was closer to 1 in 200. Elderly patients are uniquely hypersensitive to serotonin stimulation of vasopressin secretion. A similar effect is probably also responsible for the severe fatal hyponatremia caused by use of the recreational drug 3,4-methylenedioxymethamphetamine, Ecstasy. which has substantial serotoninergic activity and activated hypothalamic magnocellular neurons in rats.

### Pulmonary Disorders

Various pulmonary disorders have been associated with SIADH; other than tuberculosis, acute pneumonia, and advanced chronic obstructive lung disease, however, the occurrence of hypo-osmolality has been noted only sporadically. One reported case of pulmonary tuberculosis suggested that tuberculous lung tissue might synthesize vasopressin, but reports of the reset osmostat with advanced pulmonary tuberculosis and others indicated nonsomotic stimulation of secretion of vasopressin. Hypoxia stimulated secretion of vasopressin in animals, but in humans hypercarbia was more associated with abnormal water retention.

Elevated vasopressin levels may be limited to the initial days of hospitalization, when respiratory failure is most marked. Therefore, with SIADH in nonmalignant pulmonary disease, pulmonary disease is obvious with severe dyspnea or extensive radiographically evident infiltrates and the inappropriate antidiuresis is usually limited to the period of respiratory failure. Mechanical ventilation can cause inappropriate secretion of vasopressin and can worsen SIADH caused by other factors. The mechanism is thought to be decreased venous return.

### Other Causes

In patients with acquired immunodeficiency syndrome (AIDS) or AIDS-related complex and human immunodeficiency virus (HIV) infection, the incidence of hyponatremia has been reported to be as high as 30% to 38% in adults and children. Although there are many potential etiologic mechanisms, including dehydration, adrenal insufficiency, and pneumonitis, 12% to 68% of AIDS patients with hyponatremia appear to meet criteria for a diagnosis of SIADH. Not unexpectedly, some of the medications used to treat these patients may cause the hyponatremia through direct renal tubular toxicity or induced SIADH.

Elderly patients sometimes experience SIADH without any apparent underlying etiologic factor, and the incidence of hyponatremia in geriatric patients suggests that the normal aging process may be accompanied by abnormalities of regulation of secretion of vasopressin. Such an effect may potentially account for the fact that drug-induced hyponatremia occurs much more frequently in older patients. In a series of 50 consecutive elderly patients meeting criteria for SIADH, 60% remained idiopathic despite rigorous evaluation, leading the authors to conclude that extensive diagnostic procedures were not warranted in such elderly patients if routine history, physical examination, and laboratory evaluation failed to suggest an underlying cause.

Some well-known stimuli for vasopressin secretion are notable primarily because of their exclusion from Table 9-3. Despite unequivocal stimulation of vasopressin secretion by nicotine, cigarette smoking has been associated with SIADH rarely and primarily in psychiatric patients who have several other potential causes of inappropriate vasopressin secretion.
Random hypersecretion of vasopressin. Low or even undetectable plasma vasopressin levels despite classic clinical characteristics of SIADH. A reset osmostat system whereby vasopressin is secreted at an abnormally low threshold of plasma osmolality but otherwise displays a normal response to normal or low but nonsuppressible levels of vasopressin can cause sufficient impairment of free water excretion to produce hypo-osmolality, depending on the volume expansion produced as a result of water retention was unequivocally shown by Leaf and co-workers since the original cases studied by Bartter and Schwartz.

Contribution of Natriuresis to the Hyponatremia of SIADH

It is surprising that no correlation has been found between any of these patterns of secretion of vasopressin and the various causes of SIADH.

The best physiologic example of a reset osmostat is pregnancy, as discussed earlier. Perhaps the most perplexing aspect of the reset osmostat pattern is its occurrence in patients with tumors, which suggests that in some of these cases a tumor-related mechanism may affect pituitary vasopressin secretion. The pattern of SIADH that occurs without measurable vasopressin secretion is not yet well understood, but the positive response of one such patient to a vasopressin V2 receptor antagonist suggests that it may represent increased renal sensitivity to low circulating levels of vasopressin.

It is surprising that no correlation has been found between any of these patterns of secretion of vasopressin and the various causes of SIADH. It seems likely that in many cases a heterogeneous group of CNS processes are involved, including osmotic and nonosmotic and stimulatory and inhibitory pathways, rather than a single dominant cause.

**Pathophysiology**

**Sources of Vasopressin Secretion**

Elevated plasma levels of vasopressin can be broadly divided into those associated with paraneoplastic (ectopic) secretion of vasopressin and those associated with pituitary hypersecretion of vasopressin. There is substantial cumulative evidence that tumor tissue can synthesize vasopressin, but it is not certain whether all tumors associated with SIADH do so, because only about half of small cell carcinomas have been found to contain vasopressin immunoreactivity and many of the tumors listed in Table 9-3 have not been so studied.

**Pituitary Vasopressin Secretion: Inappropriate versus Appropriate**

In most cases of SIADH, vasopressin secretion originates from the posterior pituitary. This is also true of more than 90% of all cases of hyponatremia, including hypovolemic and hypervolemic hyponatremia. This raises the question: What is inappropriate secretion of vasopressin?

Secretion of vasopressin in response to a hypovolemic stimulus is clearly physiologically appropriate, but when it leads to symptomatic hyponatremia it can be considered inappropriate for the osmolality. Despite these semantic difficulties, the diagnosis of SIADH should rest on the original criteria and should specifically exclude other clinical conditions known to cause impairments in free water excretion even when these are mediated by secondary stimulation of vasopressin. Without maintaining these distinctions, arguable as some may be, the definition of SIADH becomes too broad to retain any practical clinical utility.

In most cases of SIADH, plasma levels of vasopressin are within normal physiologic ranges and abnormal only relative to the osmolality (Fig. 9-9). This is important for two main reasons:

1. Because the well-known vasoconstrictive effects of vasopressin do not come into play until much higher plasma levels are achieved (see "Physiology of Secretion of Vasopressin and Thirst"), hyponatremia can be ascribed to vasopressor effects of vasopressin.
2. Normal or low but nonsuppressible levels of vasopressin can cause sufficient impairment of free water excretion to produce hypo-osmolality, depending on exogenous fluid intake, as in psychiatric patients with polydipsia.

**Patterns of Vasopressin Secretion**

Studies of plasma vasopressin levels in patients with SIADH during graded increases in plasma osmolality produced by hypertonic saline administration have defined four patterns of secretion (Fig. 9-10):

1. Random hypersecretion of vasopressin.
2. Inappropriate nonsuppressible basal vasopressin release but normal secretion in response to osmoregular changes above basal plasma osmolality.
3. A reset osmostat system whereby vasopressin is secreted at an abnormally low threshold of plasma osmolality but otherwise displays a normal response to relative changes in osmolality.
4. Low or even undetectable plasma vasopressin levels despite classic clinical characteristics of SIADH.

The first pattern, unregulated vasopressin secretion, is often observed in patients with paraneoplastic vasopressin production. Resetting of the osmotic threshold for vasopressin secretion has been well described with volume depletion and edema-forming states with effective arterial blood volume, but most patients with a resetting osmostat are clinically euolemic and may have SIADH. It has been suggested that chronic hypo-osmolality itself may over time reset the intracellular threshold for osmoreceptor firing, but in animals chronic hyponatremia did not significantly alter the osmotic threshold for vasopressin secretion.

The best physiologic example of a reset osmostat is pregnancy, as discussed earlier. Perhaps the most perplexing aspect of the reset osmostat pattern is its occurrence in patients with tumors, which suggests that in some of these cases a tumor-related
Although a negative sodium balance occurs during the development of hyponatremia in patients with SIADH, eventually urinary sodium excretion simply reflects daily sodium intake.\textsuperscript{122} Thus, renal sodium wasting is excretion of sodium despite hyponatremia, but in reality there is a new steady state in which there is a neutral sodium balance.

Studies of long-term antidiuretic-induced hyponatremia in dogs and rats indicated that a large proportion of the hyponatremia was attributable to secondary Na\textsuperscript{+} losses rather than to water retention.\textsuperscript{123,124} However, the natriuresis did not actually worsen the hyponatremia; rather, it allowed volume regulation of ECF. Because of the secondary natriuresis in patients with SIADH, expanded plasma or ECF volumes are not found with tracer dilution techniques.\textsuperscript{125} Intrinsic renal mechanisms produced both diuresis and natriuresis in response to increases in renal perfusion pressures (so-called pressure diuresis) when vasopressin-infused animals were continually fluid loaded.\textsuperscript{126} However, it has not yet been proved that this mechanism is sensitive enough to detect the relatively mild degrees of volume expansion that accompany diuretic hyponatremias.

Another possibility is that natriuresis is mediated through increases in circulating natriuretic peptides such as atrial natriuretic peptide (ANP), which are elevated in SIADH into ranges capable of promoting renal sodium excretion.\textsuperscript{127} These possibilities are not mutually exclusive.

The degree to which hyponatremia may occur primarily as a result of natriuresis is controversial. Cerebral salt-wasting syndrome was first proposed by Peters and colleagues in 1950,\textsuperscript{128} as an explanation for the natriuresis and hyponatremia that sometimes accompany intracranial disease, particularly subarachnoid hemorrhage (SAH), in which up to one third of patients experience hyponatremia. After the description of SIADH in 1957, such patients were generally assumed to have hyponatremia secondary to vasopressin hypersecretion with a secondary natriuresis.\textsuperscript{129} However, clinical and experimental data have suggested that some patients with SAH and other intracranial diseases indeed have a primary natriuresis leading to volume contraction rather than SIADH.\textsuperscript{130,131,132} and the elevated plasma vasopressin levels may be physiologically appropriate for the degree of volume contraction. Some studies indicate that there is insufficient evidence of hypovolemia despite ongoing natriuresis,\textsuperscript{133,134} whereas others argue that the combined measures used to estimate ECF volume do support hypovolemia.\textsuperscript{135}

With regard to the potential mechanisms of natriuresis, both plasma and CSF ANP levels are elevated in many patients with SAH\textsuperscript{117} and have been found to correlate variably with hyponatremia in patients with intracranial diseases.\textsuperscript{136,137} However, because SIADH is also frequently associated with elevated plasma ANP levels, this finding does not prove causality. In other disorders of hyponatremia related to Na\textsuperscript{+} wasting (e.g., Addison's disease) and diuretic-induced hyponatremia, infusion of saline restores normal ECF volume and plasma tonicity by shutting off the secondary vasopressin secretion. In SAH, however, large volumes of isotonic saline sufficient to maintain plasma volume did not change the incidence of hyponatremia.\textsuperscript{138} In contrast, mineralocorticoid therapy to inhibit natriuresis reduced the incidence of hyponatremia in patients with SAH,\textsuperscript{139} but elderly patients with SIADH responded similarly to mineralocorticoid therapy.\textsuperscript{140} It seems most likely that SAH and other intracranial disorders represent a mixed disorder in which some patients have both exaggerated natriuresis and inappropriate vasopressin secretion; which effect predominates in the clinical presentation depends on their relative intensities as well as the effects of concomitant therapy.

The possibility that ANP-induced natriuresis may exacerbate hyponatremia is not confined to intracranial diseases, and it has been suggested that ectopic ANP production may contribute to or even cause the hyponatremia accompanying some small cell lung cancers.\textsuperscript{141} In several studies, small cell lung carcinoma was reported to produce ANP in addition to, or in some cases instead of, vasopressin.\textsuperscript{142} However, hyponatremia appeared to correlate more with plasma vasopressin levels than plasma ANP levels.\textsuperscript{143} Consequently, such cases may represent a mixture of inappropriate secretion of both hormones, and the ANP may further exacerbate the secondary natriuresis produced primarily by vasopressin-induced water retention.

**Renal Adaptation**

In addition to excretion of osmoles to bring volumes back toward normal, some adaptations allow excretion of more water. As stated earlier, vasopressin stimulates water retention by increasing the activity and content of aquaporin-2 water channels in the renal collecting duct epithelium. Chronic action of vasopressin in SIADH produces dramatic increases above normal of aquaporin-2 content and insertion into the epithelial cell membranes. This increases the efficiency of water retention and worsens the pathology. However, when vasopressin induces volume expansion and hypotonicity, the volume expansion and hypotonicity per se, by ill-defined mechanisms, act on the tubular cells of the collecting duct to decrease the content and action of aquaporin-2, thus decreasing the amount of water resorbed in spite of high vasopressin. This renal escape is another adaptation (besides natriuresis) that allows a patient with persistent SIADH to achieve a new steady state of Na\textsuperscript{+} and water balance with a low serum sodium level.\textsuperscript{144,145}

### Clinical Manifestations of Hyponatremia Disorders

Regardless of the etiology of hypo-osmolality, most clinical manifestations are similar. Non-neurologic symptoms are relatively uncommon, but a number of cases of rhabdomyolysis have been reported, presumably secondary to osmotically induced swelling of muscle fibers. Hypo-osmolality is primarily associated with a broad spectrum of neurologic manifestations, ranging from mild nonspecific symptoms (e.g., headache, nausea) to more significant disorders (e.g., disorientation, confusion, obtundation, focal neurologic deficits, seizures).\textsuperscript{146} This neurologic symptom complex has been termed hyponatremic encephalopathy\textsuperscript{147} and primarily reflects brain edema resulting from osmotic water shifts into the brain because of decreased effective plasma osmolality.

Significant neurologic symptoms generally do not occur until serum [Na\textsuperscript{+}] falls below 125 mEq/L, and the severity of symptoms is roughly correlated with the degree of hypo-osmolality.\textsuperscript{148,149} Individual variability is marked, however, and for any single patient the level of serum [Na\textsuperscript{+}] at which symptoms appear cannot be predicted. When the brain has volume-adapted through solute losses, thereby reducing brain edema, neurologic symptoms may even be virtually absent.\textsuperscript{150} From animal studies, the rate of serum [Na\textsuperscript{+}] decline is often more strongly correlated with morbidity and mortality than is the actual magnitude of the decrease.\textsuperscript{151} The reason is that the volume-adaptation process takes a finite period of time to complete, and the more rapid the decline in serum [Na\textsuperscript{+}], the more brain edema is accumulated before the brain can volumeregulate.

Thus, there is a much higher incidence of neurologic symptoms as well as a higher mortality rate in patients with acute hyponatremia than in patients with chronic hyponatremia.\textsuperscript{152,153} For example, the most dramatic cases of death related to hyponatremic encephalopathy have generally been reported in postoperative patients in whom hyponatremia developed rapidly as a result of intravenous infusion of hypotonic fluids.\textsuperscript{154,155} In such cases, nausea and vomiting are frequently overlooked as potential early signs of increased intracranial pressure; however, because hypo-osmolality does not have any direct effects on the gastrointestinal tract, unexplained nausea or vomiting in a hypo-osmolal setting should be assessed to have a CNS origin. Similarly, critically ill patients with unexplained seizures should be immediately evaluated for possible hyponatremia, because as many as one third of such patients have [Na\textsuperscript{+}] below 125 mEq/L as the cause of the seizure activity.\textsuperscript{156} Underlying neurologic disease and non-neurologic metabolic disorders (e.g., hypoxia,\textsuperscript{157} acidosis, hypercalcemia) can raise the level of plasma osmolality at which CNS symptoms occur.

In the most severe cases of hyponatremic encephalopathy, death results from respiratory failure after tentorial cerebral herniation and brain stem compression. One quarter of patients with severe postoperative hyponatremic encephalopathy manifested hypoxic/hypoxic respiratory failure, the expected result of brain stem compression, but three quarters had pulmonary edema as the apparent cause of the hypoxia.\textsuperscript{158}

Studies of acute hyponatremia after marathon races have shown hypoxia and pulmonary edema in association with brain edema.\textsuperscript{159} These results therefore suggest that hypoxia resulting from noncardiogenic pulmonary edema may be an early sign of developing cerebral edema even before the brain stem compression and tentorial herniation.

Clinical studies also suggest that menstruating women\textsuperscript{160} and young children\textsuperscript{161} may be particularly susceptible to the development of neurologic morbidity and mortality during hyponatremia, especially in the acute postoperative setting.\textsuperscript{162} However, other studies have failed to corroborate these findings.\textsuperscript{163,164} Consequently, the true clinical incidence and the mechanisms responsible for these sometimes catastrophic cases are not certain.
Therapy of Hypo-osmolar Disorders

Despite some areas of continuing controversy regarding correction of osmolality in hypo-osmolar patients, a relative consensus has evolved concerning the most appropriate treatment of this disorder. The following recommendations are summarized in the diagnostic and therapeutic flow diagram shown in Figure 9-11 (Figure Not Available).

Initial Evaluation

ECF volume status determines treatment of hyponatremia. If volume is expanded, the treatment of the underlying disease should take precedence over correction of plasma osmolality. This treatment often involves diuretic therapy, which should simultaneously improve plasma toxicity by stimulating excretion of hypotonic urine. If hyponatremia is present, the patient must be considered to have depletion-induced hypo-osmolality, in which case volume repletion with isotonic saline (0.9% NaCl) at a rate appropriate for the estimated fluid deficit should be initiated.

If diuretic use is known or suspected, the isotonic saline should be supplemented with potassium (30 to 40 mEq/L) even if serum [K+] is not low because of the propensity for total body potassium depletion to occur in such patients. Although generally the hypo-osmolar patient is clinically euovolemic, if a possibility of depletion-induced, rather than dilution-induced, hypo-osmolality exists, it is then appropriate to treat the patient initially with isotonic saline whether or not signs of hypovolemia are present. Improvement in and eventual correction of the hyponatremia verify solute and volume depletion. If the SIADH rather than solute depletion is present, administration of a limited volume (e.g., 1 to 2 L) of isotonic saline produces Na⁺ and water excretion without significantly changing plasma osmolality.

A patient who meets the essential criteria for SIADH but has a low urine osmolality should be observed with a trial of modest fluid restriction. If the hypo-osmolality is attributable to transient SIADH or severe polydipsia, the urine remains dilute and the plasma osmolality is fully corrected as free water is excreted. If the patient has the reset osmostat form of the disorder, however, the urine becomes concentrated at some point before the plasma osmolality and serum [Na⁺] return to normal ranges.

If either primary or secondary adrenal insufficiency is suspected, glucocorticoid replacement should be initiated immediately after the completion of a rapid ACTH stimulation test. A prompt water diuresis after initiation of glucocorticoid treatment supports a diagnosis of glucocorticoid deficiency, but absence of a quick response does not necessarily negate this diagnosis because several days of glucocorticoid replacement are sometimes required for normalization of plasma osmolality. If hypothyroidism is suspected, thyroid function tests should be performed. Replacement therapy is usually withheld pending these results unless the patient is obviously myxedematous.

If renal failure is present in a patient with hypo-osmolality, a more extensive evaluation of renal function is necessary before a course of treatment is selected.

Acute Treatment

For any significantly hyponatremic patient, one must decide how quickly the plasma osmolality should be increased and to what level. This decision depends on the risks of uncorrected hyponatremia and the risks of the correction. It has become clear that correcting severe hyponatremia too rapidly is dangerous because it is sometimes associated with pontine and extrapontine myelinolysis, a brain demyelinating disease that causes severe neurologic morbidity and mortality.

Consequently, appreciation of the appropriate therapy of this disorder requires understanding this disease as well as the pathophysiology underlying hyponatremic encephalopathy.

Pontine and Extrapontine Myelinolysis

It has become apparent that the demyelinating disease of central pontine myelinolysis (CPM) occurs with a significantly higher incidence in patients with hyponatremia and in both animal and human studies brain demyelination has been associated with the correction of existing hyponatremia. In animal models of chronic hyponatremia, this pathologic disorder is probably precipitated by the brain dehydration that has been demonstrated to occur after correction of serum [Na⁺] toward normal ranges. MRI in animals has shown that chronic hypo-osmolality predisposes to opening of the blood-brain barrier in rats after rapid correction of hyponatremia and that the disruption of the blood-brain barrier is highly correlated with subsequent demyelination. Opening the blood-brain barrier can lead to subsequent myelinolysis through an influx of complement, which is toxic to oligodendrocytes that manufacture and maintain myelin sheaths of neurons.

Although there has been considerable debate in the literature, studies in both patients and experimental animals support the concept that both the rate of correction of hyponatremia and the total magnitude of the correction over the first few days determine the risk of demyelination. In rats, an initial rate of correction of hyponatremia less than 20 mEq/L in 24 hours involved less risk, and clinical data indicate that the initial magnitude of correction represents the major risk factor related to subsequent neurologic morbidity and mortality. Initial reports implicated increases in serum [Na⁺] greater than 25 mEq/L over the first 24 to 48 hours of treatment, but later studies suggested the occurrence of CPM with increases in serum [Na⁺] of more than 12 mEq/L in 24 hours or 18 mEq/hour in 48 hours. Although overcorrection of hyponatremia to supranormal levels is also clearly a risk factor for neurologic deterioration, both clinical and experimental studies have found that demyelination occurred after corrections to serum [Na⁺] levels still below normal ranges. Both experimental studies and clinical reports have demonstrated that demyelination occurs regardless of the method used to correct the hyponatremia.

The susceptibility to demyelination after correction of hyponatremia is strongly influenced by the severity and duration of the preexisting hyponatremia. The more severe and prolonged the hyponatremia, the more solute loss that occurs during the process of brain volume regulation, and the loss of larger amounts of solute impairs the ability of the brain to buffer volume in response to subsequent increases in plasma osmolality. Clinical studies show that CPM rarely occurs in patients with a starting serum [Na⁺] above 120 mEq/L, does not appear to occur in patients with psychogenic polydipsia in whom hyponatremia develops acutely as a result of massive water ingestion, and is corrected rapidly by diuresis of the excess fluid.

Other independent risk factors for the occurrence of CPM are chronic alcoholism and malnutrition. It seems likely that the threshold for increases in serum [Na⁺] that increase the risk for CPM is lower in alcoholic and malnourished patients, and a case report of myelinolysis in a patient with beer potomania in whom the rate of correction stayed within the recommended guidelines supports this likelihood. Interestingly, uremia appears to protect hyponatremic patients from myelinolysis after rapid correction of hyponatremia, purportedly because urea acts as an intracellular osmolyte to stabilize intracellular volume and thereby reduces the degree of
brain dehydration produced after rapid correction of hyponatremia. 

The term central pontine myelinolysis is historically correct but anatomically too limited. Demyelination after correction of hyponatremia frequently occurs in white matter areas of the brain other than thepons. This occurrence led to the term osmotic demyelination syndrome (PMS) but the term pontine and extrapontine myelinolysis (PEM) would be more accurate. Apropos of the widespread nature of the neuropathologic lesions, a much broader range of neurologic disorders is now being reported in patients after correction of hyponatremia, including cognitive, behavioral, and neuropsychiatric disorders, presumably as a result of demyelination in subcortical, corpus callosal, and hippocampal white matter, and movement disorders, as a result of demyelination in the basal ganglia.

The presence of positive MRI findings strongly supports a diagnosis of PEM. But scans often fail to demonstrate the characteristic demyelinating lesions because scans are usually negative until sufficient time has passed (generally 3 to 4 weeks) after correction of hyponatremia and onset of neurologic symptoms. Although most cases of osmotically induced PEM have been reported with rapid correction of hyponatremia, the disorder has also been reported with severe hyponatremia in both animal models and patients. It is clear that one cannot predict with any degree of certainty which patients will develop demyelination. Many patients undergo rapid and large changes of serum Na⁺ without subsequent neurologic complications, which is true of experimental animals as well. Consequently, overly rapid correction of hyponatremia should be viewed as a factor that puts patients at risk for PEM but does not inevitably precipitate this disorder.

**Individualization of Therapy**

From the previous discussions of hyponatremic encephalopathy and PEM, it follows that optimal treatment of hyponatremia must entail balancing the risks of hyponatremia against the risks of correction for each patient individually. Three factors should be considered when one is making a treatment decision for a hypo-osmolar patient: (1) the severity of the hyponatremia, (2) the duration of the hyponatremia, and (3) the patient's neurologic symptoms.

**Acute Hyponatremia**

Cases of acute hyponatremia, arbitrarily defined as 48 hours in duration, are usually symptomatic if the hyponatremia is severe (i.e., < 120 mEq/L). These patients are at greatest risk for neurologic complications from the hyponatremia but rarely have demyelination, presumably because sufficient brain volume regulation has not yet occurred. Consequently, serum Na⁺ in such patients should be corrected relatively quickly.

Figure 9-11 (Figure Not Available) emphasizes that hypo-osmolar patients should always be evaluated quickly for the presence of neurologic symptoms so that appropriate therapy can be initiated, if indicated, even while other results of the diagnostic evaluation are pending. Postoperative patients, and particularly young women and children in some studies, appear to be at somewhat greater risk for rapidly progressing hyponatremic encephalopathy. They should be treated especially promptly, and administration of hypotonic fluids should be avoided in such patients postoperatively.

**Chronic Asymptomatic Hyponatremia**

Conversely, patients with chronic hyponatremia (arbitrarily defined as more than 48 hours in duration) who have minimal neurologic symptoms are at little risk from complications of hyponatremia itself, but demyelination can develop after rapid correction because of greater degrees of brain volume regulation through electrolyte and osmolyte losses. There is no indication to correct the hyponatremia rapidly, regardless of the initial serum Na⁺, and these patients should be treated by slower-acting therapies, such as fluid restriction.

**Chronic Symptomatic Hyponatremia**

Although the first two extremes have clear treatment indications, most hypo-osmolar patients have hyponatremia of indeterminate duration and varying degrees of neurologic symptoms. Such patients should be treated promptly, because of their symptoms, but with methods that allow a controlled and limited increase of hypo-osmolality. Some studies have suggested that correction parameters should consist of a maximal rate of correction of serum Na⁺ in the range of 1 to 2 mEq/L per hour as long as the total magnitude of correction does not exceed 25 mEq/L over the first 48 hours. Others recommend even more conservative parameters, with maximal correction rates of 0.5 mEq/L per hour or less and magnitudes of correction that do not exceed 12 mEq/L in the first 24 hours and 18 mEq/L in the first 48 hours.

A reasonable approach for treatment of individual patients would therefore entail choosing correction parameters within these limits, depending on the symptoms. In patients who are only minimally symptomatic, one should proceed at the lower recommended limits of 0.5 mEq/L per hour or less. In patients with more severe neurologic symptoms, initial correction at a rate of 1 to 2 mEq/L per hour (or even 3 to 5 mEq/L per hour in comatose or seizing patients who are at risk for imminent ventricular herniation and respiratory arrest) would be more appropriate.

Regardless of the initial rate of correction chosen, acute treatment should be interrupted when any of three end-points is reached:

1. The patient's symptoms are abolished.
2. A safe serum Na⁺ (> 120 mEq/L) is achieved.
3. A total magnitude of correction of 20 mEq/L is achieved.

At such a point, the active correction should be stopped and the patient treated with slower-acting therapies (e.g., oral rehydration, fluid restriction), depending on the cause of the hypo-osmolality. From these recommendations, it follows that serum Na⁺ levels must be carefully monitored at frequent intervals (at least every 4 hours) during the active phases of treatment to adjust therapy to keep the correction within these guidelines.

**Interventional (Active) Therapies for Acute Corrections**

Controlled limited corrections can be accomplished with either isosmotic or hypertonic saline infusions, depending on the etiology of the hypo-osmolality. Patients with volume-depletion hypo-osmolality (e.g., clinical hypovolemia, diuretic use, or urine Na⁺ < 30 mEq/L) usually respond well to isosmotic (0.9%) NaCl. However, patients with diuretic-induced hyponatremia are especially susceptible to rapid corrections for several reasons:

1. Such patients are usually only minimally volume-depleted.
2. They are often small, elderly women with correspondingly small plasma volumes.
3. With cessation of diuretic therapy, these patients often have a free water diuresis as their urinary diluting defect dissipates.
4. The hypokalemia that frequently accompanies the hyponatremia in such patients appears to be an additional risk factor for demyelination after correction.

Consequently, in the absence of marked neurologic symptoms, such patients should simply be prescribed a regular sodium diet (4 to 8 g/day) and should discontinue diuretics. If isotonic saline is infused, it should be done so judiciously (e.g., 50 to 75 mL/hour) with potassium replacement. Patients with euovolemic hypo-osmolality (including those with SIADH) generally do not respond to isotonic NaCl and are best treated with hypertonic (3%) NaCl solution given by continuous infusion. An initial infusion rate can be estimated by multiplying the patient's body weight in kilograms by the desired rate of increase in serum Na⁺ by approximately 1 mEq/L per hour and an infusion of 35 mL/hour increases serum Na⁺ by about 0.5 mEq/L per hour.

In patients with known cardiovascular disease, furosemide can be used to treat volume overload. It cannot be emphasized too strongly that it is necessary only to correct plasma osmolality acutely to a safe range rather than completely to normotremia.
Several days of restriction are usually necessary before a significant increase in plasma osmolality occurs. Only fluid, not salt, should be restricted. The degree of restriction required depends on urine output plus insensible fluid loss (generally, discretionary, nonfood fluids should be limited to 500 mL a day rapidly with vasopressin receptor antagonists will still be at risk for PEM, which has already been documented in animals, without measurable vasopressin levels) and should provide better therapy for patients with chronic hyponatremia. Patients whose hyponatremia is corrected too quickly diminish free water diuresis. 

Chronic Treatment

Fluid Restriction

The treatment of chronic SIADH entails a choice among several suboptimal therapeutic regimens. Any drugs known to be associated with SIADH should be discontinued or changed. Continued fluid restriction represents the least toxic treatment choice and is the preferred treatment for most cases of mild to moderate SIADH. Several points should be remembered when this approach is used:

1. All fluids, not only water, must be included in the restriction.
2. The degree of restriction required depends on urine output plus insensible fluid loss (generally, discretionary, nonfood fluids should be limited to 500 mL a day below the average daily urine volume).
3. Several days of restriction are usually necessary before a significant increase in plasma osmolality occurs.
4. Only fluid, not salt, should be restricted.

Because of the ongoing natriuresis, patients with chronic SIADH often have a negative total body [Na⁺] balance and therefore should be maintained with relatively high NaCl intakes unless otherwise contraindicated. Failure to improve after several days of confirmed negative fluid balance prompts reconsideration of other possible causes, including solute depletion and clinically inapparent hypovolemia.

Although it has not been confirmed in humans, in animals the expanded volume and hypotonicity of SIADH decrease the vasopressin-induced increase of aquaporin-2 water channels in the collecting duct and this escape allows more water excretion. 

When hyponatremia is caused primarily by polydipsia, ideally therapy should be directed at reducing fluid intake to normal ranges. Unfortunately, fluid restriction has proved difficult to accomplish in many cases. Patients with a reset thirst threshold are resistant to fluid restriction because of the thirst resulting from stimulation of brain thirst centers at higher plasma osmolalities. In some cases, the use of alternative methods to ameliorate the sensation of thirst (e.g., wetting the mouth with ice chips or using sour candies to increase salivary flow) can help to reduce fluid intake.

Fluid intake in patients with psychogenic causes of polydipsia is driven by psychiatric factors that have responded variably to behavioral modification and pharmacologic therapy. Several reports have suggested potential efficacy of the antipsychotic drug clozapine as an agent to reduce polydipsia and to prevent recurrent hyponatremia in at least a subset of these patients.

Pharmacologic Therapy

Pharmacologic intervention is reserved for refractory cases in which the degree of fluid restriction required to avoid hypo-osmolality is so severe that the patient is unable or unwilling to maintain it. Pharmacologic intervention should also be avoided initially in patients with SIADH that is secondary to tumors because successful treatment of the underlying malignant lesion often eliminates or reduces the inappropriate vasopressin secretion.

When pharmacologic management is necessary, the preferred drug is the tetracycline derivative demeclocycline. 

Pharmacologic intervention is reserved for refractory cases in which the degree of fluid restriction required to avoid hypo-osmolality is so severe that the patient is unable or unwilling to maintain it. Pharmacologic intervention should also be avoided initially in patients with SIADH that is secondary to tumors because successful treatment of the underlying malignant lesion often eliminates or reduces the inappropriate vasopressin secretion.

When pharmacologic management is necessary, the preferred drug is the tetracycline derivative demeclocycline. 

This agent causes nephrogenic diabetes insipidus, thereby decreasing urine concentration even in the presence of high plasma vasopressin levels. Appropriate doses of demeclocycline range from 600 to 1200 mg/day administered in divided doses. Treatment must be continued for several days to achieve maximal diuretic effects; consequently, one should wait 3 to 4 days before deciding to increase the dose. Demeclocycline can cause reversible azotemia and sometimes nephrotoxicity, especially in patients with cirrhosis. Renal function should therefore be monitored on a regular basis in patients receiving demeclocycline, and the medication discontinued if increasing azotemia is noted.

Other agents, such as lithium, have similar renal effects but are less desirable because of inconsistent results and significant side effects. Urea has also been described as an alternative mode of treatment for SIADH as well as other hyponatremic disorders.

Several drugs appear to decrease vasopressin hypersecretion in some cases (e.g., diphenhydantoin, opiates, ethanol), but responses have been erratic and unpredictable. Potential exceptions are agonists selective for kappa opioid receptors, which appeared to be more specific for inhibition of vasopressin hypersecretion in animal studies and in clinical trials successfully produced an aquaresis in patients with cirrhosis.

Vasopressin Antagonists

An antagonist of the kidney vasopressin V2 receptors would be the ideal agent for treatment of dilutional hyponatremia. 

Previous attempts at using peptide vasopressin receptor antagonists in humans were frustrated by species variability with regard to partial agonistic effects of such compounds. Several nonpeptide V2 receptor antagonists have been described that appear to overcome these problems. As of this writing, several of these compounds were already in clinical trials with promising results. It thus appears that we are poised to begin a new era in both the evaluation and treatment of patients with SIADH.

When selective V2 receptor antagonists are eventually approved for clinical use, clinical trials should enable investigators to answer some long-standing questions about the role of vasopressin receptor activation in producing antidiuresis as well as other potential effects in various disease states (e.g., hyponatremic patients without measurable vasopressin levels) and should provide better therapy for patients with chronic hyponatremia. Patients whose hyponatremia is corrected too rapidly with vasopressin receptor antagonists will still be at risk for PEM, which has already been documented in animals, but appropriate dosing and monitoring should allow successful adherence to the same guidelines for limited controlled correction that apply to other methods of treatment.
OXOTOCIN

Synthesis of oxytocin and electrophysiology of oxytocinergic neurons have been described earlier in the chapter. The normal physiologic regulation of oxytocin secretion and action is complicated by the fact that secretion and function of oxytocin vary markedly among different experimental mammals. There are sites of synthesis in the ovary and in various tissues of the uterus that differ among species. Because it is difficult to study pregnant women and human tissue, physiologic regulation of oxytocin secretion and function is less well known in humans than other species. The classic roles of oxytocin are uterine myometrial contraction at parturition and smooth muscle activation promoting milk let-down with nursing. 2, 5

Parturition

The isolation of oxytocin was followed quickly by the description of its ability to stimulate uterine contractions, and this was soon followed by clinical use of oxytocin as a uterotonic agent. Indeed, oxytocin is the most potent known stimulator of myometrial contraction. The uterine myometrial cells have intrinsic contractile activity, and it is necessary during pregnancy to maintain the uterus in a quiet state. In most species, this is accomplished by action of progesterone and by relaxin (produced by the corpus luteum and decidual tissue), which decrease uterine contractility. Late in pregnancy, parturition is initiated by synchronizing the myometrial activity so that contractions become contractions, with softening and dilatation of the cervix and rupture of the fetal membranes. Parturition is then completed by separation of the placenta and involution of the uterus.

Estrogen activates many of the events required to initiate and progress through parturition, whereas progesterone tends to inhibit these events. In several species, progesterone withdrawal may be an initiating factor of parturition. The mechanism of the increase of the estrogen/progesterone ratio, however, may vary considerably among species. In sheep, an increase in corticotropin-releasing hormone and vasopressin in the fetus near term produces a remarkable increase in ACTH and cortisol. The increased cortisol leads to decreased progesterone in the ewe, and this triggers parturition. In mice, the corpus luteum makes progesterone and estradiol throughout pregnancy. There is also a decrease in progesterone at delivery, but the progesterone decrease is not dependent on glucocorticoids.

In some species, such as mice and sheep, the corpus luteum is maintained throughout pregnancy and secretes progesterone. Uterotropism at the time of parturition causes an abrupt fall in progesterone. In humans and nonhuman primates, the corpus luteum is present only during the first trimester and estrogen and progesterone rise throughout pregnancy, although the rate of increase of estrogen is greater than that of progesterone as parturition approaches. A decreasing effect of progesterone and increasing estrogen in humans may be more important at a paracrine level in fetal membranes, where progesterone is inactivated and estrogen synthesis increased. It has also been reported that estrogen receptor increases in human fetal membranes at parturition whereas progesterone receptor remains stable. Thus, the cycle of events that are observed in plasma in some species may be reproduced distally as a paracrine action in primates.

In various species, several other hormones play a role in initiation of or completion of parturition, or both, including prostaglandins, endothelins, adrenergic agonists, glucocorticoids, and cytokines. The role of oxytocin in the complex interplay of these various agents is not well understood in humans.

The importance of prostaglandins is increasingly recognized in numerous species. The predominant prostaglandin, prostaglandin F 

2 (PGF 

2 ), is responsible for uteristoraxation of the corpus luteum in sheep and rodents, causing the decrease in progesterone described earlier. In fetal membranes, cyclooxygenase enzymes that stimulate synthesis of prostaglandins from arachidonic acid are increased at parturition and release prostaglandins from the uterus to act on the ovary to cause luteolysis. Prostaglandins may also help promote a controlled inflammatory response in the cervix to assist in thinning and dilatation.

In nonhuman primates, as parturition approaches, synchronized nocturnal contractions begin but revert to asynchronous contractions during the day. These synchronized nocturnal contractions correlate with plasma oxytocin levels and are blocked by administration of an oxytocin antagonist. Similarly, the response of the uterus to administered oxytocin in the monkey is greater at night than during the day, and fetal dehydroepiandrosterone sulfate (DHEAS) levels are also highest at night. DHEAS secretion by the fetus may be an important source of estrogen because the DHEAS is converted to estradiol by the placenta. Estradiol stimulates oxytocin synthesis in decidua and also stimulates synthesis of PGF 

2 in decidua.

As mentioned earlier, at this time there is an increase in estrogen receptors, which would favor the action of DHEAS to increase the estrogen response. Furthermore, there is a feed-forward mechanism whereby increased PGF 

2 stimulates oxytocin, which in turn stimulates increased production of PGF 

2 . Teleologically, it makes sense that the developing fetus, upon reaching maturity, would be a controlling factor in the initiation of labor. It was previously noted that in the sheep there is an absolute requirement for the hypotalamic-pituitary-adrenal axis to initiate labor, and the interaction of fetal DHEAS, estrogen, oxytocin, and progesterone provides a potential similar mechanism in primates (and humans?).

Although the role of oxytocin in the initiation of parturition is still debated, it is agreed that oxytocin is released explosively in a pulsatile fashion after the initiation of parturition and as parturition continues. This release of oxytocin is brought about by vaginal and cervical dilatation and is known as the Ferguson reflex. In animals with multiple births, the reflex release of oxytocin after one fetus passes the cervicovaginal canal may assist in the delivery of subsequent pups, but in humans this second phase of delivery is usually only to deliver the placenta. and may be more important in stimulating a clamping down of the uterine muscle to decrease blood loss.

Interestingly, studies with mice to "knock out" synthesis of oxytocin have found that in oxytocin-deficient mice parturition is initiated on time and proceeds normally. whereas mice deficient in cyclooxygenase and in prostaglandins have markedly prolonged labor. Administration of PGF 

2 to the deficient animals resulted in delivery of viable pups at the appropriate time.

From this brief review, it is clear that parturition is a complicated cascade of events that interact with each other at parturition and feed forward with cross-stimulation. It is not surprising that physiologic events as important to the species as pregnancy and parturition would have many redundant systems to ensure survival of the species. In addition, the complicated interaction of these various redundant systems and the feedback nature of the responses make it unlikely that interrupting any single hormonal response after parturition is initiated would be sufficient to inhibit completion of delivery.

In all of these discussions, there has been an obvious lack of understanding of the role of cytochrome aminopeptidase (oxytocinase) in the physiology of pregnancy in humans. If this enzyme developed as a protective mechanism, one would assume that oxytocin secretion by the neurohypophysis was increased throughout pregnancy, but the presence of this enzyme and the inability to study the hypothalamus in vivo make this possibility uncertain. However, the presence of the oxytocin-associated neurophysin throughout pregnancy supports this point of view. Furthermore, the presence of circulating cytochrome aminopeptidase provides a teleologic explanation for the development of oxytocin synthesis and secretion in fetal membranes in a manner that serves a paracrine function possibly protected from the degradative activity of the cytochrome aminopeptidase in plasma.
Lactation

The hypothalamic-pituitary hormones critical to lactation are prolactin and oxytocin. Prolactin storage and secretion from the anterior pituitary and its action to promote milk production are described in Chapter 8. Oxytocin is critical for milk secretion through the characteristic milk let-down response. Each of these hormones is influenced and regulated by gonadal steroid hormones. The milk-producing unit of the breast is the alveolar system, with multiple clusters of milk-producing cells surrounded by specialized myoepithelial cells. The alveoli are directly connected to ductules, and the ducts converge and lead to the nipple.

Milk is synthesized in the glandular cells of the alveoli. Oxytocin receptors are localized on glandular cells and oxytocin in the systemic circulation acts on these receptors to cause myoepithelial contraction. Oxytocin also acts on myoepithelial cells along the duct to shorten and widen the ducts, enhancing milk flow through the ducts to the nipple. Interestingly, in humans there are also oxytocin receptors on the epithelium of the gland and ducts, and oxytocin may have other indirect effects that enhance milk transport.

When an infant begins sucking at the breast, an afferent signal is transmitted from the mechanoreceptors or tactile receptors in the breast to the spinal cord and from the spinal cord to the lateral cervical nucleus. These ascending fibers cross in the medulla and eventually ascend to the oxytocinergic magnocellular neurons in the supraoptic and the paraventricular nucleus. Numerous neurotransmitters and neuropeptides are activated by different inputs to stimulate or inhibit the magnocellular neurons. Oxytocin itself is a regulator of oxytocin neurons. Oxytocin released by the dendrites acts on the same magnocellular neuron in an autocrine fashion, and an oxytocin antagonist infused into the hypothalamus decreases milk yield.

Although the exact role of each of these neurotransmitters or neuropeptides and their cooperative and competitive interactions has not been clarified in any animal (especially in humans), it is agreed that the final result is synchronous pulsatile depolarization of oxytocin neurons. The pulsatile release of oxytocin by the posterior pituitary can produce a pumping action on the alveoli, promoting maximal emptying of milk from the alveoli. Complete emptying of the breast is important to increase milk yield and if milk is not released from the breast there is involution of synthetic and secretory capacity.

The importance of oxytocin in maintaining milk secretion has been demonstrated by transgenic mice with a neurophysin construct that inhibited oxytocin synthesis. These animals delivered their young normally and had normal milk production, but there was no milk release despite normal suckling. The pups died of dehydration with no milk in the stomach. Administration of oxytocin to these oxytocin-deficient mice restored the ability to secrete milk and allowed the pups to survive. Similarly, oxytocin may promote successful lactation in women who have difficulty with lactation and milk production.

Whereas in most species suckling must occur for the sequence of events leading to milk let-down and, indeed, let-down may require several minutes of suckling. Suckling is, however, important in women because only suckling causes release of prolactin and suckling causes pulsatile release of oxytocin, whereas artificial massage of the breast produces continuous secretion of oxytocin. In women, if oxytocin is not secreted, only 20% to 30% of stored milk is released during nursing and secretion of oxytocin can be markedly inhibited by stress.

The role of steroids in oxytocin secretion is complex. Estrogen stimulates oxytocin release by dendrites, and progesterone withdrawal in an estrogen-primed animal stimulates oxytocin synthesis. Changes in these steroid hormones at the time of parturition probably modulate the lactation response both by modulating oxytocin synthesis and secretion and by modulating oxytocin receptors. Vaginal delivery and the pulses of oxytocin that are produced during the second phase of labor may enhance the pulses of oxytocin that later occur with suckling and lactation.

As breast-feeding continues in humans, the basal levels of oxytocin decrease but pulses of oxytocin in response to suckling continue and may increase. Women with diabetes insipidus have been able to breast-feed infants, and this has caused some to question the importance of oxytocin in humans. As noted earlier, however, oxytocin secretion may be preserved in the absence of vasopressin in patients with diabetes insipidus, even in those with traumatic section of the stalk.

The literature supports an interaction of oxytocin and prolactin. Receptors for prolactin have been described in the supraoptic nucleus, and prolactin enhances oxytocin release at the level of the pituitary and enhances synthesis of oxytocin when applied to magnocellular neurons. The importance of these findings has been questioned because prolactin probably does not cross the blood-brain barrier. Systemic administration of prolactin has been reported to stimulate release of oxytocin and immunoneutralization of prolactin to eliminate suckling-induced oxytocin release. Oxytocin was also reported to act as a prolactin-releasing factor. There may be a reinforcing interaction of oxytocin to stimulate release of prolactin and prolactin to enhance the release of oxytocin, producing a feedforward interaction.
Behavior

In addition to the actions in the periphery to stimulate milk let-down and uterine contraction during parturition, oxytocin is reported to have numerous actions on the CNS. The separate functions, peripheral and central, are mirrored anatomically by two sets of neurons: the magnocellular neurons and the smaller parvocellular neurons, respectively. Because the role of oxytocin secreted peripherally is to regulate physiologic events of reproduction, most of the emphasis regarding CNS effects has been on various aspects of maternal behavior. Maternal behavior is coincident with parturition and lactation in most mammalian species and is not seen with other physiologic events that produce changes only in gonadal steroid secretion. Therefore, it is likely that something in addition to or in place of gonadal steroids is active in inducing maternal behavior.

The most studied species regarding maternal behavior are sheep and rats. Studies have found that oxytocin is increased in various areas of the brain that are thought to be sites of regulation of maternal behavior, such as the olfactory lobes of rats. In some species, injection of oxytocin into the cerebral ventricles can initiate maternal behavior and injection of oxytocin antagonists can inhibit such behavior. However, the findings regarding oxytocin and reproductive behavior vary markedly among species and among strains within a given species.

Oxytocin has also been reported to play a role in males, but this is much less certain. In rodents, oxytocin administered in the CNS induced arousal and penile erection; in some species, oxytocin has been reported to increase sperm transport.
Other Actions of Oxytocin

In a number of reproductive tissues in a variety of species, oxytocin stimulates prostaglandin production. In several species, oxytocin is produced in peripheral organs rather than in the hypothalamus. There is some evidence for action of oxytocin in the menstrual cycle of humans, although this evidence is not as clear as that for the involvement of oxytocin in the estrous cycle of ruminants. Nonetheless, it is suggested that oxytocin may participate in a paracrine or autocrine fashion between cells in the human corpus luteum to stimulate the synthesis of progesterone.

In addition to accepted and postulated roles of oxytocin in reproductive behavior, numerous other central functions may be affected by central secretion of oxytocin, including the following:

1. Feeding behavior and satiety.
2. Gastric acid secretion.
3. Regulation of autonomic functions such as blood pressure, temperature, and heart rate.
4. Stimulation of glucagon to increase glucose.
5. Gonadotropin secretion.
7. Decreasing stress.
8. Stimulation of contractions of tubules and sperm transfer in the testes.

Some generalizations concerning these various central actions of oxytocin are that many of the purported actions are modulated by gonadal steroids, opioids appear to have a generally suppressive effect on oxytocin action, and there is great variability in the pattern of oxytocin receptors among species and considerable plasticity of the oxytocin receptors that may accompany physiologic changes. Oxytocin may have direct actions on receptors in the brain and may serve as a neurotransmitter or may modulate the response of classic neurotransmitters.
References


27. Giovannucci DR, Shnebel EL. Regulation of secretory granule recruitment and exocytosis at rat neurohypophysial nerve endings. J Physiol (Lond) 1997; 498:735751.


40(2):142145.

121.

120.

119.

118.

117.

116.

115.

114.

113.

112.

111.

110.

109.

108.

107.

106.

105.

1996; 45:273279.

104.

103.

102.

101.

100.

99.

98.

97.

96.

95.

94.

93.

92.

91.

90.

89.

88.

87.

86.

85.

84.

83.

82.

81.

80.

79.


509x796

318

317

316

315

314

313

312

311

310

309

308

307

306

305

304

303

302

301

300

299

298

297

296

295

294

293

292

291

290

289

288

287

286

285

284

283

282

281

280

279

278

277

276

275

274

273

272

271

270

269

268

267

266

265

264

263

262

261

260

259

258

257

256

255

254

253

252

251

250

249

248

247

246

245

244

243

242

241

240

239

238

237

236

235

234

233

232

231

230

229

228

227

226

225

224

223

222

221

220

219

218

217

216

215

214

213

212

211

210

209

208

207

206

205

204

203

202

201

200

199

198

197

196

195

194

193

192

191

190

189

188

187

186

185

184

183

182

181

180

179

178

177

176

175

174

173

172

171

170

169

168

167

166

165

164

163

162

161

160

159

158

157

156

155

154

153

152

151

150

149

148

147

146

145

144

143

142

141

140

139

138

137

136

135

134

133

132

131

130

129

128

127

126

125

124

123

122

121

120

119

118

117

116

115

114

113

112

111

110

109

108

107

106

105

104

103

102

101

100

99

98

97

96

95

94

93

92

91

90

89

88

87

86

85

84

83

82

81

80

79

78

77

76

75

74

73

72

71

70

69

68

67

66

65

64

63

62

61

60

59

58

57

56

55

54

53

52

51

50

49

48

47

46

45

44

43

42

41

40

39

38

37

36

35

34

33

32

31

30

29

28

27

26

25

24

23

22

21

20

19

18

17

16

15

14

13

12

11

10

9

8

7

6

5

4

3

2

1

318

317

316

315

314

313

312

311

310

309

308

307

306

305

304

303

302

301

300

299

298

297

296

295

294

293

292

291

290

289

288

287

286

285

284

283

282

281

280

279

278

277

276

275

274

273

272

271

270

269

268

267

266

265

264

263

262

261

260

259

258

257

256

255

254

253

252

251

250

249

248

247

246

245

244

243

242

241

240

239

238

237

236

235

234

233

232

231

230

229

228

319
References


320


321
References


Kahn A, Brachet E, Blum D. Controlled fall in natriemia and risk of seizures in hypotonic dehydration. Intensive Care Med 1979; 5:2371.


Kahn A, Brachet E, Blum D. Controlled fall in natriemia and risk of seizures in hypotonic dehydration. Intensive Care Med 1979; 5:2371.


Kahn A, Brachet E, Blum D. Controlled fall in natriemia and risk of seizures in hypotonic dehydration. Intensive Care Med 1979; 5:2371.


Kahn A, Brachet E, Blum D. Controlled fall in natriemia and risk of seizures in hypotonic dehydration. Intensive Care Med 1979; 5:2371.


Kahn A, Brachet E, Blum D. Controlled fall in natriemia and risk of seizures in hypotonic dehydration. Intensive Care Med 1979; 5:2371.


References


Oh MS, Carroll HJ. Cerebral salt-wasting syndrome: we need better proof of its existence. Nephron 1999; 82:110-114.


Oh MS, Carroll HJ. Cerebral salt-wasting syndrome: we need better proof of its existence. Nephron 1999; 82:110-114.


References


Dysfunction and anatomic abnormalities of the thyroid are among the most common diseases of the endocrine glands. This chapter provides an up-to-date physiologic and biochemical background and describes the various tests for evaluating patients with suspected disease of the thyroid gland.
PHYLOGENY, EMBRYOLOGY, AND ONTOGENY

Phylogeny

The phylogeny, embryogenesis, and certain aspects of thyroid function are closely interlinked with the gastrointestinal (GI) tract. The capacity of the thyroid gland to metabolize iodine and incorporate it into a variety of organic compounds occurs widely throughout the animal and plant kingdoms.

Monoiodotyrosine (3'-monoiodo-L-tyrosine [MIT]) and diiodotyrosine (3,5'-diiodo-L-tyrosine [DIT]) are present in a variety of invertebrate species, including mollusks, crustaceans, coelenterates, annelids, insects, and certain marine algae (Fig. 10-1). In these lower forms, however, no recognizable thyroid tissue is present. Thyroid tissue is confined to and is present in all vertebrates. A close link to the thyroid of higher vertebrates is evident in the ammocoete, the larval form of the lamprey. Here the endostyle is capable of carrying out iodinations, but prior to metamorphosis a protease is expressed in the endostyle that can hydrolyze the iodoprotein formed. Presumably, this permits the endostyle to lose its connection with the pharynx during metamorphosis and to assume its adult function as an endocrine organ that secretes iodothyronines, including 3,5,3',5'-tetraiodo-L-thyronine (thyroxine, T\(_4\)), and 3,5,3'-triiodo-L-thyronine (T\(_3\)) (Fig. 10-1).

The phylogenetic association of the thyroid gland and the GI tract is evident in several functions. The salivary and gastric glands, like the thyroid, are able to concentrate iodide in their secretions, although iodide transport in these sites is not responsive to stimulation by thyrotropin (also called thyroid-stimulating hormone [TSH]). The salivary gland contains enzymes that are capable of iodinating tyrosine in the presence of hydrogen peroxide (H\(_2\)O\(_2\)), although it forms insignificant quantities of iodoproteins under normal circumstances.
Structural Embryology

The human thyroid anlage is first recognizable about 1 month after conception, when the embryo is approximately 3.5 to 4.0 mm in length. The primordium begins as a thickening of epithelium in the pharyngeal floor, which later forms a diverticulum. With continuing development, the median diverticulum is displaced caudad and the primitive stalk connecting the primordium with the pharyngeal floor elongates (thyroglossal duct). During its caudal displacement, the primordium assumes a bilobate shape, coming into contact and fusing with the ventral aspect of the fourth pharyngeal pouch.

Normally, the thyroglossal duct undergoes dissolution and fragmentation by about the second month after conception, leaving at its point of origin a small dimple at the junction of the middle and posterior thirds of the tongue, the foramen caecum. Cells of the lower portion of the duct differentiate into thyroid tissue, forming the pyramidal lobe of the gland. Concomitantly, histologic alterations occur throughout the gland. Complex interconnecting cord-like arrangements of cells interspersed with vascular connective tissue replace the solid epithelial mass and become tubule-like structures at about the third month of fetal life; shortly thereafter, follicular arrangements devoid of colloid appear and eventually the follicles fill with colloid.
Functional Ontogeny

The ontogeny of thyroid function and its regulation in the human fetus are fairly well defined. Future follicular cells acquire the capacity to form thyroglobulin (Tg) as early as the 29th day of gestation, whereas the capacities to concentrate iodide and synthesize T4 are delayed until about the 11th week. Radioactive iodine inadvertently given to the mother would be accumulated by the fetal thyroid soon thereafter.

Because the capacity of the pituitary to synthesize and secrete TSH is not apparent until the 10th to 12th weeks, early growth and development of the thyroid do not seem to be TSH-dependent. Subsequently, rapid changes in pituitary and thyroid function take place. Probably as a consequence of hypothalamic maturation and increasing secretion of thyrotropin-releasing hormone (TRH), the serum TSH concentration increases between 18 and 26 weeks of gestation, after which levels remain higher than those in the mother. The higher levels may reflect a higher set-point of the negative feedback control of TSH secretion during fetal life than at maturity.

Thyroxine-binding globulin (TBG), the major thyroid hormonebinding protein in plasma, is detectable in serum by the 10th gestational week and increases in concentration progressively to term. This increase accounts, in part, for the progressive increase in the serum T4 concentration during the second and third trimesters, but increased secretion of T3 must also play a role because the concentration of unbound, or free, T4 also rises. The peripheral metabolism of T4 in the human fetus differs markedly from that in the adult both quantitatively and qualitatively. Overall, rates of production and degradation of T4 in unit per body mass exceed those in the adult by 10-fold. In addition, the enzymatic pathways by which T4 is metabolized differ from those in the adult, favoring the formation of the inactive 3,3',5'-triiodo-\(^{-}\)-thyronine (reverse T3 [rT3]) at the expense of T3.

From the clinical standpoint, several aspects of thyroid development are notable. In rare circumstances, thyroid tissue may develop from remnants of the thyroglossal duct near the base of the tongue. Such lingual thyroid tissue may be the sole functioning thyroid tissue present, and thus its surgical removal would lead to hypothyroidism. More commonly, elements of the thyroglossal duct may persist and later give rise to thyroglossal duct cysts, or thyroid tissue progenitors may migrate to occupy a place within the mediastinum.
ANATOMY AND HISTOLOGY

The thyroid gland is one of the largest of the endocrine organs, weighing approximately 15 to 20 g in North American adults. Moreover, the potential of the thyroid for growth is tremendous. The enlarged thyroid, commonly termed a goiter, can weigh many hundreds of grams.

The normal thyroid gland is made up of two lobes joined by a thin band of tissue, the isthmus. The latter is approximately 0.5 cm thick, 2 cm wide, and 2 cm high. The individual lobes normally have a pointed superior pole and a poorly defined, blunt inferior pole that merges medially with the isthmus. Each lobe is approximately 2.0 to 2.5 cm in thickness and width at its largest diameter and is approximately 4.0 cm in length. Occasionally, especially when the remainder of the gland is goitrous, a pyramidal lobe is discernible as a finger-like projection directed upward from the isthmus, generally just lateral to the midline, usually on the left. The right lobe is normally more vascular than the left, is often the larger of the two, and tends to enlarge more in disorders associated with a diffuse increase in size.

Two pairs of vessels constitute the major arterial blood supply: (1) the superior thyroid artery, arising from the external carotid artery, and (2) the inferior thyroid artery, arising from the subclavian artery. Estimates of thyroid blood flow range from 4 to 6 mL/minute per g, well in excess of the blood flow to the kidney (3 mL/minute per g). In diffuse toxic goiter due to Graves’ disease, blood flow may exceed 1 L/minute and may be associated with an audible bruit or even a palpable thrill.

The gland is composed of closely packed spherical units (follicles), which are invested with a rich capillary network. The interior of the follicle is filled with the clear, proteinaceous colloid that normally is the major constituent of the total thyroid mass. On cross-section, thyroid tissue appears as closely packed, ring-shaped structures consisting of a single layer of thyroid cells surrounding a lumen. The diameter of the follicles varies considerably, even within a single gland, but averages about 200 µm. The follicular cells vary in height with the degree of glandular stimulation, becoming columnar when active and cuboidal when inactive.

The epithelium rests on a basement membrane that is rich with glycoproteins separating the follicular cells from the surrounding capillaries. From 20 to 40 follicles are demarcated by connective tissue septa to form a lobule supplied by a single artery. The function of a given lobule may vary from that of its neighbors.

Under electron microscopy, the thyroid follicular epithelium has many features in common with other secretory cells and some peculiar to the thyroid. From the apex of the follicular cell, numerous microvilli extend into the colloid. It is at or near this surface of the cell that iodination, exocytosis, and the initial phase of hormone secretion, namely colloid resorption, occur. The nucleus has no distinctive features, and the cytoplasm contains an extensive endoplasmic reticulum laden with microsomes. The endoplasmic reticulum is composed of a network of wide, irregular tubules that contain the precursor of Tg. The carbohydrate component of Tg is added to this precursor in the Golgi apparatus, which is located apically. Lysosomes and mitochondria are scattered throughout the cytoplasm. Stimulation by TSH results in enlargement of the Golgi apparatus, formation of pseudopodia at the apical surface, and the appearance in the apical portion of the cell of many droplets that contain colloid taken up from the follicular lumen.

The thyroid also contains parafollicular cells (C cells) that are the source of the calcium-lowering hormone, calcitonin. These cells arise during embryonic development from the last pair of pharyngeal pouches but ultimately come to rest either among the cells of the follicular epithelium or in the thyroid interstitium. They differ from the cells of the follicular epithelium in never bordering on the follicular lumen and in being rich in mitochondria. The C cells undergo hyperplasia early in the syndrome of familial medullary carcinoma of the thyroid and give rise to this tumor in both its familial and its sporadic forms.
IODINE AND THE SYNTHESIS AND SECRETION OF THYROID HORMONES

Overview

The function of the thyroid gland is to generate the quantity of thyroid hormone necessary to meet the demands of the peripheral tissues. This requires the daily thyroidal uptake of sufficient iodide and its oxidation by thyroid peroxidase (TPO) to allow the synthesis of approximately 110 nmoles (85 µg) of T\(_4\), which is 65% iodine by weight. This requires the synthesis of a 660-kd glycoprotein homodimer, Tg. Tg contains specific tyrosine residues that are then iodinated at the apical portion of the thyroid cell to form mono- and diiodotyrosine (MIT and DIT) (see Fig. 10-2).

TPO-catalyzed coupling of two molecules of DIT, or one of DIT and one of MIT, leads to formation of T\(_4\) and T\(_3\), respectively, which are then stored as colloid, still as part of the Tg molecule. Pinocytosis of stored colloid leads to the formation of phagolysosomes, the colloid droplets in which Tg is digested, releasing T\(_4\), T\(_3\), DIT, and MIT as the droplet is translocated toward the basal portion of the cell. Thyroxine and T\(_3\) exit the cell into the capillaries, and DIT and MIT are deiodinated by an iodothyronine deiodinase to allow recycling of the iodide to iodinate newly synthesized Tg.

The synthesis of thyroid hormones requires the expression of a number of thyroid cellspecific proteins. In addition to Tg and TPO, the TSH receptor is also required to transduce the effects of extracellular TSH for efficient hormone synthesis. Several thyroid cellspecific proteins thyroid transcription factors 1 and 2 (TTF-1 and TTF-2) and PAX-8 stimulate transcription of the Tg and TPO genes. One or more of these proteins may also influence expression of the TSH receptor.

Although the biochemical details of these processes are beyond the scope of this discussion, those aspects with clinical relevance are detailed in the following sections. Excellent detailed reviews of these topics may be found elsewhere.
Dietary Iodine

Formation of normal quantities of thyroid hormone requires the availability of adequate quantities of exogenous iodine to allow thyroidal uptake of about 60 µg daily, taking into account the fecal losses of about 10 to 20 µg iodine of iodothyronines as glucuronides and about 100 to 150 µg as urinary iodine in iodine-sufficient populations. Plasma iodide (I⁻), the form of the element in biologic solutions, is completely filterable with about 60% to 70% of the filtered load reabsorbed passively. At least 100 µg of iodine per day is required to eliminate all signs of iodine deficiency (Table 10-1).

In North America, the daily dietary iodine intake is in the range of 150 to 300 µg daily, largely owing to the iodination of salt; in Japan, where large quantities of foods rich in iodine are consumed, intakes may be as high as several milligrams per day. Notably, iodine intake in the United States is decreasing as a result of a reduction in salt intake, with median urinary iodine of 15 µg/dL but a low urinary iodine (<5 µg/dL) in 12% of the population.¹² ¹³ ¹⁴

The daily dietary intake of iodine varies widely throughout the world, depending on the iodine content of soil and water and on dietary practice (see Table 10-1). Even in a single area, iodine intake varies among different individuals and in the same individual from day to day. Iodine may also enter the body via medications, diagnostic agents, dietary supplements, and food additives.

<table>
<thead>
<tr>
<th>TABLE 10-1 -- Recommended and Typical Values for Dietary Iodine Intake</th>
<th>µg/lday</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Recommended Daily Intake</strong></td>
<td></td>
</tr>
<tr>
<td>Adults</td>
<td>150</td>
</tr>
<tr>
<td>During pregnancy</td>
<td>200</td>
</tr>
<tr>
<td>Children</td>
<td>90120</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Typical iodine intakes</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>North America (1992)</td>
<td>75300</td>
</tr>
<tr>
<td>Chile (1981)</td>
<td>&lt;50150</td>
</tr>
<tr>
<td>Belgium (1993)</td>
<td>5060</td>
</tr>
<tr>
<td>Germany (1993)</td>
<td>2070</td>
</tr>
<tr>
<td>Switzerland (1993)</td>
<td>130160</td>
</tr>
</tbody>
</table>

As detailed later under the heading “Regulation of Thyroid Function,” iodine deficiency is common, especially in mountainous and in formerly glaciated regions of the earth. An estimated 1 billion people live in iodine-deficient areas of the world, and individuals living in such areas often develop TSH-induced compensatory enlargement of the thyroid (endemic goiter). If iodine deficiency is severe during pregnancy, fetal thyroid hormone production falls with irreparable damage to the developing central nervous system (CNS). This is manifested by varying degrees of mental retardation and is termed endemic cretinism. Thus, iodine-deficiency disorders (IDDs), including endemic goiter and cretinism, are the most common thyroid-related human illnesses indeed, the most common endocrine disorders worldwide.

Plasma iodide is partly replenished by that lost from the thyroid gland into the blood and by iodide liberated through deiodination of iodothyronines in peripheral tissues. Ultimately, however, the diet is its most important source. Iodine is ingested in both inorganic and organically bound forms. Iodide is rapidly and efficiently absorbed from the GI tract (within 30 minutes), and little is lost in the stool. In the body, iodide is confined largely to the extracellular fluid; however, it is also found in red blood cells and is concentrated in the intraluminal fluids of the GI tract, notably the saliva and gastric juice, from which it is reabsorbed, thus reentering the extracellular fluid. Iodide is also concentrated in milk.

Until it is oxidized and bound to tyrosyl residues in Tg, iodide entering the thyroid by active transport is in rapid equilibrium with the main iodide pool. The concentration of iodide in the extracellular fluid is normally 10 to 15 µg/L (10⁻⁷ M), and the content of the peripheral pool is approximately 250 µg. The thyroid gland contains the largest pool of body iodine (normally 8000 µg), most of which is in the form of DIT and MIT. Generally, this pool of iodine turns over slowly (1%/day).
Iodide Metabolism by the Thyroid Cell

Because the concentration of iodide in plasma is extremely low, a mechanism is required for the thyroid cell to concentrate the required amounts of this element. This process, called iodide trapping, is accomplished by a membrane protein, the sodium-iodide symporter (NIS). Human NIS is a 643 amino acid protein with 13 membrane-spanning domains.

The transport of iodide is an active process, depending on the presence of sodium gradient across the basolateral membrane of the thyroid cell such that downhill transport of 2 Na⁺ ions results in the entry of one iodide atom against an electrochemical gradient (see Fig. 10-2). In addition to being expressed in the basolateral membrane of the thyroid cell, NIS has also been identified in other iodide concentrating cells, including salivary and mammary glands, choroid plexus, gastric mucosa, and in the cytotrophoblast and syncytiotrophoblast. The iodide transport system generates an iodide gradient of 20 to 40 over the cell membrane and NIS also transports $\text{TCO}_4^-$, $\text{ClO}_4^-$, and SCN⁻, accounting for the utility of radioactive $\text{TCO}_4^-$ as a thyroid scanning tool and the capacity of potassium perchlorate ($\text{KClO}_4$) to block iodide uptake. In fact, these anions have a higher affinity for NIS than does iodide itself. On the other hand, the affinity of NIS for iodide is much higher than it is for the other inorganic anions, such as bromide and chloride, accounting for the selectivity of the thyroid transport mechanism.

It has been known for decades that the iodide-concentrating mechanism is required for normal thyroid function, as its absence is associated with congenital hypothyroidism and goiter unless large quantities of inorganic iodide are provided. A number of families have now been identified in which various mutations in the NIS gene are associated with congenital hypothyroidism and an iodide transport defect. Transcription of the NIS gene is increased by TSH. The mechanism for this has not been completely elucidated, but studies of the rat NIS promoter suggest that there is an NIS upstream enhancer, which confers a cyclic adenosine monophosphate (cAMP) response but also contains binding sites for the thyroid specific transcription factors PAX-8 and TTF-1, as well as a degenerate cAMP response element element. Importantly, several studies have documented decreases in NIS expression in human thyroid adenomas and carcinomas that contribute to the loss of iodine uptake in neoplastic thyroid cells, which thus present as "cold" nodules on radioisotopic imaging.

A second thyroid cell protein involved in iodide metabolism, pendrin, the product of the PDS gene, has now been identified by positional cloning using genomic DNA from families with the autosomal recessive disorder, Pendred's syndrome. This is a long-recognized inherited condition in which sensorineural hearing loss is combined with varying degrees of impaired thyroid hormone synthesis, leading to goiter. Pendrin is a transmembrane protein, a member of the sulfate transport protein family. Initially thought to be a sulfate transporter, it is now recognized to transport chloride, iodide, and bicarbonate ($\text{HCO}_3^-$). Pendrin is expressed in the apical border of the thyroid cell, the inner ear, and the kidney (see Fig. 10-2). Mutations in pendrin cause an inner ear malformation, although not all patients have goiter. It is postulated that pendrin is required for iodide transport across the apical membrane of the thyrocyte into the follicular lumen, where it is then oxidized and coupled to tyrosine in Tg (see Fig. 10-2).

The presence of thyroid dysfunction in Pendred's syndrome can be ascertained by the perchlorate discharge test, which illustrates the physiologic role of pendrin in thyroidal iodine metabolism. In normal individuals, more than 90% of thyroid radiiodine is present as iodotyrosine and iodothyronine within minutes of its entry into the thyroid. It is then no longer in the intracellular iodide pool. In patients with Pendred's syndrome, or with other disorders inhibiting the iodination of tyrosine (see later topics, such as Hashimoto's thyroiditis), this process is delayed, as shown by the exit (discharge) of more than 10% of the thyroidal radioiodine within 2 hours of administration of 500 mg of $\text{KClO}_4$. Perchlorate inhibits NIS function by an as yet unidentified mechanism eliminating the iodide gradient, which is required for maintaining the radiiodide in the gland. This illustrates that both iodide transport by NIS at the basal pole of the thyrocyte and its efflux across the apical membrane by pendrin are required for thyroid hormone synthesis. Deafness in patients with Pendred's syndrome is due to formation of a common cavity in the upper coils of the cochlea with dilatation of the vestibular aqueducts, not to the hypothyroidism per se.

In addition to being brought into the thyroid gland by active transport from the extracellular fluid, thyroidal iodide is generated by the deiodination of iodotyrosines liberated during the hydrolysis of Tg. A portion of this iodide is oxidized and used to iodinate tyrosine, and the remainder is lost from the gland as the iodide leak. This conservation process is interrupted when antithyroid drugs which inhibit iodide oxidation, such as methimazole (MMI), carbimazole (CB) or propylthiouracil (PTU) are given, thus further enhancing the effectiveness of these thyroid peroxidase inhibitors in blocking thyroid hormone synthesis.
Iodide Oxidation and Organification of Iodide

Within the thyroid gland, iodide participates in a series of reactions that lead to the synthesis of the active thyroid hormones. The first of these involves oxidation of iodide and incorporation of the resulting intermediate into the hormonally inactive iodotyrosines MIT and DIT, a process termed organification. Iodide is normally oxidized rapidly, immediately appearing in organic combination in Tg. The iodinations that lead to formation of iodotyrosines occur within Tg rather than on the free amino acids.

Oxidation of thyroidal iodide is mediated by the heme-containing protein TPO. The complementary DNA (cDNA) for human TPO encodes 933 amino acids with a molecular size of 103 kd, 10% of which is due to carbohydrate. The protein contains a membrane spanning region near the COOH terminus, and it is oriented in the apical membrane of the thyroid cell with residues 1 to 844 in the follicular lumen (see Fig. 10-2). TPO is the major thyroid microsomal antigen, and recombinant human TPO is now used for the detection of anti-thyroid microsomal antibodies, commonly present in the serum of patients with Hashimoto’s thyroiditis.

Because TPO is a heme protein, organic iodinations require molecular oxygen and are inhibited by cyanide and azide. In vitro, TPO, in the presence of H₂O₂, iodinates Tg as well as other proteins. The reaction catalyzed by peroxidase in vitro has many properties of the iodination reaction in vivo, including inhibition by PTU and MMI and by high concentrations of iodide (the Wolff-Chaikoff effect). The evanescent product of the peroxidation of iodide (i.e., the active iodinating form) may be free hypoiodous acid, iodine, or iodinium (I⁺). The H₂O₂ that serves as the oxidant of iodide is generated through the auto-oxidation of flavin enzymes acting as NADHand particularly NADPHoxidases. In this way, generation of H₂O₂ is linked to electron transfers due to substrate oxidations within the thyroid. Radioautographic and histochemical evidence suggests that the iodination reactions occur at the cell colloid interface (see Fig. 10-2). Thus mitochondrial systems provide a source of H₂O₂, cell membranes contain TPO, and the cytoplasmic fraction the regulatory inhibitors of organic iodinations.

The rate of organic iodinations depends on the degree of thyroid stimulation by TSH (see later). Iodinations are susceptible to inhibition by a number of pharmacologic agents, including the thiourea derivatives, PTU, MMI, and CB, which are inhibitors of peroxidase and have intrinsic reducing activity. Defects in the organic binding mechanism cause goitrous congenital hypothyroidism or, if less severe, goiter without hypothyroidism. In some families, thyroidal TPO is absent. In others, the defect may reside in inadequate production of H₂O₂ or in abnormalities in Tg that render it less readily iodinated (see Chapter 12).
Iodothyronine Synthesis

The MIT and DIT formed via oxidation and organic binding of iodide are precursors of the hormonally active iodothyronines T₄ and T₃. Because noniodinated thyronine cannot be demonstrated in Tg, T₄ and T₃ must arise from iodinated tyrosine precursors. Synthesis of T₄ from DIT requires the fusion of two DIT molecules to yield a structure with two diiodinated rings linked by an ether bridge and is catalyzed by TPO. Concomitantly, a residual dehydroalanine is formed at the site of the DIT residue contributing the phenolic hydroxyl group (beta or outer ring). This process is termed the coupling reaction.

Efficient synthesis of T₄ and T₃ in the thyroid requires Tg. The large (>260 kb) Tg gene is on chromosome 8. The Tg messenger RNA is 8 to 8.5 kb in length and encodes a 330-kd 12S subunit that is 10% carbohydrate by weight. There are 134 tyrosyl residues in the 660-kd homodimer. Only 25 to 30 of these are iodinated, but only residues 5, 1290, and 2553 form T₄ and residue 2746, T₃. The T₄-forming, readily iodinated, and iodothyronine-forming acceptor residues of Tg from different species are in a Glu/Asp-Tyr or a Thr/Ser-Tyr/Ser sequence, suggesting an important role of primary sequence in these reactions.

There are three to four T₄ molecules in each molecule of human Tg under conditions of normal iodination (25 atoms per Tg molecule, 0.5% iodine by weight), but only about one in five molecules of human Tg contains a T₃ residue. In Tg from patients with untreated Graves’ disease, the content of T₃ residues remains approximately the same, but the number of T₃ residues doubles to an average of 0.4 per molecule. This difference is independent of the iodination state of the Tg and is a consequence of thyroidal stimulation. Because the coupling reaction is catalyzed by TPO, virtually all agents that inhibit organic binding also inhibit coupling.
Storage and Release of Thyroid Hormone

The thyroid gland is unique among the endocrine glands by virtue of the large store of hormone it contains and the low rate at which the hormone turns over (1%/day). This aspect of thyroid hormone economy has homeostatic value, in that the reservoir provides prolonged protection against depletion of circulating hormone should synthesis cease. In normal humans, administration of antithyroid agents for as long as 2 weeks has little effect on the serum $T_4$ concentration. There are approximately 250 µg $T_4$/g wet weight in normal human thyroid or 5000 µg of $T_4$ in a 20-g gland. This is sufficient to maintain a euthyroid state for at least 50 days.

When $T_4$ is released rapidly in an uncontrolled fashion during subacute or painless thyroiditis, this can cause significant transient thyrotoxicosis. $T_g$ is present in the plasma of normal individuals at concentrations up to 50 ng/mL, leaving the thyroid gland through the lymphatics. However, peripheral hydrolysis of $T_g$ does not contribute significantly to the $T_4$ and $T_3$ in the circulation, even during thyroiditis, when large quantities of this protein are released.

The first step in thyroid hormone release is the endocytosis of colloid from the follicular lumen by two processes: macropinocytosis by pseudopods formed at the apical membrane, and micropinocytosis by small coated vesicles that form at the apical surface (see Fig. 10-2). Both processes are stimulated by TSH, but the relative importance of the two pathways varies among species. Micropinocytosis is thought to predominate in humans.

Following endocytosis, endocytotic vesicles fuse with lysosomes, and proteolysis is catalyzed by cathepsin - and ß-like thiol proteases, all of which are active at the acidic pH of the lysosome. The iodotyrosines released from $T_g$ are rapidly deiodinated by an NADPH-dependent iodotyrosine deiodinase, and the released iodine is recycled. The $T_4$ is released from $T_g$, but it is not clear how its transfer into the plasma is regulated. Release is acutely stimulated by TSH, as may be the $5'$-monodeiodination of small amounts of $T_4$ to $T_3$ by the types 1 and 2 iodothyronine deiodinases (D1 and D2), which are both expressed in human thyroid.

Although basal and TSH-stimulated conversion of $T_4$ to $T_3$ is easily demonstrated in the perfused canine thyroid gland, the contribution of thyroidal $T_4$ deiodination to $T_3$ secretion in humans under physiologic conditions is not known. The 15'/1 ratio of $T_4$ to $T_3$ in human $T_g$, as compared with the 10'/1 ratio of the secreted hormones, suggests minor $T_g$ monodeiodination. However, stimulation of D2-catalyzed $T_4$ 5'-deiodination in Graves’ thyroid may enhance that pathway and contribute to the relative increase in the ratio of $T_3$ to $T_4$ production in that condition.

$T_g$ proteolysis and $T_4$ release are inhibited by several agents, the most important of which is iodide. Inhibition of hormone release is responsible for the rapid improvement that iodide induces in hyperthyroid patients. The mechanism by which this effect is mediated is uncertain, but iodide inhibits the stimulation of thyroid adenylate cyclase by TSH and by the stimulatory immunoglobulins of Graves’ disease. Increasing iodination of $T_g$ also increases its resistance to hydrolysis by acid proteases in the lysosomes. Lithium inhibits thyroid hormone release, although its mechanism of action is poorly understood and may differ from that of iodide.
Role and Mechanism of Thyrotropin Effects

All steps in the formation and release of thyroid hormones are stimulated by TSH secreted by the pituitary thyrotrophs (see Chapter 8). Thyroid cells express the TSH receptor (TSHR), a member of the glycoprotein G protein-coupled-receptor family. The deduced amino acid sequence of this protein predicts a large extracellular NH₂-terminal domain, seven membrane-spanning domains, and an intracellular domain that transduces the signal to the Gs (adenyl cyclase) and GqII (phospholipase C [PL-C]) pathways (Table 10-2).¹ The TSHR also binds thyroid-stimulating antibody (TSAb) and thyroid-blocking antibodies (TBAb) (see Chapter 11).² In addition, the closely related luteinizing hormone (LH) and chorionic gonadotropin (CG) also bind to and activate TSHR signaling. CG accounts for the physiologic hyperthyroidism of early pregnancy.²³⁴⁵⁶⁷ Nonetheless, the persistence of a functional TSHR that is not down-regulated by autophosphorylation can explain the hyperthyroidism of Graves’ disease and that associated with TSH-producing thyrotroph tumors.²⁸²⁹ Interestingly, certain “activating mutations,” either germline or somatic, have been identified in the membrane spanning or intracellular portions of the TSHR molecule that cause generalized or nodular hyperfunction.³⁰³¹ Even more subtle changes may occur. In one family, a replacement of lysine 183 with arginine in the extracellular domain increased the activation of TSHR by human chorionic gonadotropin (hCG), causing recurrent gestational hyperthyroidism.³² Curiously, the TSHR may be cleaved into A and B subunits to varying extents, but the physiologic significance of cleaved and uncleaved receptors is unclear.³³

Studies of the rate-limiting intracellular reactions stimulated by TSH are complicated by the fact that the intrathyroidal events are stimulated by protein kinase A (PKA)-related mechanisms in one species or cell model and by PL-C-directed pathways in another [see Table 10-2].³⁴ The effects on

<table>
<thead>
<tr>
<th>Table 10-2 -- Thyroid Cell Functions Stimulated by Thyrotropin</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Function Affected</strong></td>
</tr>
<tr>
<td>Iodide Metabolism</td>
</tr>
<tr>
<td>Increase I in follicular lumen</td>
</tr>
<tr>
<td>Delayed increase in NIS expression</td>
</tr>
<tr>
<td>Increase thyroid blood flow</td>
</tr>
<tr>
<td>Increase in I efflux from thyroid cell</td>
</tr>
<tr>
<td>Thyroid Hormone Synthesis</td>
</tr>
<tr>
<td>Hydrogen peroxide</td>
</tr>
<tr>
<td>Thyroglobulin and TPO synthesis</td>
</tr>
<tr>
<td>NADPH via pentose-phosphate pathway</td>
</tr>
<tr>
<td>Thyroid Hormone Secretion</td>
</tr>
<tr>
<td>Pinocytosis of thyroglobulin</td>
</tr>
<tr>
<td>Release of thyroglobulin into plasma via basolateral membrane</td>
</tr>
<tr>
<td>Mitogenesis</td>
</tr>
</tbody>
</table>

iodide kinetics include both an early stimulation of iodide efflux into the follicular lumen and a later increase in the Vₘ₉₅ for iodide transport. The latter is likely due to enhanced expression of NIS.³⁶³⁷³⁸

Peroxidase generation in human thyrocytes is activated by Ca²⁺ and diacylglycerol, although 10-fold higher concentrations of TSH are required for activation of this limb of the TSHR pathways than for activation of adenylyl cyclase. TSH increases the levels of both TPO messenger RNA (mRNA) and protein, even though no cAMP response element-binding protein (CREB) sequences have been identified in the promoter of this gene.³⁹ The effect on TPO may be secondary to cAMP stimulation of thyroid cell-specific proteins, such as TTF-1, TTF-2, or PAX-8 (see earlier), but this remains to be shown. Similarly, transcription of the Tg gene is also stimulated by cAMP through indirect pathways, perhaps involving the same mechanism.⁴⁰

Interestingly, TSH can increase TSH mRNA in human thyroid cells, although at high TSH concentrations there may be a modest decrease in receptor expression. Nonetheless, the persistence of a functional TSHR that is not down-regulated by autophosphorylation can explain the hyperthyroidism of Graves’ disease and that associated with TSH-producing thyrotroph tumors.⁴¹ TSH, via cAMP, also stimulates the ingestion and hydrolysis of colloid and the release of T₄ (and T₃) from the thyroid cell. Thyroid cell proliferation is stimulated by cAMP, phorbol esters, and epidermal growth factor (EGF) through tyrosine kinases.⁴² However, cAMP causes proliferation while maintaining differentiated function, whereas EGF and phorbol esters lead to dedifferentiation. Similarly, insulin-like growth factor I (IGF-I) and fibroblast growth factor (FGF) stimulate cell division and dedifferentiation, although species differences exist in the effects of these agents.⁴³⁴⁴ In human thyrocytes, the events stimulated by the adenylyl cyclase pathway are probably the most crucial for thyroid growth and explain the goitrous changes associated with prolonged TSH stimulation, such as those occurring during iodine deficiency.
THYROID HORMONES IN PERIPHERAL TISSUES

Plasma Transport

The metabolic transformations of thyroid hormones in peripheral tissues determine their biologic potency and regulate their biologic effects. Consequently, an understanding of thyroid physiopathology requires a knowledge of the pathways of thyroid hormone metabolism.

A wide variety of iodothyronines and their metabolic derivatives exist in plasma. Of these, \( T_4 \) is highest in concentration and the only one that arises solely from direct secretion by the thyroid gland. In normal people, \( T_3 \) is also released from the thyroid but about 80% is derived from the peripheral tissues by the enzymatic removal of a single 5'-iodine atom (outer ring or 5'-monodeiodination) from \( T_4 \). The remaining iodothyronines and their derivatives are generated in the peripheral tissues from \( T_3 \) and \( T_4 \). Principal among them are 3,3',5'-triiodothyronine (T3), 3,3'-diiodothyronine (3,3'-T2), and 3,3',5'-triiodothyronine (3,3',5'-T3). Trace concentrations of other diiodothyronines, moniodothyronines, and conjugates thereof with glucuronic or sulfuric acid are also present.

The major iodothyronines are poorly soluble in water and thus bind reversibly to plasma proteins. The plasma proteins with which \( T_3 \) is mainly associated are TBG and transthyretin (TTR), formerly termed \( T_4 \)-binding prealbumin (TBPA), and albumin (Table 10-3). About 75% to 80% of \( T_3 \) is bound by TBG, with the remainder bound by TTR and albumin.

Thyroxine-Binding Globulin

TBG is a glycoprotein with a molecular mass of about 54 kDa, about 20% of which is carbohydrate. The gene that encodes the protein is on the X chromosome. The protein sequence of TBG resembles that of the SERPIN family of serine antiproteases. Because there is one iodothyronine binding site per TBG molecule, the \( T_4 \) or \( T_3 \) binding capacity of TBG in normal human serum is equivalent to its concentration, approximately 270 nmol/L (21 µg \( T_4 \)/dL). The half-life of the protein in plasma is about 5 days, and the metabolic clearance rate (MCR) is approximately 800 mL/day.

Congenital deficiencies of TBG are common, occurring in 1/5000 newborns, and are associated with the complete absence of the protein in males. Other abnormalities of TBG can alter the susceptibility to heat denaturation or the capacity to bind thyroid hormone. One such variant has been described in Australian aborigines, and a TBG protein with increased heat lability occurs in Africans. All of these abnormalities are inherited in an X-linked fashion.

The glycosylation of TBG influences its clearance from the plasma and its behavior during isoelectric focusing. Four to six bands are present; after exposure to neuraminidase, however, these differences are lost, indicating that they are due to variations in the numbers of sialic acid residues. In estrogen-treated patients, the prevalence of the more acidic bands of TBG is increased. The more highly sialylated TBG, compared with the more positively charged TBG, is cleared more slowly from plasma because increased sialylation inhibits the hepatic uptake of glycoproteins. The sera of pregnant patients, women receiving oral contraceptives, and individuals of aboriginal and African origin have abnormal TBG isoelectric patterns.

Other abnormalities of TBG can alter the susceptibility to heat denaturation or the capacity to bind thyroid hormone. One such variant has been described in Australian aborigines, and a TBG protein with increased heat lability occurs in Africans. All of these abnormalities are inherited in an X-linked fashion.

The metabolic clearance rate of TBG is inversely related to its binding capacity. The serum TBG concentration is highest in children and lowest in pregnant women. TBG concentration is also decreased in patients receiving drugs that induce hepatic enzymes, such as phenobarbital and drugs used to treat tuberculosis. In addition, TBG concentrations are increased in patients with liver disease, malnutrition, and pregnancy. TBG is significantly lower in patients with Congenital TBG deficiencies, occurring in 1/5000 newborns, and are associated with the complete absence of the protein in males. Other abnormalities of TBG can alter the susceptibility to heat denaturation or the capacity to bind thyroid hormone. One such variant has been described in Australian aborigines, and a TBG protein with increased heat lability occurs in Africans. All of these abnormalities are inherited in an X-linked fashion.

Another post-translational modification affecting TBG occurs in patients with sepsis or in patients after cardiopulmonary bypass surgery. TBG is subjected to

---

**TABLE 10-3 -- Comparison of the Major Human Thyroid Hormone-Binding Proteins**

<table>
<thead>
<tr>
<th>Protein Type</th>
<th>Thyroxine-Binding Globulin</th>
<th>Transthyretin</th>
<th>Albumin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mol wt of holoprotein (kDa)</td>
<td>54,000</td>
<td>54,000 (4 subunits)</td>
<td>66,000</td>
</tr>
<tr>
<td>Plasma concentrations (µmol/L)</td>
<td>27</td>
<td>4.6</td>
<td>640</td>
</tr>
<tr>
<td>( T_4 ) binding capacity as µg ( T_4 )/dL</td>
<td>21</td>
<td>350</td>
<td>50,000</td>
</tr>
<tr>
<td>Association constants of the major binding site (L/M)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( T_4 )</td>
<td>( 1 \times 10^{10} )</td>
<td>( 7 \times 10^7 )</td>
<td>( 7 \times 10^9 )</td>
</tr>
<tr>
<td>( T_3 )</td>
<td>( 5 \times 10^8 )</td>
<td>( 1.4 \times 10^7 )</td>
<td>( 1 \times 10^5 )</td>
</tr>
<tr>
<td>Fraction of sites occupied by ( T_4 ) in euthyroid plasma</td>
<td>0.31</td>
<td>0.02</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Distribution volume (L)</td>
<td>7</td>
<td>5.7</td>
<td>7.8</td>
</tr>
<tr>
<td>Turnover rate (%/d)</td>
<td>13</td>
<td>59</td>
<td>5</td>
</tr>
<tr>
<td>Distribution of iodothyronines (%/protein)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( T_4 )</td>
<td>68</td>
<td>11</td>
<td>20</td>
</tr>
<tr>
<td>( T_3 )</td>
<td>80</td>
<td>9</td>
<td>11</td>
</tr>
</tbody>
</table>

TBG excess have normal amounts of highly sialylated TBG, as do men and nonpregnant women. Because TBG is the principal \( T_4 \) - and \( T_3 \) -binding protein, changes in TBG or in its binding are paralleled by changes in total plasma \( T_4 \) and \( T_3 \) even though \( T_4 \) and \( T_3 \) production is little changed.

---

cleavage by a serine protease released from polymorphonuclear leukocytes resulting in the release of a 5-kD COOM terminal loop with a consequent decrease in affinity for T₄. An analogous reaction has been described for CBG that releases cortisol at the site of inflammation. 10 It has been postulated that the released T₄ might play a critical role in the response to injury, perhaps by providing a supply of iodine for antibacterial purposes. 11 12 The cleaved TBG (49 kd) circulates, and because it binds T₄ with lower avidity, this may explain the increased ratio of free T₄ to bound T₄ in acute illness, even when TBG saturation studies or immunoassays indicate a normal TBG concentration (see “Thyroid Function during Fasting or Illness”).

Transthyretin

TTR exists in part as a complex with retinol (vitamin A)-binding protein, hence its name. It consists of four identical polypeptide chains, with a total molecular mass of approximately 55 kDa, and is not glycosylated. Its concentration in plasma is approximately 4 mmol/L (250 µg/mL). Each mole of TTR binds 1 mole of T₃ with high affinity, and a second T₃ molecule is bound with lower affinity at high concentrations of T₄. 59 Binding of T₄ by TTR is independent of the association with retinol-binding protein. Its half-life in plasma is normally about 2 days, but this decreases during illness. 60

TTR is expressed in the choroid plexus and is the major thyroid hormone-binding protein in the cerebrospinal fluid (CSF). 61 Targeted TTR gene disruption in mice shows that aside from the predictable 50% decrease in total plasma T₄ concentration and reciprocal increase in the free T₄ fraction (since TTR is a major T₄-binding protein in mice), the absence of the gene causes no developmental abnormality. There is also no evidence of impaired uptake of T₄ into the brain, leaving the role of TTR in CSF undefined in regard to thyroid physiology. 61 62

Variant forms of TTR are associated with familial amyloidotic polyneuropathy. 63 64 In affected families, the TTR monomer has one of several different point mutations and TTR accumulates in amyloid tissue deposits. Neither thyroid dysfunction nor altered vitamin A metabolism has been reported, although there is altered affinity of some of the mutant proteins for T₄. Families with both high-affinity TTR and a few with increased TTR levels have been reported. 65

Comparison of Thyroxine and Triiodothyronine Binding by Thyroxine-Binding Globulin and Transthyretin

The TBG-binding site has an affinity for T₄ that is about 20-fold less than that for T₃ (see Table 10-3). TBG binds both the dextroisomer of T₄ and the naturally occurring levosomer. Deamination of the iodothyronine molecule reduces binding to TBG and increases the affinity for TTR, the acetic and propionic acid analogues of T₄ and T₃ bind poorly, if at all, to this protein. Binding of T₄ by TBG is inhibited by phenytoin, 65 salicylate, 66 salsalate, 65 67 furosemide, 65 fenofenac, 65 and miltoane. The affinity of these compounds for TBG is much weaker than that of T₄ or T₃ by TBG, but their concentration in plasma may be sufficient to interfere with T₄ and T₃ binding and reduce total hormone levels. Inhibitors of the T₄ TTR interaction include salicylate, saltsalate and some of its congeners, penicillin, and plant flavonoids. 68

Albumin

The affinity of albumin for T₄ and T₃ binding is much lower than that of either TBG or TTR, but the high concentration of this protein results in the binding of 10% of the plasma thyroid hormones (see Table 10-3). Changes in albumin concentration per se have little influence on the total hormone levels unless they are accompanied by alterations in TBG and TTR, all of which are synthesized in the liver. Hepatic failure or nephrotic syndrome leads to a decreased plasma concentration of all three, and the albumin concentration serves as a surrogate for estimating TBG concentrations.

The role of albumin in thyroid physiology becomes chemically important in patients with familial dysalbuminemic hyperthyroxinemia (FDH). 68 69 In this autosomal dominant disorder, the plasma contains high amounts of a usually minor albumin variant that binds T₄ (but not T₃) with increased avidity. This increases total T₄ levels, but free T₄ and total free T₃ levels remain normal in an otherwise euthyroid patient. In one family, the albumin variant was reported to bind T₃, but not T₄, with 40-fold higher affinity, resulting in dysalbuminemic hypothyriodothyroxinemia. 69

Other Plasma Thyroid Hormone-Binding Proteins

Between 3% and 6% of plasma T₄ and T₃ are bound to lipoproteins. 70 The T₄-binding lipoprotein is 27-kDa homodimer with an affinity for T₄ that is lower than that of TBG. This binding is of uncertain physiologic significance but may play a role in targeting T₄ delivery to specific tissues.
Free Thyroid Hormones

Because most of the circulating \( T_4 \) and \( T_3 \) is bound to TBG, its concentration and degree of saturation are the major determinants of the free fraction of \( T_4 \). Binding of the thyroid hormones to the plasma proteins alters their metabolism. The negligible urinary excretion of \( T_3 \) and \( T_4 \) is due to the limited filterability of the hormone-binding protein complexes at the glomerulus. The volume of distribution and rate of turnover of the hormones are also affected by their protein associations.

In vitro, the interaction between the thyroid hormones and their binding proteins conforms to a reversible binding equilibrium that can be expressed by conventional equilibrium equations. For the formulations that follow, \( T_4 \) is used as the prototype, with the understanding that similar interactions apply in the case of \( T_3 \). The interaction between \( T_4 \) and TBG can be expressed as follows:

\[
T_4 + TBG \overset{K}{\rightleftharpoons} T_4 \cdot TBG
\]

where TBG is the unoccupied binding protein, \( K \) is the equilibrium association constant for the interaction, and \( T_4 \) is the concentration of free \( T_4 \).

\( T_4 \) TBG is \( T_4 \) bound to TBG (almost equal to 66% of total \( T_4 \)). This interaction can also be expressed by the mass action relationship, wherein

\[
\frac{T_4 \cdot TBG}{(T_4)(TBG)} = K
\]

Rearranging

\[
\frac{T_4}{T_4 \cdot TBG} = \frac{1}{(TBG)K}
\]

Thus, the free fraction of \( T_4 \) is inversely proportional to the concentration of unoccupied TBG-binding sites. Estimates of the free \( T_4 \) concentration in serum can be generated by direct or indirect assay. For example, with the aid of radiolabeled \( T_4 \), the proportion that is unbound by protein is determined by dialysis, and the concentration of free \( T_4 \) can then be calculated as the product of the total hormone concentration and the fraction that is free. In normal serum, the free \( T_4 \) is approximately 0.02% of the total (20 pmol/L, 1.5 ng/dL). The approximately 20-fold lower affinity of TBG for \( T_4 \) results in a higher proportion of \( T_3 \) free (0.30%).

It is the free hormone that is available to the tissues for intracellular transport and feedback regulation, that induces its metabolic effects, and that undergoes degradation. The bound hormone acts merely as a reservoir. It follows that the concentration of the free hormone is the determinant of the metabolic state, and it is this concentration that is defended by homeostatic mechanisms. If an increase in the overall net binding affinity for \( T_4 \) occurs, the free \( T_4 \) concentration can be maintained at normal levels only if the bound \( T_4 \) increases. This is true whether or not the causative factor is an increase in the concentration of TBG or the presence of abnormal \( T_4 \)-binding proteins.

The plasma concentration of \( T_4 \) is determined by its rate of entry into, and exit from, the plasma. The MCR relates the quantity of \( T_4 \) removed from the plasma per unit time to the quantity available for removal (i.e., its plasma concentration). Thus,

\[
\text{MCR} = \frac{D}{[P]}
\]

where MCR is the metabolic clearance rate (volume/time), \( D \) is the absolute disposal or removal rate (amount/time), and \([P]\) is the plasma concentration (amount/volume). Transposing,

\[
[P] = \frac{D}{\text{MCR}}
\]

However, under steady-state conditions, the production rate (PR) of \( T_4 \) and the disposal rate (D) are equal. Hence,

\[
[P] = \frac{\text{PR}}{\text{MCR}}
\]

Thus, for any level of \( T_4 \) production, be it increased, normal, or decreased, the total plasma \( T_4 \) level varies inversely with its MCR. However, if only the free \( T_4 \) leaves the plasma and enters the cells while the bound \( T_4 \) is confined largely to the intravascular space, changing the fraction of total \( T_4 \) that is free, by changing the fraction that is available to the tissues, changes the MCR in a parallel manner. This explains, in part, why a primary increase in thyroid hormone binding, such as occurs when TBG concentrations are increased during pregnancy or by administration of excess estrogen, transiently reduces the free \( T_4 \) level and its clearance, causing an increase in the plasma total \( T_4 \) concentration.

The transient decrease in free thyroid hormones also reduces the negative feedback on the hypothalamic-pituitary-thyroid axis. This results in an increase in TSH secretion with a consequent increase in thyroid hormone production as an additional compensation. This can explain a portion of the increased levothyroxine (\( L-T_4 \)) requirement in the first trimester of pregnancy and the adaptation of the normal thyroid gland that occurs during estrogen administration.

This formulation is called the free thyroid hormone hypothesis. If it is free hormone that is available for cellular entry, what is the role, if any, of the hormone-binding proteins? These proteins permit distribution of the hydrophobic thyroid hormones throughout the vascular system. For example, if a protein-free solution containing tracer \( T_4 \) is perfused through rat liver via the portal vein, there is a steep concentration gradient with a decreasing quantity of \( T_4 \) in cells as the distance from the center of the portal lobule increases. In fact, virtually all of the \( T_4 \) is taken up by the first cells to be contacted by the bolus. In contrast, if serum albumin is added to the perfusate, the distribution of tracer is uniform throughout the lobule, with only 46% of the tracer removed from the bolus. Both influx and efflux of thyroid hormone from tissues are rapid. Thus, intracellular free \( T_3 \) and \( T_4 \) are in equilibrium with the free hormone pool in plasma. In the steady state, the rate of \( T_3 \) and \( T_4 \) metabolism (not the dissociation rate from plasma proteins) determines the rate of removal of these hormones from the plasma.
Cellular Uptake and Intracellular Binding

Carrier-mediated, energy-dependent transport of thyroid hormones has now been demonstrated in many cell lines. The carrier transport system for T\textsubscript{3} and T\textsubscript{4} is saturable, stereospecific, and requires adenosine triphosphate (ATP), but the two iodothyronines typically do not compete for uptake.

The proteins mediating this transport are the Na\textsuperscript{+}/taurocholate cotransporting polypeptide (NTCP), members of the Na\textsuperscript{+} independent organic anion transporter (OATP) family and the L type amino acid transporters. Information is still available only in preliminary form, but it is likely that there will be a variety of overlapping mechanisms for thyroid hormone transport and that different transporters will be present in different cell types.

In patients receiving a low-calorie diet for 1 week, T\textsubscript{3} and T\textsubscript{4} uptake into both rapidly equilibrating (liver and kidney) and slowly equilibrating pools (muscle) is decreased. Thus, inhibition of T\textsubscript{4} (and rT\textsubscript{3}) transport into these organs may well explain the low production of T\textsubscript{3} and the elevation of rT\textsubscript{3} concentrations in the serum of sick or fasting individuals (see later). The processes regulating efflux of thyroid hormones from cells are less well understood but appear to involve a verapamil-sensitive mechanism.

Cellular-Binding Proteins

Proteins that bind T\textsubscript{4} and T\textsubscript{3} are present in the cytosol, and some studies suggest that the cytosol-binding proteins for T\textsubscript{4} and T\textsubscript{3} are distinct from one another. The intracellular free T\textsubscript{3} concentration can be estimated by evaluating the binding affinity of intracellular and nuclear binding proteins and by measuring the T\textsubscript{3} concentration in cytosol. Such analyses suggest that there is a free hormone gradient of twofold to threefold across the plasma membrane. This concentration gradient, however, is not nearly as high as that calculated for the nuclear/cytosolic ratios in liver, kidney, heart, and brain, which are 50:1 to 250:1.

The mechanisms for maintaining these gradients have not been defined and have not been verified by direct methods. It is surprising that, despite such high ratios for free T\textsubscript{3} between the nucleus and cytoplasm, only 10% of the intracellular T\textsubscript{3} is present in the nucleus except in pituitary cells, where T\textsubscript{3} is equally divided between the nuclear and cytoplasmic compartments.
Thyroid Hormone Activation and Inactivation by the Selenodeiodinases

The most important pathway for \( T_4 \) metabolism is its outer ring (S') monodeiodination to the active thyroid hormone, \( T_3 \). This reaction is catalyzed by type 1 and type 2 deiodinases (D1 and D2) \(^{(1)}\). Inner ring deiodination, catalyzed primarily by type 3 deiodinase (D3), inactivates \( T_4 \) and \( T_3 \) \(^{(2)}\). These reactions can be considered physiologically activating and inactivating pathways that control \( T_4 \) concentrations in peripheral tissues.

The structures of the three human deiodinases are similar to one another and are conserved from tadpoles to humans. All three contain selenocysteine in the active catalytic center and hence are termed selenodeiodinases \(^{(3)}\). Selenocysteine has nucleophilic properties that make it ideal for catalysis of oxidoreductive reactions such as iodothyronine deiodination and the reduction of \( \text{H}_2\text{O}_2 \) by another family of selenoenzymes, the glutathione peroxidases. \(^{(4)}\) Selenium acts as the iodine acceptor during deiodination reactions \(^{(5)}\).

Mutagenesis of selenocysteine in D1 to cysteine (e.g., replacing Se with S) reduces the reaction velocity by 100-fold. \(^{(6)}\) Synthesis of selenoproteins is a complex process because the normal STOP translation function of the UGA codon which encodes selenocysteine must be overridden by the cell. This is accomplished by a combination of a specific structural feature, the SECIS element, in the 3′-untranslated region of the mRNAs encoding these proteins together with a specific group of selenocysteine-incorporating gene products. \(^{(7)}\)

### Comparative Enzymology and Regulation of the Selenodeiodinases

**Type 1 and 2**

Type 1 deiodinase has several characteristics that distinguish it from D2 and D3 \(^{(8)}\). It can catalyze both S- and S'-deiodination of \( T_4 \) to form \( T_3 \) and r\( T_3 \), respectively, but the \( K_m \) for these reactions is about three orders of magnitude greater than that of D2 and D3 for this substrate. In fact, the preferred substrates of D1 are r\( T_3 \) (S'-deiodination) and T\(_3\)SO\(_4\) (S-deiodination). \(^{(9)}\)

Unlike the deiodinations catalyzed by D2 and D3, D1-catalyzed reactions are susceptible to inhibition by PTU \(^{(10)}\). D1 also differs from D2 in being markedly increased by excess thyroid hormone through increased gene transcription, whereas D2 mRNA and protein are reduced during thyrotoxicosis and increased during hypothyroidism. \(^{(11)}\)

D1 and D2 are also expressed in different tissues; D1 is highly expressed in human liver and kidney, but human D2 is widely distributed in skeletal and cardiac muscle, the CNS, skin, and the pituitary gland. \(^{(12)}\) Furthermore, in some cell types, the subcellular location of D1 is the plasma membrane whereas that of D2 is the endoplasmic reticulum. \(^{(13)}\)

It is likely based on its subcellular location that the \( T_3 \) produced by D1 is formed close to the cell surface, whereas that generated by D2 is near the nucleus. This may explain why the \( T_3 \) produced by D1-catalyzed reactions readily enters the plasma whereas that generated by D2 enters the nucleus. \(^{(14)}\)

**Type 3**

Type 3 deiodinase is most highly expressed in placenta and in the gravid uterus in rats; it is also found in the CNS. \(^{(15)}\) The highest expression identified to date in humans is in infantile hemangiomas. In infants with extensive visceral lesions, thyroid hormone inactivation by D3 may overwhelm the secretory capacity of the infant's thyroid, causing hypothyroidism. \(^{(16)}\) D3 expression is increased by thyroid hormone at a transcriptional level thus providing a feedback loop to maintain \( T_3 \) homeostasis.

---

**Table 10-4** -- Human Iodothyronine Selenodeiodinases

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Type 1 (Outer and Inner Ring)</th>
<th>Type 2 (Outer Ring)</th>
<th>Type 3 (Inner Rings)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physiologic Role</td>
<td>( rT_4 ) and ( T_4 ) S degradation, source of plasma ( T_3 ) especially in hyperthyroid patients</td>
<td>Provide intracellular ( T_3 ) in specific tissues, source of plasma ( T_3 )</td>
<td>Inactivate ( T_4 ) and ( T_3 )</td>
</tr>
<tr>
<td>Tissue location</td>
<td>Liver, kidney, thyroid, pituitary (?) (not CNS)</td>
<td>CNS, pituitary, brown adipose tissue, placenta, thyroid, skeletal muscle, heart</td>
<td>Placenta, CNS, hemangiomas, fetal liver</td>
</tr>
<tr>
<td>Subcellular location</td>
<td>Plasma membrane</td>
<td>Endoplasmic reticulum</td>
<td>?</td>
</tr>
<tr>
<td>Preferred substrates (position)</td>
<td>( rT_3(5') ), ( T_3(5) )</td>
<td>( T_1 ) and ( rT_3(5') )</td>
<td>( T_1 ) and ( T_3(5) )</td>
</tr>
<tr>
<td>( K_m )</td>
<td>( 10^{-7} ) ( (rT_3) ), ( 10^{-8} ) ( (T_3) )</td>
<td>( 10^{-6} ) ( (T_3) )</td>
<td>( 10^{-5} ) ( (T_2 ) and ( T_4) )</td>
</tr>
<tr>
<td>Susceptibility to PTU inhibition</td>
<td>High</td>
<td>Absent</td>
<td>Absent</td>
</tr>
<tr>
<td>Response to increased ( T_4 )</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CNS, central nervous system; PTU, propylthiouracil; \( rT_3 \), reverse \( T_3 \).

---

In normal adults, T₄ has a distribution volume of approximately 10 L (Table 10-5). Because the concentration of total T₄ in plasma is approximately 100 nmol/L (8 µg/dL), the extrathyroidal T₄ pool is approximately 1 µmol (about 800 µg). In the adult, the fractional rate of turnover of T₄ in the periphery is about 10%/day (half-life, 6.7 days). Thus, about 1.1 L of the peripheral T₄ distribution space is cleared of hormone daily, a volume containing approximately 110 nmol (85 µg) of T₄.

The kinetics of T₄ metabolism differ from those of T₃, partly because of the 10-fold to 15-fold lower affinity of T₄ for TBG. The volume of distribution of T₄ in the normal adult is about 40 L, about four times that of T₃, and its fractional turnover rate is about 60%/day. Hence, the MCR of T₄ is about 24 L/day. At a mean normal serum T₄ concentration of 1.8 nmol/L (120 ng/dL), 50-fold lower than T₃, the daily production of T₄ is approximately 50 nmol (33 µg), or about 46% of that of T₃ (see Table 10-4). The rapid MCR of reverse T₃ and a low concentration in plasma (0.25 nmol/L, 15 ng/dL) combine to yield daily production rates for rT₃ of about 45 nmol. The turnover of 3,3'-T₂ (see Fig. 10-3) is even more rapid than that of T₄.

About 40% of secreted T₄ is monodeiodinated in the 5'-position to yield T₁, and a similar fraction is deiodinated in the 5-position to yield rT₂, the latter largely by D₃ (see Fig. 10-3). With a normal T₄ production rate of approximately 110 nmol (85 µg)/day, about 44 nmol (28 µg) of T₄ and rT₂ is produced by peripheral deiodination. Thus, 80% to 85% of T₄ and all of rT₂ production in humans can be accounted for by peripheral deiodination of T₄, findings consonant with the high ratio of T₄ to T₃ (15:1) and rT₂ to rT₃ in human Tg.

Of the T₄ generated via T₃-5'-deiodination in euthyroid humans, only 20% to 25% is inhibited by PTU. This suggests that in the euthyroid state, thyroidal T₃ secretion and D₁-catalyzed T₃ production may only account for about 40% of T₃ production. Alternatively, the inhibition by PTU may be incomplete and D₂-generated T₃ thereby underestimated. Whatever the contribution, the remainder of T₃ production is derived from D₂-catalyzed T₄ deiodination, perhaps catalyzed by the large pool of D₂ in skeletal muscle. This is consistent with the increase in fractional T₄ to T₃ conversion in hypothyroid or hypothyroxinemic subjects characteristic of D₂-mediated deiodination.

In contrast, PTU inhibits about 50% of peripheral T₃ production in the hyperthyroid patient consistent with an up-regulation of D₁ and down-regulation of D₂, which would be anticipated under these circumstances.

Although much of the T₄ and rT₂ produced from T₄ in peripheral tissues exits those tissues and enters the blood, an uncertain fraction of both are degraded intracellularly prior to their exit. As discussed later, in some D₂-containing tissues a significant fraction of T₄ in the cell nucleus is derived from intracellular local T₃ generation rather than from the plasma.

Other pathways are also involved in T₄ and T₃ metabolism. T₄ and T₃ undergo glucuronidation of the phenolic hydroxyl by the UDP-glucuronyltransferases (UDPGT) (see Fig. 10-4) (Figure Not Available). This pathway is clinically significant because certain pharmacotherapeutic agents such as phenytoin, rifampin and phenobarbital may enhance glucuronide conjugation, leading to biliary excretion into the intestine. Because T₄-G and T₃-G are not easily reabsorbed from intestinal contents, the significance of this pathway is that therapy with such agents generally increases ¹-iodothyronine requirements. In patients with an intact thyroid, this is not apparent, because internal adjustments increase the thyroid hormone production rate to compensate for the accelerated biliary excretion. In patients with hypothyroidism, however, an increase in ¹-iodothyronine dosage is required. Deamination and decarboxylation reactions that produce tetrac and triac and sulfation of T₄, and T₃ at the phenolic hydroxyl account for an as-yet-unidentified fraction of T₃ and T₄ metabolism in humans (see Fig. 10-4) (Figure Not Available).

T₃ is metabolized mainly by 5'-monodeiodination, either by D₃ orrafter sulfation in the liverby D₁. T₃ is metabolized by 5'-monodeiodination, primarily by D₁ (see Table 10-4). Both pathways yield 3,3'-T₂ (see Fig. 10-3), which is then rapidly degraded to moniodothyronines and thyronine. Thyronine and the iodide which escapes uptake by the thyroid gland are excreted in the urine.

Sources of Intracellular Triiodothyronine

In view of the differential tissue distribution of the various deiodinases, their various Kₗ values, and their differential regulation, it is not surprising that tissues may derive intracellular T₃ via several deiodinative pathways. Because T₃ regulates gene expression, it is especially relevant to analyze the quantity and source of nuclear T₃ in various tissues (Fig. 10-6).

In rat kidney and liver, D₁-expressing tissues, most nuclear T₃ is derived from plasma T₄. In the rat cerebral cortex, pituitary gland, and brown fat, all of which express D₂, half or more of intracellular T₃ is generated locally from T₄ within the tissue. This may be due in part to the differences in the subcellular localization between D₂ and D₁ that were mentioned earlier. In the rat, tissues depending on D₂ for nuclear T₃ are those in which a constant supply of thyroid hormone is crucial for either normal development (cerebral cortex), thyroid function (pituitary), or survival during cold stress (brown adipose tissue). These tissues are also characterized by a high
degree of saturation of the nuclear T3 receptors in comparison with other tissues (liver, kidney) in which nuclear T3 receptor sites are only about 50% occupied at normal serum T3 concentrations (see Fig. 10-6). This arrangement allows multiple levels of regulation of thyroid hormone action.

Intracellular D2-catalyzed T3 production has important implications for thyroid hormone physiology. First, because the T3 produced from T4 occupies a significant fraction of the receptors in those tissues, changes in either serum T4 or T3 can change receptor occupancy. However, because a fall in T4 also increases D2 protein half-life by decreasing the rate of ubiquitination and its proteasomal degradation, a rise in D2 activity mitigates the impact of a reduction of serum T4 in D2-expressing tissues, helping to maintain T3 homeostasis. Second, the requirement for both T3 and T4 for normal saturation of pituitary gland and CNS T3 receptors permits a response of the hypothalamic-pituitary axis to a reduction in plasma T4, which is the earliest manifestation of iodine deficiency or primary hypothyroidism (see "Regulation of Thyroid Function").

Because the D2 gene is positively regulated by cAMP, D2 activity and T3 production increase rapidly in brown adipose tissue under stimulation by the sympathetic nervous system during exposure to cold. This response is critical to adaptive thermogenesis during cold exposure in the human neonate and lifelong in the rodent.

Figure 10-6 Schematic diagram of the origin of the specifically bound nuclear triiodothyronine (T3) in various rat tissues. Data are derived from studies in which double-isotope labeling techniques were used to estimate the sources of specifically bound nuclear T3. In tissues having a receptor saturation significantly greater than 50%, the additional T3 is provided by D2-catalyzed T4 to T3 conversion. T3 in rat plasma is derived from thyroid secretion (40%) with the remainder from D1-and D2-catalyzed T4 to T3 conversion. BAT, brown adipose tissue; PIT, pituitary gland.
Physiologic, Pathologic, and Pharmacologic Influences on Thyroid Hormone Deiodination

There are a number of circumstances in which thyroid hormone activation is either inhibited or the degradation of thyroid hormone is accelerated via deiodination pathways. There are several categories: (1) physiologic, (2) pathologic, and (3) pharmacologic (Table 10-6).

Physiologic Influences

Physiologic activation of T₄ to T₃, particularly by D1, is impaired during fetal life. In experimental animals, however, D3 activity is increased in the fetus, particularly in the skin and in other tissues as well, contributing to a reduced serum T₃ level in the fetal state. In contrast, D2 action is sufficient to provide intracellular T₃ in those tissues, such as the brain, where thyroid hormone is absolutely required for normal development.

Careful chronic activation of D2 or D3 has been demonstrated in several animal models. These include a short-lived burst of D2 activity in the cochlea of the newborn mouse and the increased D3 activity in the retina during metamorphosis in the tadpole. Similar, although not yet documented, events may occur in humans.

Pathologic Influences

Alterations in iodothyronine deiodination occur during fasting and illness. This group of conditions is associated with a marked decrease in T₃, elevated serum rT₃ levels, and a decrease in T₃ clearance. These changes in thyroid hormone metabolism probably reflect both decreased T₄ transport and D1 and D2 actions. Tumor necrosis factor- (TNF-) and interleukin-1 decrease D1 expression in isolated hepatocytes, and TNF- decreases D2 expression in human skeletal muscle cells. TNF and IL might be playing a role in the fasting or illness-induced reductions in T₃ activation. These events are part of the global response that reduces plasma T₃ during illness (see "Thyroid Function during Fasting or Illness").

Selenium deficiency may be due to endemic deficiency (e.g., as in western China) or to the effects of protein-restricted diets. Because of the differential retention of selenium in various organs, selenium deficiency predominantly affects the liver and kidney, organs expressing D1. Under these circumstances, the serum T₄ level and the serum T₃/T₄ ratio are increased, a compensatory response to decreased T₃ production via the D1 pathway during this condition.

In Zaire, where both iodine and selenium are deficient, replenition of selenium prior to replenition of iodine led to a deterioration in overall thyroid function, presumably because of the acceleration of T₂ degradation by D1, and perhaps by D3, in the selenium-deficient subject. Hepatic dysfunction may reduce T₃ activation as a consequence of the decrease in D1-containing cells or because of the effects of the accompanying illness on deiodinase activity. Changes in deiodinase function with altered thyroid states have been reviewed in Table 10-4.

Pharmacologic Influences

A number of commonly used drugs have significant effects on thyroid hormone deiodination. PTU causes a dose-related inhibition of T₃ production, which is most readily seen in the thyrotoxic patient in whom T₄ to T₃ conversion via the D1 pathway is markedly increased. This inhibitory capacity is not shared by methimazole or carbimazole.

TABLE 10-6 -- Factors Altering the Peripheral Activation or Inactivation of Thyroid Hormones

<table>
<thead>
<tr>
<th>Factor or Condition</th>
<th>Tissue Uptake</th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physiologic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fetus</td>
<td>?</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pathologic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fasting</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Illness</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Selenium deficiency</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>Hepatic dysfunction</td>
<td></td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thyrotoxicosis</td>
<td></td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypothyroidism</td>
<td></td>
<td>?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hemangioma</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pharmacologic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Propylthiouracil</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>Amiodarone</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iopanoic or iopodic acid</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Growth hormone</td>
<td></td>
<td>(?)</td>
<td>(?)</td>
<td></td>
</tr>
<tr>
<td>Gold thioglucose</td>
<td></td>
<td>N</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>Thyroid hormone</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The antiarrhythmic drug amiodarone shares sufficient structural similarity with T₄ that it can inhibit deiodination of T₄ and rT₃ (Fig. 10-7). Patients receiving this agent develop a compensatory increase in plasma T₃ to maintain serum T₃ in the normal range. There is also a corresponding increase in TSH within the first weeks of therapy that gradually returns to normal as the thyroid axis reequilibrates. The T₃ and rT₃ MCRs are reduced by 20% to 25%, with a reduction in the fractional T₃ to T₄ conversion rate of about 50%. There is direct evidence of inhibition of D1 activity in rats given this agent, but no studies have examined its effects on D2. The mechanism is probably due to competitive inhibition either by the drug or by one of its metabolites, but inactivation of D1 and accelerated D2 degradation may also occur. Amiodarone also inhibits the active transport of T₄ and T₃ into hepatocytes, and the drug or one of its products may interfere with T₃ binding to
thyroid hormone receptors.

The effects of amiodarone resemble those observed with the iodoaniline derivatives formerly used for visualization of the gallbladder (see Fig. 10-7). Iopanoic or iopodic acid inhibits the deiodinases by competing with the iodothyronine substrates. This makes these agents useful in the acute treatment of patients with severe hyperthyroidism, in whom they cause a rapid decrease in T3.

High dosages of glucocorticoids (10 times replacement) acutely reduce the T3/T4 ratio in plasma, suggesting that T4 to T3 conversion is blocked. The rT3/T4 ratio increases, suggesting that D3 action is also increased. These effects resolve during chronic therapy such that thyroid function is little affected, and thyroid hormone requirements are not increased by chronic glucocorticoid therapy.

Figure 10-7 Comparison of the chemical structure of thyroxine (T4) with the structures of two agents that block the deiodination of the iodothyronines. The inhibition of T4 to triiodothyronine (T3) conversion, which occurs in patients receiving amiodarone, may be due to the drug itself or to a metabolic product. Iopanoic acid and related iodoanilines are competitive inhibitors of all three iodothyronine deiodinases.

Recombinant growth hormone increases circulating T3 and T4 levels and reduces rT3/T4 ratios. This result is also seen in patients with hypothyroidism who are receiving L-thyroxine, indicating it is a peripheral effect. In the chick, growth hormone reduces D3 activity, suggesting that a similar effect on this enzyme may be the explanation for its effect in humans; however, direct evidence is lacking.
Mechanism of Thyroid Hormone Action

Thyroid hormone acts by binding to a specific nuclear DNA-bound thyroid hormone receptor (TR), usually as a heterodimer with retinoid X receptor (RXR) at specific sequences (thyroid hormone response elements (TREs)) dictated by the DNA binding-site preferences of the RXR-TR complex. The general mechanism by which nuclear receptor-activating ligands, such as T3, produce their effects is discussed in Chapter 4.

T3 has a 15-fold higher binding affinity for TRs than does T4, which explains its function as the active thyroid hormone. In humans, two TR genes (and) are found on different chromosomes (TR on chromosome 17, TR on chromosome 3). Several alternatively spliced gene products from each of these genes form both active and inactive gene products. The active proteins are TR-1 and TRs 1, 2, and 3. The structure of the TRs conforms to a protein with three major functional domains, one binding DNA, one binding ligand, and two major transcriptional activation domains. These activation domains of the TRs and are similar to, but not identical with, the major differences in the amino-terminal portion of the molecule.

There are tissue-specific preferences in expression of the various TRs, suggesting that they subserve different functions in various tissues. In general, TR-2 is down-regulated in the thyroid gland, where regulation of thyroid function occurs. In addition to differences in the amino-terminus between TR-1 and 2, the two proteins are under the regulation of different promoters that can function in tissue-specific patterns.

TR-2 is down-regulated by T3, whereas TR-1 mRNA expression is not affected. TR-2 is also expressed in the cochlea. TR-1 is expressed in all tissues, although its mRNA is especially highly expressed in the kidney, liver, brain, and heart. TR-1 mRNA is also expressed in the brain and at lower levels in skeletal muscle, lungs, and heart. TR-3 mRNA is expressed at very low levels but is more abundant in the liver or kidneys and lungs in comparison with other tissues.

It is instructive to examine the effects of gene targeting of TR and TR in order to understand their different physiologic roles. A disruption of the TR gene (both TR-1 and 2) causes deafness, a marked reduction in feedback sensitivity of hypothalamic-pituitary-thyroid axis, and a decrease in hepatic D1. Thus, in these mice, both TSH and thyroid hormone levels are elevated. However, TSH levels are further increased if thyroid hormone levels are reduced, indicating a feedback suppression of TSH release, albeit attenuated by this genetic manipulation. Phenotypically, thyroid function in these mice resembles that in families with resistance to thyroid hormone (RTH) in which TR mutations markedly reduce its binding affinity for T3. Curiously, the clinical manifestations in these individuals may resemble thyrotoxicosis for reasons discussed in Chapter 12.

Despite evidence of impaired feedback regulation, there is relatively little abnormality in the brain and heart of the TR-deficient mice. The effect of a TR-1 disruption in the mouse is quite different. The predominant phenotypic effects are modest bradycardia and hypothermia. These animals have subtle but significant abnormalities in myocardial function and electrical activity and, despite the basal hypothyroidism, have normal responses to cold exposure.

These studies have led to the generalization that feedback regulation of thyroid hormone effects, along with cochlear development, are functions of TR, whereas cardiac functions and energy metabolism are probably regulated by TR. However, this is a generalized feature, and further experimental studies of these animals, as well as animals with specific mutations in the TR gene analogous to those in families with thyroid hormone resistance, are currently in progress. It is also likely that small differences in the ligand-binding domains of TR and TR will allow design of thyroid hormone analogues selective for one or the other of these receptors.

The binding of T3 to the TR-TRE complex of a gene positively regulated by T3 initiates a conformational change in the TR such that one or more repressors of transcription are dissociated from the receptor and replaced by coactivator proteins. Some of these coactivating proteins induce DNA acetylation, or histone acetyltransferase activity, making the neighboring thyroid hormoneregulated genes dissociate from nucleosomes and available for binding of the transcriptional initiation complex.

Another group of genes, as exemplified by those encoding the TSH subunit, the -glycoprotein subunit, and TRH, are negatively regulated by T3. Negative regulation by thyroid hormone is still poorly understood but may involve specific negative TREs located in the promoter or even the coding regions of target genes. In one scenario, the complexing of T3 with DNA-bound TRs recruits corepressors and histone deacetylases to these binding sites. An example of this mechanism is that regulating the subunit of TSH. In the absence of T3, a gene containing a TR bound to a negative TRE is activated by association of coactivator and histone acetylating activator proteins, which then dissociate as a result of T3. Other mechanisms of negative regulation by T3 have also been suggested, including TR trapping and sequestration of coactivators to prohibit positive regulation of these genes.

Other potential mechanisms of thyroid hormone action by interaction with the membrane are under investigation. The pivotal role of T3 in the activation of thyroid hormone-dependent genes indicates that T3 serves primarily as a prohormone. Because T3 down-regulates D2 and inactivates D1, however, it might be considered an active hormone in the sense that it negatively regulates its own activation.
addition, there is an inverse relationship between the glandular organic iodine level and the rate of hormone formation. Such autoregulatory mechanisms serve to stabilize the rate of hormone synthesis despite fluctuations in the availability of iodine. Stability in hormone production is achieved, in part, because the large intraglandular store of hormone buffers the effect of acute increases or decreases in hormone synthesis. Autoregulatory mechanisms within the gland, in turn, tend to maintain the constancy of the thyroid hormone pool.

Finally, the hypothalamic-pituitary feedback mechanism senses variations in the availability of free thyroid hormones, however small, and acts to correct them. There is a close relationship between the hypothalamus, the anterior pituitary gland, the thyroid gland, and still higher centers in the brain, with the function of the entire complex being modified in a typical negative-feedback manner by the availability of the thyroid hormones. Additional hormones and neuropeptides also influence this axis (see Chapter 7 and Chapter 8).

Thyrotropin-Releasing Hormone Synthesis and Secretion

TRH, a modified tripeptide (pyroglutamyl-histidyl-proline-amide), is derived from a large pre-pro-TRH molecule that contains five progenitor sequences. The TRH peptides are released from the pre-pro molecule by a peptidase that acts at flanking lysine/arginine residues. TRH is expressed in the hypothalamus, the brain, the C cells of the thyroid gland, the beta cells of the pancreas, the myocordium, the reproductive organs (including prostate and testes), and the spinal cord. The parvocellular region of the paraventricular nuclei of the hypothalamus is the source of the TRH that regulates TSH secretion. The 5'-flanking region of the gene encoding TRH has sequences for mediating responses to glucocorticoids and cAMP. In addition, at least two elements in this region of the gene can confer negative regulation of thyroid hormone receptor complexes.

TRH travels in the axons of the peptidergic neurons through the median eminence and is released close to the hypothalamic-pituitary portal plexus. The neuron bodies that produce TRH are innervated by catecholamine, neuroepitopeptide Y, and somatostatin-containing axons, all of which potentially influence the rate of synthesis of the pre-pro-TRH molecule.

A complex series of interactions with neuropeptides involving leptin, agouti-related peptide (AgRP), and melanocyte-stimulating hormone (-MSH) regulates pre-pro-TRH synthesis in the rat. The acute decrease in TRH synthesis that occurs in the starved rodent and leads to central hypothyroidism can be reversed by leptin infusion. Leptin directly increases pro-TRH biosynthesis in rat hypothalamic neurons, as does -MSH, whereas neuroepitopeptide Y and melanocortin-4 receptor (MC4R) suppress TRH synthesis as well. Thus, a fall in leptin, directly and indirectly, can reduce TRH synthesis.

Although leptin and -MSH can also activate the human TRH promoter through MC4R and the leptin receptor (ObRb), there is no acute reduction of TSH in fasted humans despite a decline in leptin levels. In humans, fasting causes a slight decrease in the amplitude of the pulsatile TSH release, which may be due to a decrease in TRH synthesis owing to reduced leptin. In addition, whereas a human leptin geneinactivating mutation causes morbid obesity and hypogonadism, it does not cause central hypothyroidism. Thus, current results suggest that the major role of leptin in regulating thyroid function in rodents in response to fasting is not shared by humans.

T3 suppresses the levels of pre-pro-TRH mRNA by T3 in the hypothalamus but normal feedback regulation of pre-pro-TRH mRNA synthesis by thyroid hormone requires a combination of T3 and T4 in the circulation, with the latter giving rise to T3 via T3 5'-deiodination in the CNS. In rats, there is a dense expression of D2 in the specialized ependymal cells (tanyocytes) in the inferior portion of the third ventricle. These cells have processes extending into the median eminence and arcuate nucleus, where active conversion of T3 to T3 releases T3 in the region of the hypothalamic-pituitary portal system. Thus, part of the negative feedback induced by T3 may be generated both indirectly at the level of the paraventricular nucleus by suppressing TRH and at the median eminence and arcuate nucleus at a point where neuropeptides and T3 enter the pituitary portal system.

TRH binds to a receptor in the thyrotroph membrane, and calcium and cyclic guanosine monophosphate (cGMP) are the second messengers for induction of the thyroid response. The calcium is derived from endoplasmic reticulum, owing to increases in inositol triphosphate (IP₃) secondary to G protein activation of PLC-C. Both phorbol ester and calcium ionophores can stimulate TSH gene transcription. In addition to inhibiting the synthesis of pre-pro-TRH mRNA, thyroid hormone also blocks the capacity of TRH to stimulate TSH release from the thyrotroph. The mechanism for this effect is unknown, but the stimulating effects of both phorbol esters and calcium ionophores are blocked by prior incubation with thyroid hormone.

Exogenous TRH elicits the secretion of prolactin at threshold doses that are the same as those for stimulation of TSH secretion. As with TSH, the prolactin response to TRH is modified by the prevailing levels of thyroid hormones, although not to as marked an extent. The role of TRH as a physiologic modulator of prolactin secretion is uncertain, however. For example, nursing increases the serum prolactin concentration, but the serum TSH concentration is unchanged.

TSH is the major regulator of the morphologic and functional states of the thyroid gland. It is a glycoprotein secreted by thyrotrophs in the anteromedial portion of the adenohypophysis composed of an subunit of 14 kd (92 amino acids) that is common to LH, follicle-stimulating hormone (FSH), and hCG as well as a specific subunit, a 112-amino acid protein synthesized in thyrotrophs. The peptide sequence cysteine-alanine-glycine-tyrosine-cysteine (CACYC) is highly conserved in the subunits of TSH, FSH, LH, and hCG and is required for heterodimerization with the subunit. An an

In normal thyrotrophs and in thyrotroph tumors, synthesis of the subunit is in excess, indicating that the quantity of the subunit is rate-limiting for TSH secretion. TRH
increases and thyroid hormone suppresses the transcription of both subunits; these are the most important influences on TSH synthesis.

The physiologic glycosylation of TSH involves addition of preformed asparagine-linked oligosaccharides in the rough endoplasmic reticulum, modifications in proximal and distal Golgi apparatus, and the appearance of the intact, folded hormone in the secretory granules. The glycosylation of the subunits protects them from intracellular degradation and permits normal folding of the protein chains so that internal disulfide linkages are correctly formed. Glycosylation is also required for full biologic activity, and sialylation protects circulating TSH from interaction with hepatic galactose receptors, thus increasing its half-life. The biologic activity of TSH in the serum of patients with pituitary tumors or hypothalamic disorders is inappropriately low compared with immunologic activity, suggesting the formation of an abnormal product. Long-term administration of TRH can enhance the biologic activity of TSH in patients with hypothalamic hypothyroidism and may lead to increased thyroid hormone levels, suggesting that this is due to TRH deficiency. Thus, in humans, TRH regulates not only TSH subunit synthesis but also post-translational processing.

Levels of subunit in serum range from 0.5 to 5 µg/L but are elevated in postmenopausal women. In normal serum, TSH is present at concentrations between 0.5 and 5 mUL. The level is increased in hypothyroidism and reduced in hyperthyroidism (see later). The plasma TSH half-life is about 30 minutes, and production rates in humans are 40 to 150 mU/day.

Circulating TSH displays two types of variations. Pulsatile variations are characterized by fluctuations at 1- to 2-hour intervals. The magnitude of TSH pulsations is decreased during fasting, illness, or after surgery. Circadian variations are characterized by a nocturnal surge that precedes the onset of sleep and appears to be independent of the cortisol rhythm and fluctuations in serum and in T4 and T3 concentrations. When the onset of sleep is delayed, the nocturnal TSH surge is enhanced and prolonged, and the early onset of sleep results in a surge of lesser magnitude and shorter duration.

The degree of thyroid hypofunction after destruction of the hypothalamus is less severe than that following hypophysectomy, and residual thyroid function in the former circumstance can be altered by raising or lowering the concentration of thyroid hormones in the blood. Thus, thyroid hormones mediate the feedback regulation of TSH secretion, and TRH determines its set-point. The relationship between circulating T4 and TSH and pituitary TSH release is illustrated in Fig. 10-10. The acute inhibition of TSH release by in vivo administration of physiologic quantities of T3 is mediated by the T3 produced by D2 in the pituitary gland (and perhaps the hypothalamus), since it is blocked by a general delodinase inhibitor, but not by PTU, the specific D1 inhibitor. A decrease in either plasma T3 or T4 causes an increase in TSH secretion because both T3 directly, and T4 via intrapituitary and intracerebral T3 to T4 conversion, contribute to T3 in the hypothalamus and pituitary gland (see Fig. 10-10). It follows that exogenous T3 is an effective suppressor of TSH secretion because (1) it is converted to plasma T3 and (2) it serves as the prohormone for T3 in the CNS and the pituitary gland. There is a linear relationship between the serum T3 concentration and the log of the TSH (Fig. 10-11). Thus, the serum T3 concentration is an exquisitely sensitive indicator of the thyroid state of most patients.

Somatostatin (SRIH, or somatotropin release-inhibiting hormone), acting through inhibitory G protein (Gi), decreases TSH secretion in vitro and in vivo, but prolonged treatment with a somatostatin analogue does not cause hypothyroidism. Similar acute effects occur during dopamine infusion and administration of bromocriptine, a dopamine agonist. Both of these agents inhibit adenylyl cyclase. Conversely, blockade of the dopamine receptor by metoclopramide increases the basal serum TSH concentration in both euthyroid and hypothyroid patients. These findings suggest that dopamine is a regulator of TSH secretion, indicating that compensatory mechanisms negate these acute effects. The neuroendocrine regulation of TSH secretion is detailed in Chapter 7.

### Table 10-7 -- Endogenous and Exogenous Agents That May Suppress Thyrotropin Secretion

<table>
<thead>
<tr>
<th>Thyroid hormones and analogues</th>
<th>Dopamine and dopamine agonists</th>
<th>Somatostatin and somatostatin analogues</th>
<th>Dobutamine</th>
<th>Glucocorticoids (acute, high-dose)</th>
<th>Interleukin-1, interleukin-6</th>
<th>Tumor necrosis factor-</th>
<th>Bexarotene (retinoid X receptor agonist)</th>
<th>Phenyltoin</th>
</tr>
</thead>
</table>

A number of drugs or hormones can suppress TSH secretion (Table 10-7). Glucocorticoids given in high doses acutely suppress TSH secretion transiently, although prolonged therapy is not associated with central hypothyroidism. Patients with Cushing's disease have subnormal TSH production but with minimal effects on T4 production. Bexarotene, an RXR agonist used for treatment of T cell lymphoma, suppresses TSH sufficiently to cause central hypothyroidism, presumably by reducing TSH gene transcription.
Iodine Deficiency

The response of vertebrates to a deficiency of iodine is designed to conserve this limited resource and to improve the efficiency of its utilization. These adjustments occur at the hypothalamic, pituitary, thyroid, and peripheral tissue levels. Removal of iodine from the diet causes a rapid decrease in serum T\textsubscript{4} concentrations and a simultaneous increase in serum TSH. Interestingly, no detectable decrease in T\textsubscript{3} occurs, suggesting that the signal to increase TSH must derive from a decrease in the T\textsubscript{3} generated intracellularly from T\textsubscript{4} in the pituitary gland, the hypothalamus, or both. TSH increases NIS, Tg, and TPO synthesis and iodine organization and Tg turnover. Because of the decrease in iodide supply and in the DIT/MIT ratio, the T\textsubscript{4}/T\textsubscript{3} ratio in Tg decreases and the rate of thyroidal T\textsubscript{3} secretion is probably increased despite a decline in T\textsubscript{4} secretion.

TSH also stimulates cell division, leading to goiter. In the rat model, the fall in plasma T\textsubscript{4} increases D2 from fivefold to 20-fold in the CNS, hypothalamus, and pituitary gland, increasing the efficiency of T\textsubscript{4} conversion to T\textsubscript{3}. With moderately severe iodine deficiency, D3 in the CNS is also reduced, prolonging the mean residence time of T\textsubscript{3} in that organ. This permits serum T\textsubscript{3} to remain normal and the CNS T\textsubscript{3} to be only moderately reduced even with up to a 10-fold decrease in circulating T\textsubscript{4}.

Despite the TSH elevation and nearly undetectable serum T\textsubscript{4}, growth, oxygen (O\textsubscript{2}) consumption and thermal homeostasis can be maintained. In humans, these compensatory alterations in thyroid function come into operation when total iodine intake falls below 75 µg/day. This situation obtains in many countries in Europe and South America as well as for several hundred million individuals in areas of iodine deficiency in China, India, Indonesia, and Africa.

These changes in serum hormones have been well documented in humans in areas of iodine deficiency and in patients with NIS mutations. The physiologic response to iodine deficiency is similar to that occurring during the development of primary hypothyroidism in humans. It is also reproduced when the efficiency of iodide trapping and organification is reduced in patients with Hashimoto’s disease or in patients with Graves’ disease receiving thiourea drugs. The physiologic rationale for this series of events is clear. T\textsubscript{3} has approximately three times the potency of the prohormone T\textsubscript{4} and contains only three iodine atoms. In terms of metabolic potency, this results in a fourfold more efficient use of iodine. Maintenance of normal circulating T\textsubscript{3} levels provides hormone for tissues in which nuclear T\textsubscript{3} is completely derived from the plasma.
Besides being protected against iodine deficiency, the thyroid gland is protected against an excess of iodine that might otherwise lead to hyperthyroidism. As with the response to iodine deficiency, there are multiple levels of defense against this eventuality. The usual source of excess iodine is pharmaceutical, with angiographic dyes, amiodarone, and povidone-iodine the most common sources (Table 10-8).

Effects of Increased Iodine Intake on Thyroid Hormone Synthesis

The quantity of iodine undergoing organification displays a biphasic response to increasing doses of iodide, at first increasing and then decreasing as a result of a relative blockade of organic binding. This decreasing yield of organic iodine from increasing doses of iodide (the Wolff-Chaikoff effect) results from a high concentration of inorganic iodide within the thyroid cell. The susceptibility to the Wolff-Chaikoff effect can be increased by (1) stimulation of the iodide-trapping mechanism (as in patients with Graves' disease) or during persistent TSH stimulation or (2) in patients with impairment of iodine organification, as may occur during thiourea drug treatment, in Hashimoto's disease, or in thyroid glands previously irradiated by either $^{131}$I or external beam therapy (e.g., for Hodgkin's disease).

In such situations, goiter and hypothyroidism can develop if excess iodide is given for long periods. The mechanism for inhibition of organification may involve the effects of high iodide concentrations on TPO-catalyzed organifications as well as a consequence of the formation of one or more inhibitory iodolipids in the thyroid cells.

Consequently, thyroidal iodide falls to levels insufficient to maintain the full Wolff-Chaikoff effect. This adaptation prevents the development of hypothyroidism, iodide goiter, or myxedema. Of note, it does not occur in the third trimester fetus; thus, chronic high iodine intake during pregnancy must be avoided because it causes fetal hypothyroidism and compensatory, obstructive goiter (Fig. 10-13).

Effects on Thyroid Hormone Release

An important practical effect of pharmacologic doses of iodine is the prompt inhibition of thyroid hormone release. Iodine decreases not only the fractional turnover of thyroidal iodide but also the $T_4$ secretion rate. This effect is the mechanism by which iodine rapidly lowers the serum $T_4$ concentration in patients with diffuse toxic goiter or toxic nodules (see Chapter 11). The mechanism by which iodide inhibits secretion of thyroid hormones is unknown, but the effect is mediated at the thyroid cell level rather than through an action on TSH. Iodine also diminishes the hypervascularity and hyperplasia that characterize the diffuse toxic goiter of Graves' disease. This effect facilitates surgical therapy for the disorder.
Thyroid Function in Pregnancy and in the Fetus and Newborn

Pregnancy affects virtually all aspects of thyroid hormone economy (Table 10-9). Total serum T₄ and T₃ concentrations rise to levels twice those of nonpregnant women owing to the increase in TBG concentration in the first trimester (Fig. 10-14). Free T₄ and T₃ levels also increase slightly during the first trimester but return to normal by about 20 weeks of gestation and remain so until delivery. This increase is due to hCG, which is a weak agonist for the TSH receptor. The slight decrease in serum TSH during the first trimester indicates that the free T₄ and T₃ changes are not dependent on the hypothalamic-pituitary axis. This decrease in serum TSH is all the more surprising, since it coincides with a number of events that act to increase maternal requirements for thyroid hormone.

In addition to an increase in serum TBG, there is also an increased plasma volume as well as accelerated inactivation of T₄ and T₃ by D3 expression in placenta and, perhaps, in the uterus as well. On the basis of changes in requirements for T₃ during gestation in women with primary hypothyroidism, the estimated increase in T₃ production required during this period is approximately 50%.

During pregnancy, the requirement for increased T₄ production increases iodine requirements. This need is compounded by the fact that the increased glomerular filtration rate during pregnancy enhances renal iodide clearance, leading to higher fractional urinary excretion of circulating iodide. In addition, maternal iodine intake must be increased to supply the iodide requirements of the fetal thyroid gland during the second and third trimesters (see Table 10-9). If these increased requirements for iodide are not met, serum T₄ levels fall, TSH levels rise, and goiter ensues. This series of events is well documented in areas of endemic iodine deficiency or a borderline iodine supply, such as in Brussels.

In that city, 70% of pregnant women carefully monitored throughout their pregnancy had a 20% or greater increase in thyroid volume during gestation. This contrasts with the lack of goiter during pregnancy in studies in North America. After delivery, the changes in thyroid hormone gradually return to normal, with serum TBG values reaching their normal levels 6 to 8 weeks post partum.

Pregnancy exerts a number of effects on the immune system. These changes may have striking effects on the natural history of patients with autoimmune thyroid disease, including both Graves' disease and Hashimoto's thyroiditis. In general, thyroid stimulation in women with Graves' disease is exacerbated during the first trimester but then is reduced gradually during the second and third, only to exacerbate again in the first several months post partum. Other than the increase in T₄ requirements in hypothyroid pregnant women, there is no evidence of a clinically significant change in patients with Hashimoto's disease during pregnancy, except that thyroid antibody levels fall. On the other hand, the marked rebound in the immune system occurring in the postpartum period leads to a phase of acute thyroid cell destruction, postpartum thyroid disease (PPTD), in about 30% of these patients (see Chapter 12). The basal metabolic rate (BMR) increases during the second trimester owing to the increase in the total mass of body tissue consequent to the pregnancy. The changes of pregnancy, together with the decreased peripheral vascular resistance, vasodilatation, and modest tachycardia, may suggest thyrotoxicosis (see Table 10-9). It is important that physicians remember that such changes are normal in pregnancy, especially when treating the hyperthyroid pregnant patient.

Fetal Thyroid Function

Fetal thyroid function begins at about the end of the first trimester. Thereafter, there are steady increases in fetal TBG and in total T₄ and T₃. Throughout gestation, serum TSH values are greater than are present in maternal circulation and higher than would be expected in adults with normal thyroid function. This indicates that there is increasing hypothalamic-pituitary resistance to T₄ during fetal development; it is speculated that this may be a consequence of increased TRH secretion.

Circulating T₄ levels remain relatively low, in contrast to the fetal free T₄ concentrations that approximate those in the maternal circulation from gestational age 28 weeks onward. This is explained primarily by the high D3 in fetal tissues, especially the liver.
Maternal-Fetal Interactions

The fetal pituitary-thyroid axis functions as a unit that is essentially independent of the mother. Transplacental passage of TSH from mother to fetus is negligible, but the same is not true of maternal T₄. In infants with congenital hypothyroidism caused by either TPO deficiency or athyreosis, serum concentrations of T₄ in umbilical cord blood are usually one third to one half of normal.

Thus, at least when the maternal-fetal concentration gradient is high, significant transfer of T₄ to the fetal circulation occurs. This transfer may be significant, given the capacity of the fetal brain to increase the efficiency of T₄ to T₃ conversion. Furthermore, T₄ can be found in coelomic and amniotic fluids prior to the onset of thyroid function.

The major factor limiting T₄ and T₃ transport from mother to fetus is the D3 expressed in the placenta. Blocking D3 activity in a perfused human placental lobule markedly increases the quantity of T₄ crossing into the equivalent of the fetal circulation. Placental D3 activity is present throughout gestation, although its activity, expressed per milligram of placental protein, decreases. Nonetheless, the increase in placental size results in progressive increases in total D3 content throughout gestation. This can account, at least in part, for the sudden decrease in maternal -thyroxine requirements that occurs immediately upon delivery.

Thyroid Function in the Newborn

Mean total T₄ levels in cord sera are 150 nmol/L (12 µg/dL). Serum TBG concentrations are elevated, but not as high as in the maternal serum. At term, free T₄ concentrations are slightly lower than those in the mother. Cord serum T₃ concentrations are low (0.8 × 10⁻⁹ M, 50 ng/dL), and rT₃ and T₃, S are elevated. After delivery, the neonate’s serum TSH level increases rapidly to a peak at 30 minutes of extrauterine life, returning to its initial value within 48 hours. This neonatal TSH surge is thought to occur in response to marked reduction in environmental temperature after delivery. Serum T₄, T₃, and Tg concentrations increase rapidly during the first few hours after delivery and are in the hyperthyroid range by 24 hours of life. The TSH surge doubtless contributes to the increase in serum T₃ concentration, but enhancement of extrathyroidal conversion of T₄ to T₃ by D1 or D2 is thought to be a major factor as well. The adrenergic stimulation of D2 in brown adipose tissue may also contribute to the increase in serum T₃.

Serum T₃ concentrations increase during the first 24 hours of postnatal life as a result of the increased T₄ but decrease to normal values by the fifth postnatal day. By the 10th day or so, the serum T₄ and T₃ concentrations are lower but still exceed normal adult values. Serum T₄ levels are slightly higher in the first year of life but gradually fall to the normal adult range.

Premature infants have an immature hypothalamic-pituitary-thyroid axis with low levels of T₄, T₃, and TSH. Serum T₄, TBG, and free T₄ all tend to correlate with gestational age. Preterm infants do have a TSH surge after delivery, but this is attenuated relative to that of full-term infants. In addition, when prematurity is accompanied by complications, such as respiratory distress syndrome or nutritional problems, serum T₄, and especially T₃, may fall to low levels because of a combination of reduced TBG production, immaturity of the thyroid gland, and suppression of the hypothalamic-pituitary axis owing to illness. These changes are, in many respects, similar to those in adults who are severely ill. The physician must take all of these issues into account when evaluating the thyroid status of the preterm infant, particularly given the increased prevalence of congenital hypothyroidism in this age group.

Thyroid hormone production rates are higher per unit of body weight in neonatal infants and children than in adults. The daily -thyroxine requirement is about 10 µg/kg in the newborn, decreasing to about 1.6 µg/kg in the adult.
Aging and the Thyroid Gland

Reproducible thyroid function abnormalities are observed in aging patients. In the healthy patient, there is a normal free T₄ but a relatively lower serum TSH than in younger individuals. In addition, although some disagree, it appears that serum T₃ levels decline, especially in individuals older than age 100 years. The serum T₃/T₄ ratio also tends to be reduced in individuals in their eighth and ninth decades; in addition, the daily secretion rate of TSH is reduced.

Although these changes resemble those occurring in patients who are ill, serum rT₃ concentrations are not nearly as elevated as they are in hospitalized individuals, and whether the reductions in serum T₃ are pathologic or physiologic is still unclear. A decreased requirement of about 20% for thyroid hormone replacement in the hypothyroid elderly becomes apparent in the eighth decade.
Thyroid Function during Fasting or Illness

A number of changes may take place in thyroid function during nutritional deprivation or illness. The changes that are induced in these two conditions are similar and, therefore, are best discussed in concert. Many of the changes involve alterations in thyroid hormone metabolism and are alluded to in that discussion.

Effects of Nutritional Deprivation

Both short-term and long-term alterations in nutritional state affect various aspects of thyroid hormone economy, especially peripheral hormone metabolism, as previously discussed. When euthyroid lean or obese subjects are starved, the serum total and free T₃ levels decrease to subnormal levels. In general, the serum total and free T₃ concentration remains essentially unchanged, or total T₄ may decrease slightly because of a modest decrease in iodothyronine-binding proteins. As serum T₃ concentrations decrease, concentrations of rT₃ increase reciprocally, usually to values about twice normal, not because of a major increase in the production of rT₃, but because of a decrease in its clearance. The abnormal T₃ and rT₃ concentrations in serum are quickly restored to normal by administration of small quantities (200 kcal) of carbohydrate. Similar quantities of protein have no effect on the serum T₃ level but may lower the serum rT₃ level. Calories given as fat are ineffective.

Despite the decrease in free T₃ concentration with starvation, the basal serum TSH and the free T₄ concentration and its response to TRH infusion are essentially unchanged. This occurs despite a decrease in leptin, suggesting that this hormone is less crucial in maintaining TRH production in humans than in rats despite its positive stimulation of the human TRH promoter.

Basal oxygen consumption and heart rate decline, nitrogen balance returns toward normal, and peripheral steroid metabolism shifts toward the pattern seen in hypothyroidism. In some, but not all studies, these changes are partially reversed by administration of exogenous T₃ while fasting continues. The decrease in T₃ during fasting is viewed by many as a beneficial energy-sparing and nitrogen-sparing adaptation.

Chronic malnutrition, as in protein-calorie malnutrition and anorexia nervosa, is also associated with a decreased serum T₂ concentration. Serum T₆ levels also tend to be slightly decreased, but serum TSH concentrations and their response to exogenous TRH are usually normal. In contrast, overfeeding, particularly with carbohydrate, increases T₃ production rate, increases the serum T₃ level, lowers the serum rT₃ concentration, and increases basal thermogenesis.

Effects of Illness

The changes in circulating thyroid hormones in illness resemble those during fasting, except that they may be much more severe. In addition, in the severely ill patient, there is often a suppression of the pituitary hormone release, which is either endogenous, because of loss of hypothalamic input, or worsened by some agents, such as dopamine and glucocorticoids, often given to very ill patients. This condition has come to be called the euthyroid sick syndrome, nonthyroid illness, or the low T₃ syndrome. These terms refer to the global pattern of changes in thyroid physiology that occur during illness.

The changes in thyroid function are a continuum, with the abnormalities becoming progressively more severe in accordance with the patient’s clinical condition. The disruption in thyroid function in sick patients can be arbitrarily divided into three stages (Table 10-10):

<table>
<thead>
<tr>
<th>Severity of Illness</th>
<th>Thyroid-Related Hormone</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Free T₃</td>
</tr>
<tr>
<td>Stage 1 (mild)</td>
<td>Normal</td>
</tr>
<tr>
<td>Stage 2 (moderate)</td>
<td>Increased</td>
</tr>
<tr>
<td>Stage 3 (severe)</td>
<td>Reduced</td>
</tr>
</tbody>
</table>

In addition to the central changes in TSH regulation and the abnormalities in peripheral hormone metabolism, patients may also have abnormalities in thyroid function that may be attributed to changes in the circulating binding proteins and, especially in the case of sepsis, those caused by the marked reduction the TBG affinity due to the serpin cleavage of the carboxy-terminal fragment of TBG. In addition to these endogenous changes, the TSH-suppressive effects of therapeutic agents (e.g., dopamine, dobutamine, glucocorticoids) on the central thyroid axis may complicate and exacerbate the abnormality. Illness and surgery decrease the nocturnal pulsatile TSH surge, presumably by reducing TRH release.

After the initial recognition of this syndrome, a large number of studies were performed using the starved rat as a model. As we have come to recognize, thyroid physiology in the rat, especially during fasting, differs from that in humans. Therefore, many of the studies pointing to an effect of illness or starvation to impair D₁ activity, either through reductions in D₁ or a D₁ cofactor, may have limited applicability in humans. In addition, the role of leptin appears to be much more important in maintaining TRH synthesis in rodents than in humans. Patients with severe illness have high leptin levels despite suppression of the central hypothalamic-pituitary-thyroid axis. In humans, rT₃ production remains normal as long as TSH secretion is maintained. Therefore, the elevation in rT₃ indicates an impairment of its clearance and rT₃ uptake into D₁-expressing tissues is impaired during illness or fasting. This decreased transport is due either to ATP depletion or, perhaps, to substances that compete with rT₃ for cellular entry.

Because the serum T₃ can fall to undetectable levels during illness, it seems likely that all three pathways for T₃ production, TSH secretion and T₄ outer ring deiodination by D₁ and D₂, may be reduced. Again, it is not certain whether abnormalities, particularly in D₂, are a result of (1) decreased enzyme levels (e.g., as in skeletal muscle induced by TNF-α) (2) decreased uptake into the slowly equilibrating D₂-containing tissue pool, or (3) a reduction in D₂ due to accelerated proteolysis of this enzyme via the ubiquitin-proteasome pathway. T₃ uptake into rapidly equilibrating pools, such as those in the liver and kidney, is also reduced during fasting or...
in illness.

Although serum TSH concentrations in severely ill patients are reduced, an increase in TSH above the normal range may appear during recovery, with the elevation in TSH concentration persisting until circulating free T4 and T3 levels return to normal. This pattern can be confusing if the elevated TSH concentration is associated with the still-reduced concentrations of free T4. Such patients meet all laboratory criteria for primary hypothyroidism with the exception of the clinical context. Follow-up generally reveals a normalization of TSH and T4 within 1 to 2 months.

Despite the severity of the abnormalities, particularly in serum T3, it is still debatable whether therapeutic intervention should be initiated even in the most severely ill patients. This is because most controlled studies have not shown beneficial effects of T4 or T3 supplementation in such individuals. The one exception is the possible beneficial effect of T3 therapy in patients after coronary artery bypass grafting; one study showed a positive effect but a second showed no beneficial effect.

A promising alternative strategy has been initiated in a series of studies of patients with prolonged critical illness. In an attempt to correct the inappropriate catabolism and failure of fat mobilization of prolonged critical illness, infusions of GH-releasing peptide-2 (GHRP-2) and TRH have been given for 5-day periods in a randomized fashion. Increases in TSH, T3, and T4 (as well as in IGF-I) insulin and IGF-binding proteins 1, 3, and 5 and leptin occurred and were associated with positive effects on osteocalcin, and decreases in the urinary/urea creatinine ratio.

These results suggested that restoration of somatotroph and thyrotroph function simultaneously by replacement of deficient neuroendocrine peptides had significant beneficial effects on the disordered metabolism of severe chronic illness that cannot be achieved by isolated and supraphysiologic replacement of either thyroid hormone or growth hormone alone. Whether these encouraging results will be associated with an improved outcome remains to be demonstrated, but the results suggest that in prolonged critical illness the central hypothyroidism is not beneficial.

The Thyroid Axis and Neuropsychiatric Illness

Patients with neuropsychiatric disease can present with any of a number of abnormalities in thyroid function. Patients with bipolar disorders may show slight elevations in serum TSH and reductions in free T4, whereas patients with severe depression have slightly elevated serum T4 and reduced serum TSH levels. Other acutely psychotic patients may have either high or low serum TSH concentrations and tend to have elevated free T4 levels.

The etiologic mechanism of these minor abnormalities is not clear, but the thyroid function test results of such patients may resemble the results of those with primary thyroid disease and must be differentiated from these.
Hormonal Effects on Thyroid Function

Glucocorticoids

The acute administration of pharmacologic doses of glucocorticoid eliminates pulsatile release of serum TSH concentrations in normal patients, presumably by reducing TRH release. With continued administration, there is an escape from this suppression (Table 10-11). Pharmacologic doses of glucocorticoid decrease serum T₃ concentration in normal and hyperthyroid patients as well as in hypothyroid patients maintained on L-thyroxine. The latter finding and the accompanying increase in rT₃ production suggest that glucocorticoids increase D₃ activity. The decreases in TBG and TTR have only modest effects on total T₄ concentrations.

Primary adrenal insufficiency may be associated with reduced serum T₄ and elevated serum TSH concentrations, suggesting the coexistence of primary hypothyroidism. However, treatment of the adrenal insufficiency can lead to complete resolution of the abnormalities in thyroid function, suggesting that in some patients they are a consequence of glucocorticoid deficiency rather than primary thyroid disease. Nevertheless, the prevalence of primary hypothyroidism is increased in patients with autoimmune hypoadrenalism, so the two causes must be differentiated (see Chapter 37). Similarly, thyroid autoimmunity can develop in patients successfully treated for Cushing's disease.

Gonadal Steroids

Estrogen increases TBG sialylation and half-life in serum. Estrogen administration to postmenopausal women causes an increase of 15% to 20% in TSH. Presumably, this increases T₄ secretion, since total T₄ increases and free T₄ is unchanged. It is not certain whether this is a transient or persisting phenomenon. Estrogen also increases the L-thyroxine requirement in patients with primary hypothyroidism. Again, the cause of this change is unclear.

Growth Hormone

Growth hormone increases the serum free T₃ and decreases free T₄ in both L-thyroxine-treated and normal individuals, suggesting either suppression of D₃ activity or increased T₄ to T₃ conversion. This change would reduce requirements for L-thyroxine in patients receiving exogenous hormone.

### Table 10-11 -- Effects of Hormones on Thyroid Function

<table>
<thead>
<tr>
<th>Hormone</th>
<th>Effect on Thyroid Function</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Glucocorticoids</strong></td>
<td></td>
</tr>
<tr>
<td>Excess</td>
<td>Decrease TSH, TBG, TTR (high-dose)</td>
</tr>
<tr>
<td></td>
<td>Decrease serum T₃/T₄ and increase rT₃/T₄ ratios</td>
</tr>
<tr>
<td></td>
<td>Increase rT₃ production (D3)</td>
</tr>
<tr>
<td></td>
<td>Decrease T₄ and T₃ secretion in Graves' disease</td>
</tr>
<tr>
<td>Deficiency</td>
<td>Increase TSH</td>
</tr>
<tr>
<td><strong>Estrogen</strong></td>
<td>Increase TBG sialylation and half-life in serum</td>
</tr>
<tr>
<td></td>
<td>Increase TSH in postmenopausal women</td>
</tr>
<tr>
<td></td>
<td>Increase T₄ requirement in hypothyroid patients</td>
</tr>
<tr>
<td><strong>Androgen</strong></td>
<td>Decrease TBG</td>
</tr>
<tr>
<td></td>
<td>Decrease T₃ turnover in women and reduce T₄ requirements in hypothyroid patients</td>
</tr>
<tr>
<td><strong>Growth Hormone</strong></td>
<td>Decrease D3 activity</td>
</tr>
<tr>
<td></td>
<td>D3, type 3 deiodinase; T₃, T₄, reverse T₃ and T₄ TBG, thyroxine-binding globulin; TSH, thyrotropin; TTR, transthyretin.</td>
</tr>
</tbody>
</table>
LABORATORY ASSESSMENT OF THYROID STATUS

In considering the laboratory assessment of the patient with known or suspected thyroid disease, the physician should seek to arrive at both functional and anatomic diagnoses. Laboratory determinations will confirm whether there is an excess, normal, or an insufficient supply of thyroid hormone to verify the inferences from the clinical history and physical examination. The second is to ascertain the presence or absence of anatomic abnormalities in the thyroid gland itself.

Laboratory evaluation can be divided into five major categories:

1. Tests that assess the state of the hypothalamic-pituitary-thyroid axis.
2. Estimates of the free $T_3$ or $T_4$ concentrations in the serum.
3. Tests that reflect the impact of thyroid hormone on tissues.
4. Tests for evidence of the presence of autoimmune thyroid disease.
5. Tests that provide information about thyroidal iodine metabolism.

Techniques for evaluation of anatomic abnormalities of the thyroid gland (thyroid ultrasound) and thyroid isotopic scanning are covered in Chapter 13.

Tests of the Hypothalamic-Pituitary-Thyroid Axis

Thyrotropin

While they are an inherently indirect reflection of thyroid hormone supply, tests that assess the state of the hypothalamic-pituitary-thyroid axis play a critical role in the diagnosis of thyroid disease. This is because the rate of TSH secretion is exquisitely sensitive to plasma concentrations of free thyroid hormones, thus providing a precise and specific barometer of the patient's thyroid status (see Fig. 10-11). Exceptions to this rule do occur (see later) but are rare. For example, the feedback of TSH secretion in premature infants and children is less sensitive to free $T_4$ than is that of the adult, probably due to higher TRH secretion rates.

Immunometric assay technology now makes it possible to define the normal range for serum TSH and, hence, to ascertain (1) when thyroid function is inadequate and (2) when the hormone supply is excessive. This assay uses the TSH molecule as a link between a TSH antibody bound to an inert surface (e.g., particles, the side of a test tube) and a second antibody directed against a different TSH epitope that is labeled with a detectable marker (e.g., an enzyme, or a chemiluminescent reagent) (see Chapter 6). Thus, the signal generated is proportional to the concentration of TSH in the serum. This technique is more specific, sensitive, and rapid than radioimmunoassay.

The normal range of serum TSH concentration varies slightly in different laboratories but is most commonly 0.5 to 5 mU/L or 0.3 to 4.0 mU/L, depending on the TSH reference preparation and assay used. Not all immunometric TSH assays are equally sensitive and specific. A useful functional categorization is in terms of the minimal detectable TSH that can be quantified with a less than 20% coefficient of variation.

The term generation has been employed to categorize each assay with respect to its sensitivity. Each successive generation offers about a 10-fold improvement in sensitivity (Fig. 10-15). The first generation (TSH radioimmunoassay) has lower limits of detectability of approximately 1 mU/L, whereas the third generation assay has minimal detectable limits of about 0.004 mU/L. A minimally suitable TSH assay should be able to quantitate concentrations of TSH of 0.1 mU/L with a coefficient of variation of less than 20%, thus falling into the second-generation or third-generation category.

An artifically elevated result may be obtained using the immunometric technique with serum containing heterophilic antimouse immunoglobulin G antibodies (HAMA). Such antibodies may substitute for TSH and cause falsely high values.

In patients with hyperthyroidism (excess thyroid hormone secretion) and/or thyrotoxicosis (excess thyroid hormone from any cause), the TSH level is virtually always suppressed. The values fall into two general categories: (1) when the lower limit of normal and 0.1 mU/L and (2) less than 0.1 mU/L. Individuals in the former category are usually really asymptomatic (subclinical hyperthyroidism), whereas those in the latter category almost invariably have symptomatic thyrotoxicosis and a significant elevation in free $T_4$.

Patients with hypothyamic or pituitary hypothyroidism typically have normal, not suppressed, serum TSH level, and sometimes the TSH is even slightly elevated. In the latter case, the circulating TSH generally has reduced biologic activity because of abnormal glycosylation reflecting the disordered hypothalamic-pituitary relationship. Patients with primary hypothyroidism have serum TSH concentrations that range from minimally elevated to 1000 mU/L.

In general, the degree of TSH elevation correlates with the clinical severity. Patients with serum TSH values in the range of 6 to 15 mU/L have few, if any, symptoms and the serum free $T_3$ or free $T_4$ index (FT$_4$) or FT$_3$ index may be low-normal or even slightly reduced. Serum free $T_4$ concentrations are typically normal. Such findings indicate early thyroidal decompensation with a compensatory increase in TSH secretion. A detailed discussion of the various conditions associated with abnormal serum TSH concentrations follows a description of the quantitation of serum thyroid hormones.
Thyrotropin-Releasing Hormone Stimulation Test

The availability of sensitive TSH assays has virtually eliminated the need for the TRH test except in rare circumstances. If a reliable immunometric TSH assay is not available, a TRH infusion test may be performed. The extent of increased serum TSH concentration after TRH administration correlates well with the basal serum TSH concentration. Exogenous TRH thus acts as an amplifier to exaggerate any underlying abnormality in TSH secretion.

A dose of 200 to 400 µg is administered as a single intravenous (IV) infusion over 5 minutes. In normal individuals, the serum TSH concentration rises rapidly, reaches a peak in 20 to 30 minutes, and returns to basal values in 2 or 3 hours. In practice, specimens for TSH analysis need be drawn only just before and 30 minutes after TRH administration. Normal increments range between 5 and 30 mIU/L and average about 12 mIU/L. Responses to TRH are decreased by pharmacologic glucocorticoid treatment, levodopa, dopamine infusions, somatostatin or dopamine analogues, and responses are augmented by the dopaminergic antagonists metoclopramide and domperidone.

The negative feedback inhibition of basal TSH secretion or TRH-induced TSH release is so exquisitely sensitive that doses of exogenous hormone insufficient to increase the serum T₄ or T₃ concentrations above the normal range nonetheless decrease the TSH response to TRH. Conversely, small decreases in serum T₄ concentrations cause increased basal serum TSH concentrations and an increased response to TRH. Thus, an elevation in both serum TSH and free T₄ indicates (1) a TSH-producing pituitary tumor, (2) RTH (resistance to thyroid hormone), or (3) hyperthyroidism with an artifactual increase in TSH. TRH does not increase the artifactually elevated TSH but usually increases that in patients with RTH. The TRH response of thyrotroph tumors is variable.

A differentiation between hypothalamic and pituitary causes of central hypothyroidism by TRH testing is not possible. Other diagnostic modalities such as magnetic resonance imaging (MRI) are required for a definitive evaluation of such patients (see Chapter 8).
Quantification of Serum Thyroid Hormone Concentrations

Total Thyroxine and Triiodothyronine Levels

Quantitation of the circulating thyroid hormone concentration is essential to confirm that the thyroid status abnormality suggested by an abnormal TSH result is accurate and to document its severity. Sensitive and specific radioimmunoassays are available for measuring the total concentrations of \( T_4 \) and \( T_3 \) and some of their metabolic by-products. Because thyroid status correlates with free rather than total hormone concentration, the physician must also estimate the free hormone concentration.

by measuring it directly, or by measuring it indirectly (as in a free hormone index), or by measuring the concentration of TBG, the major carrier protein for \( T_4 \) and \( T_3 \) in human plasma.

The degree of abnormality in the free \( T_4 \) level generally correlates with the severity of the hormone excess or deficiency, whereas the serum TSH concentration indicates the impact of this abnormality in a specific patient. Total concentrations are measured in whole serum by immunoassays. Errors peculiar to tests for thyroid hormones result from competition for the labeled antigen between the specific antibody and the plasma-binding proteins unless binding of the hormone to these proteins is inhibited by one of various agents. Endogenous antibodies to \( T_4 \) or \( T_3 \) are present in rare patients and invalidate the assay (see Chapter 6).

The normal range for total \( T_4 \) in healthy, euthyroid adults with a normal circulating TBG concentration is 64 to 142 nmol/L (5 to 11 µg/dL). Normal serum \( T_3 \) concentrations are 1.1 to 2.9 nmol/L (70 to 190 ng/dL). At birth (cord serum), \( T_3 \) concentrations are about 50% of those in normal adults, but within a few hours this number rises abruptly, peaking at about 24 hours at concentrations in the low thyrotoxic range for adults. The \( T_3 \) concentration gradually decreases during the first few weeks of life but remains about 25% higher than values in adults through early adolescence. Serum \( T_3 \) values may decrease with age even in healthy individuals.

Radioimmunoassays for \( rT_4 \), \( T_3 \), triac, tetrac, and the di-o-dothyronines are of primary interest in the research setting because these iodothyronines are derived from the circulating \( T_4 \) or \( T_3 \), both of which can be easily quantitated. An exception may be compound W, an as-yet-unidentified product of \( T_3 \) metabolism in the fetal circulation that appears in maternal sera. If this is indeed the case, measurements of compound W may serve as a much needed index of the state of fetal thyroid function to monitor the effects of maternal antithyroid drug therapy on fetal thyroid function.

Free Thyroxine and Free Triiodothyronine

The most accurate and direct measurements of the concentrations of free \( T_4 \) and free \( T_3 \) in serum are obtained by assay of these hormones in a dialysate or ultrafiltrate of serum. Alternatively, serum can be enriched with tracer amounts of the labeled hormone, and the concentration of the isotope in the dialysate or ultrafiltrate is expressed as a fraction of that in undiluted serum. The absolute concentration of free hormone is the product of the total hormone concentration and the fraction that is dialyzable. As mentioned, about 0.02% of \( T_4 \) and 0.3% of \( T_3 \) are free (see Table 10-5). The normal ranges for free \( T_4 \) are 9 to 30 pmol/L (0.7 to 2.5 ng/dL); for free \( T_3 \), 5 to 8 pmol/L (0.2 to 0.5 ng/dL).

Because \( T_3 \) is the major secretory product of the thyroid gland and correlates most closely with the serum TSH, in most situations a free \( T_3 \) estimate is all that is required to ascertain the state of thyroid secretion or supply. A bewildering array of methods exist to quantify free \( T_4 \) or \( T_3 \) in whole serum that involve automation. Although many automated tests are said to be able to measure free \( T_4 \) directly, they do not; it appears that results in sera with abnormal binding proteins are not accurate (see Chapter 6).

There are two general categories of methods: (1) comparative free \( T_4 \) methods and (2) so-called free \( T_3 \) index methods.

Three general approaches are used:

- Two-step labeled hormone methods
- One-step labeled analogue methods
- Labeled antibody approaches

In general, two-step labeled hormone back-titration methods, compared with one-step analogue methods, are less subject to artifacts caused by abnormal binding proteins, changes in albumin and TBG, or increases in free fatty acids. Both methods are now being replaced by those designed to quantify the unoccupied TBG-binding sites. Although these methods are usually satisfactory in ambulatory patients with uncomplicated thyroid disorders, all general approaches may be subject to artifacts from endogenous antibodies to \( T_3 \), abnormal binding proteins, or illness. Thus, the clinician must be wary if the \( T_4 \) result by any method does not agree with the clinical state and the TSH. In such cases, another method should be used to estimate the free \( T_4 \) (e.g., quantitation of \( T_4 \) in a dialysate or ultrafiltrate), the free \( T_3 \) index should be measured, or the result should be ignored.

Free Thyroxine Index

Another means of partially circumventing the technical difficulties inherent in direct estimates of free \( T_4 \) concentrations is to determine the thyroid hormone-binding ratio (THBR) and to multiply this result by the total \( T_4 \) or \( T_3 \) concentration to estimate the free thyroid hormone concentration, termed the FT4 I or the FT3 I. Because the normal THBR is 1.0, the FT4 I has a normal range in units identical to that of THBR.

Because the THBR is proportional to the free fraction of the endogenous thyroid hormones in serum, it can be multiplied by the total \( T_4 \) or \( T_3 \) concentration to estimate the free thyroid hormone concentration, termed the FT4 I or the FT3 I. Because the normal THBR is 1.0, the FT4 I has a normal range in units identical to that
of the total T₄, or T₃ (e.g., 64 to 142 for SI units and 5 to 11 in gravimetric terms). A schematic demonstration of the relationships between total and free T₄, occupied and unoccupied TBG-binding sites, and the THBR is shown in Figure 10-16 for euthyroid individuals with variations in TBG concentrations and in Figure 10-17 for subjects with a constant TBG and alterations in serum thyroid hormone production rates.

Estrogen, pregnancy, and severe illness, compared with hyperthyroidism and hypothyroidism, are more common causes of changes in total T₄ concentrations (Table 10-12). In the euthyroid person, only about one third of the available binding sites on TBG are occupied by T₄, and the free T₄ fraction is 2 times 10⁻⁴ of the total. During pregnancy, the TBG-binding capacity, serum T₄, and the number of unoccupied TBG binding sites almost double, leading to almost a 50% reduction of the free T₄ fraction. This is reflected in a reduced THBR. If

![Figure 10-16 Pattern of changes in total serum thyroxine (T₄) concentrations and thyroid hormone-binding ratio (THBR) in euthyroid patients with alterations in circulating concentrations of thyroxine-binding globulin (TBG). To convert T₄ from nmol/L, to µg/dL (total) or pmol/L (free), divide by 12.87.](image)

...Continued...

The THBR is linearly related to the free fraction of thyroid hormones except at the extremes of its range. To derive the maximum information, therefore, it is important to consider both the calculated FT₄ and the pattern of the deviations of total hormone and THBR from normal. When concentrations of binding proteins are altered, the deviation of the total T₄ measurements from normal is in the opposite direction to that of the THBR (see Fig. 10-16, central panels). In this circumstance, one should be suspicious that an alteration in TBG, rather than an alteration in thyroid hormone production, is responsible for the abnormality in the total thyroid hormone level. However, when an elevated T₄ level is due to increased T₄ secretion or exogenous sources, the concentration of unoccupied TBG-binding sites is reduced and both the free fraction and the THBR are increased. Therefore, the free T₄ and the FT₄ are increased in parallel and to a greater extent than would be suggested from the change in the total hormone level.

The changes in hypothyroidism are in the opposite direction but of lower magnitude. The concentration of unoccupied TBG-binding sites does not increase greatly when the serum T₄ level decreases from a normal mean of 100 nmol/L to 30 nmol/L, and the decrease in the free T₄ fraction and the THBR are not large. Thus, the reduced FT₄ is hypothyroidism is due predominantly to the decrease in T₄ rather than to a decrease in its free fraction. The central panels in Figure 10-17 show the parallel deviations from normal hormone levels and THBR when the thyroid hormone supply is altered.

Simultaneous abnormalities in both TBG and thyroid hormone production may also occur. Hyperthyroidism should be suspected during pregnancy when the T₄ concentration is very high and the THBR is not subnormal. Similarly, a serum T₄ concentration in the lower portion of normal range for a nonpregnant woman, accompanied by a significant reduction in the THBR, indicates hypothyroidism.

![Figure 10-17 Pattern of changes in total serum thyroxine (T₄) concentration and thyroid hormone-binding ratio (THBR) in patients with hyperthyroidism or hypothyroidism with normal serum thyroxine-binding globulin (TBG) concentration.](image)

One should keep several caveats in mind when interpreting these results. The use of labeled T₄ in these tests in some assays can produce difficulties in three situations: (1) in patients with FDH (familial dysalbuminemic hyperthyroxinemia), (2) in the presence of endogenous antibodies directed against T₄, and (3) in sick patients. In patients with FDH, the abnormal albumin binds T₄, but not T₃, with increased avidity. Therefore, these patients have an elevated level of total T₄ and a reduced free fraction of T₄ but not T₃. An alternative approach to assessing the FT₄ is thyroid hormone binding in serum is to measure TBG either by saturation analysis or by radioimmunoassay. Normal concentrations of TBG by radioimmunoassay are about 270 nmol/L (1.0 to 1.5 mg/dL) and are only slightly higher in women than in men. However, the elastase released from leukocytes during infection (see earlier) reduces the binding affinity of TBG for T₄ (and T₃) but does not change its immunoreactivity or its binding capacity. With this proviso, the serum TBG concentration result can be employed in one of two ways:

1. Normalization of the T₄/TBG (or T₃/TBG) ratio yields values that correlate reasonably well with the FT₄ or FT₃.

2. Derivation of an FT₄ from concentrations of TBG and total T₄ and the association constant for the interaction between the two.

<table>
<thead>
<tr>
<th>TABLE 10-12</th>
<th>Circumstances Associated with Altered Binding of Thyroxine by Thyroxine-Binding Globulin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increased Binding</td>
<td>Decreased Binding</td>
</tr>
<tr>
<td>Pregnancy</td>
<td>Androgens</td>
</tr>
<tr>
<td>Neonatal state</td>
<td>Large doses of glucocorticoids</td>
</tr>
<tr>
<td>Estrogens and hyperestrogenemic states</td>
<td>Active acromegaly</td>
</tr>
<tr>
<td>Tamoxifen</td>
<td>Nephrotic syndrome</td>
</tr>
<tr>
<td>Oral contraceptives</td>
<td>Major systemic illness</td>
</tr>
<tr>
<td>Acute intermittent porphyria</td>
<td>Genetic factors</td>
</tr>
<tr>
<td>Infectious and chronic active hepatitis</td>
<td>Asparaginase</td>
</tr>
<tr>
<td>Biliary cirrhosis</td>
<td>Genetic factors</td>
</tr>
<tr>
<td>Perphenazine</td>
<td></td>
</tr>
</tbody>
</table>
HIV infection
HIV, human immunodeficiency virus.

In most instances, values calculated in this manner correlate with the FT$_4$ determined by other techniques, although the T$_4$ /TBG ratio suggests a subnormal FT$_4$ in some euthyroid patients with an elevated TBG. [321]
Abnormal Thyrotropin or Thyroid Hormone Concentrations in Blood

There are many causes of abnormal TSH levels that should be considered by the clinician (Table 10-13). The clinical status and FT₄ levels allow assignment of the etiologic mechanism for these. Assay of the FT₃ level is rarely required but is included for completeness.

### Suppressed Thyrotropin Levels

The most common cause of a reduction in serum TSH is an excess supply of thyroid hormone due to either increased endogenous thyroid hormone production or excessive exogenous thyroid hormone. Because the concentration of TSH is inversely proportional to the degree of thyroid hormone excess, patients with clinical symptoms invariably have serum TSH concentrations below 0.1 mU/L. In such patients, the serum FT₄ is usually increased.

A low iodine intake may lead to clinical thyrotoxicosis, a suppressed TSH level, but only a high-normal FT₃. Therefore, in order to make a diagnosis of T₃ thyrotoxicosis, an FT₃ level is required. When the thyroid hormone supply is only slightly in excess of the requirement for that patient, serum

<table>
<thead>
<tr>
<th>Expected TSH (mU/L)</th>
<th>Clinical Thyroid Status</th>
<th>Free T₃ Index</th>
<th>Free T₄ Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thyrotropin Reduced</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Hyperthyroidism of any cause</td>
<td>&lt;0.1</td>
<td>N, (I)</td>
<td>N, (I)</td>
</tr>
<tr>
<td>2. &quot;Euthyroid&quot; Graves' disease</td>
<td>0.20.5</td>
<td>N, (I)</td>
<td>N</td>
</tr>
<tr>
<td>3. Autonomous nodule or multinodular goiter</td>
<td>0.20.5</td>
<td>N, (I)</td>
<td>N</td>
</tr>
<tr>
<td>4. Exogenous thyroid hormone excess</td>
<td>&lt;0.10.5</td>
<td>N, N</td>
<td></td>
</tr>
<tr>
<td>5. Thyroiditis (subacute or painless)</td>
<td>&lt;0.10.5</td>
<td>N, N</td>
<td></td>
</tr>
<tr>
<td>6. Recent thyrotoxicosis due to any cause</td>
<td>&lt;0.10.5</td>
<td>N, N</td>
<td></td>
</tr>
<tr>
<td>7. Illness with or without dopamine infusion</td>
<td>&lt;0.15.0</td>
<td>N, N</td>
<td></td>
</tr>
<tr>
<td>8. First trimester of pregnancy</td>
<td>0.20.5</td>
<td>N, (I)</td>
<td>N, (I)</td>
</tr>
<tr>
<td>9. Hyperemesis gravidarum</td>
<td>0.20.5</td>
<td>N, (I)</td>
<td>(N)</td>
</tr>
<tr>
<td>10. Hydatidiform mole</td>
<td>0.10.4</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>11. Acute psychosis or depression (rare)</td>
<td>0.410</td>
<td>N</td>
<td>N, (I)</td>
</tr>
<tr>
<td>12. Elderly (small fraction)</td>
<td>0.20.5</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>13. Glucocorticoids (acute, high dose)</td>
<td>0.10.5</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>14. Congenital TSH deficiency</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. PIT1 deficiency</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. CAGYC mutant</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Thyrotropin Elevated</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Primary hypothyroidism</td>
</tr>
<tr>
<td>2. Recovery from severe illness</td>
</tr>
<tr>
<td>3. Iodine deficiency</td>
</tr>
<tr>
<td>4. Thyroid hormone resistance</td>
</tr>
<tr>
<td>5. Thyrotoxic tumor</td>
</tr>
<tr>
<td>6. Hypothalamic-pituitary disease</td>
</tr>
<tr>
<td>7. Psychiatric illnesses</td>
</tr>
<tr>
<td>8. Adrenal insufficiency</td>
</tr>
<tr>
<td>9. Artifact (endogenous antimouse-globulin antibodies)</td>
</tr>
</tbody>
</table>

Arrows indicate the nature of the abnormality in the T₃ or T₄ index. Parentheses indicate that such a result is unusual but may occur.

TSH is suppressed but clinical manifestations are subtle or absent and the FT₄ and FT₃ are in the high-normal range. Such minimal changes can occur with euthyroid Graves' disease, autonomous thyroid hormone-producing adenomas, multinodular goiters, subacute or painless thyroiditis, and the ingestion of an amount of exogenous thyroid hormone slightly greater than that required for metabolic needs. In the absence of symptoms, this condition is termed subclinical hyperthyroidism.

The hypothalamic-pituitary-axial feedback loop is usually suppressed for up to 3 months after complete resolution of the thyrotoxic state. The best test for assessing the physiologic state in such patients is the FT₃ index. A common scenario with this pattern is during follow-up of patients receiving antithyroid drugs or for Graves' disease. With time, the TSH feedback regulatory loop normalizes and TSH secretion returns and becomes appropriate for the circulating free thyroid hormone concentration.

In stage 3 illness, with or without dopamine infusion or excess glucocorticoid, TSH is suppressed, causing assessment of thyroid functional status to be difficult. Because the FT₄ level may also be reduced in such patients, astute clinical judgment is required to assign thyroid status. Since HCG can stimulate the TSH receptor, conditions in which HCG is elevated (e.g., in the first trimester of pregnancy, during severe hyperemesis gravidarum, and in patients with hydatidiform mole or choriocarcinoma) are associated with mild to moderate hyperthyroidism and suppression of TSH. In euthyroid patients, TSH returns to normal in the second and third trimesters. A persistent and severely suppressed TSH level (<0.1 mU/L) in the pregnant patient during this time suggests that hyperthyroidism is due to other causes.

Patients with psychosis or depression may have abnormalities in the hypothalamic-pituitary-thyroid axis, although these findings are rarely associated with clinical symptoms. Typically, TSH is suppressed but in some patients it may be spontaneously elevated.
with glucocorticoids may transiently suppress TSH, chronic treatment does not. Rarely, patients with genetic deficiency of the PIT1 protein or with a mutation in the TSH-subunit have severe hypothyroidism and no detectable TSH (see Chapter 12).

Elevated Thyrotropin Levels

Elevations in TSH nearly always suggest a reduced supply of T₄ or T₃, which may be permanent or transient. Acutely ill patients may have elevated serum TSH levels, as in renal insufficiency, or may experience an asynchronous return of the hypothalamic-pituitary and thyroid axes to normal as they recover from acute illness; the latter case, patients have a transient form of primary hypothyroidism.

Iodine deficiency is the most common cause of elevated TSH worldwide, but this does not occur in North America.

Patients with resistance to thyroid hormone may be clinically hyperthyroid, euthyroid, or hypothyroid. The most common laboratory pattern is a serum TSH that is normal in absolute terms but inappropriately high for the elevated FT₄ I. Individuals with a more marked pituitary than general resistance to thyroid hormone (sometimes called pituitary RTH or PRTH) have symptoms suggesting hyperthyroidism, an elevated FT₄ I, and a normal or even elevated serum TSH. These patients must be differentiated from the equally rare patients with a thyrotroph tumor in whom the persistent secretion of TSH causes hyperthyroidism (see Chapter 8).

Patients with hypothalamic-pituitary dysfunction may have clinical and chemical hypothyroidism but low-normal or even elevated serum TSH concentrations. The explanation for this paradox is that the biologic effectiveness of the circulating TSH is impaired because of abnormal glycosylation secondary to reduced TRH stimulation of the thyrotrophs. Nonetheless, the abnormal TSH is a suitable antigen in the immunometric assay.

Some patients with psychiatric illness may have elevated TSH levels for reasons that are not understood. In adrenal insufficiency, TSH levels may be modestly elevated but return to normal with glucocorticoid replacement. In patients with antimouse IgG antibodies, TSH is usually artifactually elevated, often greatly so (see Chapter 6).

Despite the utility and general efficacy of the serum TSH measurement alone as a screening tool for identifying patients with thyroid dysfunction, a patient should not receive treatment for this dysfunction solely on the basis of an abnormal TSH level. The TSH assay is an indirect reflection of thyroid hormone supply and does not, by itself, permit a conclusive diagnosis of a specific disorder of thyroid hormone production.
Concordant and Divergent Abnormalities of Serum Thyroxine and Triiodothyronine Concentrations

One suspects that millions of dollars might be saved each year by the judicious selection of thyroid tests. In the authors' experience, far too many serum T₃ measurements are performed. Serum T₄ is rarely required for the accurate evaluation of the patient with an abnormal TSH level and is almost always an indirect reflection of the serum T₃ supply. Serum T₃ results are virtually useless in the hospitalized patient. Nonetheless, the endocrinologist is often asked to interpret abnormal or discordant serum T₃ and T₄ values (Table 10-14).

**TABLE 10-14 Causes of Concordant and Divergent Changes in Serum Thyroxine (T₄) and Triiodothyronine (T₃) Levels**

<table>
<thead>
<tr>
<th><strong>T₃ and T₄ Increased</strong></th>
<th><strong>T₃ and T₄ Decreased</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Increased TBG (see Table 10-12)</td>
<td>Decreased TBG</td>
</tr>
<tr>
<td>Thyrotoxicosis</td>
<td>Thyrotoxicosis due to ingestion of liothyronine (T₃)</td>
</tr>
<tr>
<td>Thyroid hormone resistance (RTH)</td>
<td>Euthyroid patients taking desiccated thyroid or liotrix (see text on hypothyroidism)</td>
</tr>
<tr>
<td><strong>T₃ Increased, T₄ Normal or Low</strong></td>
<td><strong>T₃ Decreased, T₄ Increased</strong></td>
</tr>
<tr>
<td>Familial dysalbuminemic hyperthyroxinemia (FDH)</td>
<td>Thyrotoxicosis due to ingestion of liothyronine (T₃)</td>
</tr>
<tr>
<td>Increased TTR or TTR binding</td>
<td>Euthyroid patients taking desiccated thyroid or liotrix (see text on hypothyroidism)</td>
</tr>
<tr>
<td>Amiodarone, high-dose propranolol, or oral cholecystographic agents</td>
<td>Phenytoin, carbamazepine</td>
</tr>
<tr>
<td>Illness, especially psychiatric</td>
<td>Iodine deficiency</td>
</tr>
<tr>
<td>Amphetamine abuse</td>
<td>Phenyltoin, carbamazepine</td>
</tr>
<tr>
<td>T₃ thyrotoxicosis, thyrotoxicosis with decreased T₃-to-T₄ conversion (see Table 10-6)</td>
<td>T₄, T₃ thyrotoxicosis</td>
</tr>
<tr>
<td><strong>T₃ Normal, T₄ Increased</strong></td>
<td><strong>T₃ Normal, T₄ Decreased</strong></td>
</tr>
<tr>
<td>T₃ thyrotoxicosis</td>
<td>Most patients with significant illness or during fasting (see Table 10-10)</td>
</tr>
<tr>
<td><strong>T₃ Decreased, T₄ Increased</strong></td>
<td><strong>T₃ Decreased, T₄ Decreased</strong></td>
</tr>
<tr>
<td>Thyrotoxicosis due to ingestion of liothyronine (T₃)</td>
<td>Mild or moderate thyroid failure</td>
</tr>
<tr>
<td>Euthyroid patients taking desiccated thyroid or liotrix (see text on hypothyroidism)</td>
<td>Iodine deficiency</td>
</tr>
<tr>
<td><strong>T₄, T₃ Decreased</strong></td>
<td>Phenyltoin, carbamazepine</td>
</tr>
<tr>
<td>Severe hypothyroidism</td>
<td>T₄, T₃ decrease</td>
</tr>
<tr>
<td>Severe systemic illness (euthyroid patient)</td>
<td>Decreased T3</td>
</tr>
<tr>
<td>Decreased T3</td>
<td>Salicylates in high doses (&gt;2.0 g/day)</td>
</tr>
<tr>
<td>TBG, thyroxine-binding globulin; TTR, transthyretin.</td>
<td>TBG, thyroxine-binding globulin; TTR, transthyretin.</td>
</tr>
</tbody>
</table>

An increase in serum TBG concentration secondary to increased estrogen is the most common cause of simultaneous elevations of serum total T₄ and T₃. Other causes of TBG elevation are listed in Table 10-12. Thyrotoxicosis due to hyperthyroidism is usually associated with an increased T₃/T₄ ratio, whereas that due to thyroiditis or exogenous T₃-thyroxine is associated with a decreased T₃/T₄ ratio. In patients with RTH, serum total and free T₄ and T₃ concentrations are elevated, although typical clinical features of thyrotoxicosis are often lacking, and both basal serum TSH concentrations and the response to TRH are normal or increased.

In all these conditions, serum T₃ concentrations are normal. Certain pharmacotherapeutic agents may elevate the serum T₃ concentration by inhibiting the conversion of T₄ to T₃ by D₁ and D₂ or by interfering with the cellular uptake of T₄. Agents that inhibit peripheral T₃ production include amiodarone and oral cholecystographic agents, such as iopanoic acid, sodium iopodate, tyropanoate, and iobenzamic acid (see Fig. 10-7). Propranolol in high doses, but not other β-adrenergic agents, can inhibit D₁ and, at doses greater than 160 ng/day, can lower serum T₃. The T₃ increase is a compensatory response to the blockade of T₃ production attributed to these agents, and the T₃ and TSH eventually normalize with a T₄ level that is elevated as long as thyroid function is normal.

Decreased T₄, increased TSH

Patients receiving replacement therapy with either liothyronine (T₃) or with agents with a higher T₃/T₄ ratio than in human thyroid secretion (11/1 by weight), including liotrix and desiccated thyroid, usually have a low-normal FT₄ when doses are titered to return serum concentrations of TSH to normal. T₃ levels are transiently decreased. Patients with T₃ levels as low as 2.5 ng/dl have been reported, and in these cases FT₄ levels may be >2 ng/dl. These patients require frequent follow-up (see Table 10-12).
elevated if serum is drawn between 1 and 4 hours after the patient ingests the medication.

Increased Thyroxine, Normal Triiodothyronine

The reasons for the decreased T₄ and normal T₃ in early hypothyroidism or during iodine deficiency have been discussed earlier (see "Iodine Deficiency"). Many drugs lower the serum T₄ concentration by interfering with the binding of T₄ to plasma proteins or by accelerating T₄ metabolism or both. Therapeutic doses of phenytoin lower the serum FT₄ I concentration, sometimes into the hypothyroid range. Although high concentrations of the drug can inhibit the binding of T₄ and T₃ to TBG in vitro, a drug-induced acceleration of T₄ disposal and, perhaps, central TSH suppression appear to be responsible for the reduced T₄.

Enhancement of hepatic disposal of T₄ by induction of cytochrome CYP3A4 in patients receiving carbamazepine, rifampin, phenobarbital, and phenytoin causes increased glucuronide conjugation of T₄ and T₃. This does not pose a problem in euthyroid patients. In hypothyroid patients, however, the levothyroxine dose must be increased.

Serum T₃ concentrations usually remain at low-normal levels.

Decreased Thyroxine, Decreased Triiodothyronine

Hypothyroidism of any origin and TBG deficiency are the most common causes of parallel reductions in serum T₄ and T₃ (see Table 10-12). In addition, salicylates, especially salsalate, inhibit the binding of T₄ and T₃ by serum proteins in vitro and have comparable effects in vivo when given in high doses. Initially, serum free T₄ and T₃ concentrations increase, but the consequent increased MCRs of the hormones, combined with the suppression of the hypothalamic-pituitary axis, lead to a new equilibrium in which serum total T₄ and T₃ values are decreased and free T₄, free T₃, and TSH values return to normal.

Marked lowering of the serum T₄ concentration and moderate decreases in the serum T₃ level may also occur in patients receiving the nonsteroidal anti-inflammatory drug (NSAID) fenclofenac. Patients are clinically euthyroid, and serum TSH levels are normal. Administration of L-asparaginase has the same rapid effect owing to inhibition of TBG synthesis. Stage 3 (severe) illness reduces both serum T₄ and T₃ (see Table 10-10).
Tests That Assess the Metabolic Impact of Thyroid Hormones

Abnormalities in the supply of thyroid hormone to the peripheral tissues are associated with alterations in a number of metabolic processes that can be quantitated. Some of these may be useful in the rare patient in whom serum TSH is not an accurate barometer of thyroid status, such as the patient with RTH. Such tests may be the sole means of evaluating the metabolic response of the peripheral tissues to thyroid hormones in such patients.

**Basal Metabolic Rate**

Thyroid hormones increase energy expenditure and heat production, as manifested by weight loss, increased caloric requirement, and heat intolerance. Because it is impractical to measure heat production directly, the basal metabolic rate measures oxygen consumption under specified conditions of fasting, rest, and tranquil surroundings. Under these conditions, the energy equivalent of 1 L of oxygen is equivalent to 4.83 kcal.

Under basal conditions, approximately 25% of oxygen consumption is due to energy expenditure in visceral organs, including the liver, kidneys, and heart; 10% in the brain; 10% in respiratory activity; and the remainder in skeletal muscle. Because energy expenditure is related to functioning tissue mass, oxygen consumption is related to some index thereof, most often body surface area. Calculated in this way, basal oxygen consumption (resting energy expenditure) is higher in men than in women and declines rapidly from infancy to the third decade and more slowly thereafter.

Values in patients, calculated as a percentage of established normal means for gender and age, normally range from -15% to +5%. In severely hypothyroid patients, values may be as low as -40%. In thyrotoxic patients, these values may reach +25% to 50%. Abnormal, usually elevated values, are seen in patients with burns and with systemic disorders (e.g., febrile illnesses, pheochromocytoma, myeloproliferative disorders, anxiety, and disorders associated with involuntary muscular activity). Interestingly, the changes in resting energy expenditure correlate well with the FT₄ and TSH in hypothyroid patients who are taking varying doses of levothyroxine.

**Biochemical Markers of Thyroid Status**

Occasionally, a diagnosis of thyroid dysfunction is first suspected as a result of an abnormality in a laboratory test performed in the course of an evaluation for an unrelated medical problem. Classical examples are a markedly elevated creatine kinase MM isoenzyme or LDL cholesterol level leading to the recognition of hypothyroidism. Other similar markers are listed in Table 10-15. These tests are not useful in the diagnosis of thyroid disease, but some, such as sex hormone-binding globulin (SHBG), ferritin, or LDL cholesterol, have been used as end points in clinical studies of the responsivity of the liver to thyroid hormone in patients with RTH.
Serum Thyroglobulin

The sensitivity of modern Tg assays is 1 ng/mL or even less. The results can be artifactually altered by serum anti-Tg antibodies, and serum should be screened for Tg antibodies with a sensitive Tg-antibody immunoassay. In immunoradiometric assays, interferences lead to underestimations or false-negative values. Tg is normally present in serum, with the concentration ranging up to 90 pmol/L (50 ng/mL); mean normal values vary with the assay used but are on the order of 30 pmol/L (20 ng/mL).

<table>
<thead>
<tr>
<th>TABLE 10-15 – Biochemical Markers of Thyroid Status</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Thyrotoxicosis</strong></td>
</tr>
<tr>
<td><strong>Increased</strong></td>
</tr>
<tr>
<td>Osteocalcin</td>
</tr>
<tr>
<td>Urine pyridinium collagen cross-links</td>
</tr>
<tr>
<td>Alkaline phosphatase (bone or liver)</td>
</tr>
<tr>
<td>Atrial natriuretic hormone</td>
</tr>
<tr>
<td>Sex hormonebinding globulin</td>
</tr>
<tr>
<td>Ferritin</td>
</tr>
<tr>
<td>von Willebrand's factor</td>
</tr>
<tr>
<td><strong>Decreased</strong></td>
</tr>
<tr>
<td>Low-density-lipoprotein cholesterol</td>
</tr>
<tr>
<td>Lp(a)</td>
</tr>
</tbody>
</table>

| **Hypothyroidism**                                  |
| **Increased**                                       |
| Creatine kinase (MM isoform)                        |
| Low-density-lipoprotein cholesterol                 |
| Lp(a)                                               |
| Plasma norepinephrine                               |
| **Decreased**                                       |
| Vasopressin                                         |
| Lp(a), lipoprotein a.                                |

Concentrations are somewhat higher in women than in men and are elevated several-fold in pregnant women and in newborns. Levels are elevated in three types of thyroid disorders:

- Goiter and thyroid gland hyperfunction
- Inflammatory or physical injury to the thyroid gland
- Differentiated follicular cell-derived thyroid tumors

Values are elevated in both endemic and sporadic nontoxic goiter, and the degree of elevation correlates with the thyroid size. Transient elevations occur in patients with subacute thyroiditis and as a result of trauma to the gland during thyroid surgery or after $^{131}$I therapy. Subnormal or undetectable concentrations are found in patients with thyrotoxicosis factitia and aid in differentiating this disorder from other causes of thyrotoxicosis with a low radioactive iodine uptake (RAIU).

Antithyroglobulin antibodies interfere with measurements of the Tg concentration, thus precluding its use in patients with Hashimoto’s disease. Serum Tg concentrations are increased in patients with both benign and differentiated malignant follicular cell-derived tumors of the thyroid and do not serve to distinguish between the two. After total thyroid ablation for papillary or follicular thyroid carcinoma, Tg should not be detectable and its subsequent appearance signifies the presence of persistent or recurrent disease. Secretion of Tg is TSH-dependent. Therefore, the serum Tg level may rise when suppressive therapy is withdrawn or after injections of rTSH, and this increases the sensitivity of the marker for the detection of persistent or recurrent thyroid carcinoma, even when $^{131}$I scans are negative.
Tests for Thyroid Autoantibodies

Graves’ disease and Hashimoto’s disease are well characterized and interrelated autoimmune thyroid disorders (AITDs). Thus, circulating antibodies and T cells against one or another thyroid antigen are often present.

Three varieties of thyroid autoantibodies are useful and widely available for clinical diagnostic use receptor [Table 10-16]. In this section, antibodies to Tg and thyroid peroxidase (TPO) are discussed. Antibodies directed against the TSH receptor,

### TABLE 10-16 -- Common Thyroid Autoantibodies (Ab)

<table>
<thead>
<tr>
<th>Antigen</th>
<th>Molecular Size</th>
<th>Abbreviation</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>TSH receptor</td>
<td>100 kd</td>
<td>TSHRAb</td>
<td>Antibody that causes Graves’ disease</td>
</tr>
<tr>
<td>TSHR-blocking Ab</td>
<td></td>
<td>TSHR-blocking Ab</td>
<td>Present in some thyroiditis patients</td>
</tr>
<tr>
<td>Thyroglobulin</td>
<td>330 kd</td>
<td>TgAb</td>
<td>Often undetectable using older techniques</td>
</tr>
<tr>
<td>Thyroid peroxidase</td>
<td>107 kd</td>
<td>TPOAb</td>
<td>Useful diagnostic marker</td>
</tr>
</tbody>
</table>

TSH, thyrotropin.

the cause of hyperthyroidism in patients with Graves’ disease, are covered in greater detail in Chapter 11.

Antibodies to Thyroid Peroxidase and Thyroglobulin

Table 10-17 summarizes some advantages and disadvantages of available techniques for measuring thyroid autoantibodies. The original technique of hemagglutination has many disadvantages, including lack of IgG specificity, low sensitivity, and operator dependency. Modern assay techniques have good precision because they depend on the direct measurement of the interaction between autoantibody and autoantigen (i.e., the interaction between labeled thyroid antigen and the patient’s serum). In general, the more sensitive an assay, the more it tends to be specific and precise. However, because many normal individuals exhibit low levels of autoantibodies, the clinical specificity of the more sensitive tests is reduced and the absolute concentration becomes more important; the higher the concentration of autoantibody, the greater the clinical specificity.

The prevalence of detectable thyroid autoantibodies in various disorders is shown in Table 10-18; however, data on concentration tend to vary from assay to assay even with the use of standardization.

### TABLE 10-17 -- Advantages and Disadvantages of Different Methods for Measurement of Autoantibodies to Thyroid Peroxidase and Thyroglobulin

<table>
<thead>
<tr>
<th>Technique</th>
<th>Precision</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immunofluorescence</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Hemagglutination</td>
<td>Low</td>
<td>Low</td>
<td>Variable</td>
<td>High</td>
</tr>
<tr>
<td>ELISA</td>
<td>Variable</td>
<td>High</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Radioassay</td>
<td>High</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
</tr>
</tbody>
</table>

ELISA, enzyme-linked immunosorbent assay.

The actual standard serum preparation, however, cannot be included in every assay. Instead, a serum pool is usually compared and normalized to the original standard. Yet autoantibodies differ considerably in their affinity and epitope recognition of antigen. Hence, despite this attempt at standardization, assay results from different commercial assays may still vary considerably. When following antibody titers (e.g., after the treatment of thyroid cancer), it is important to use the same autoantibody assay.

### TABLE 10-18 -- Prevalence of Thyroid Autoantibodies (Ab)

<table>
<thead>
<tr>
<th>Group</th>
<th>TSHRAb (%)</th>
<th>hTgAb (%)</th>
<th>hTPOAb (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>General population</td>
<td>0</td>
<td>520</td>
<td>827</td>
</tr>
<tr>
<td>Graves’ disease</td>
<td>8095</td>
<td>5070</td>
<td>5080</td>
</tr>
<tr>
<td>Autoimmune thyroiditis</td>
<td>1020</td>
<td>8090</td>
<td>90100</td>
</tr>
<tr>
<td>Relatives of patients</td>
<td>0</td>
<td>4050</td>
<td>4050</td>
</tr>
<tr>
<td>Patients with IDDM</td>
<td>0</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>Pregnant women</td>
<td>0</td>
<td>14</td>
<td>14</td>
</tr>
</tbody>
</table>

IDDM, insulin-dependent diabetes mellitus; Tg, thyroglobulin; TPO, thyroid peroxidase.

among laboratories, assays for thyroid autoantibodies should be standardized (i.e., results should be expressed in relation to a widely available standard preparation, as with hormone immunoassays). Although there are no formal “international standards” for human thyroid autoantibodies, TgAb and TPOAb standard sera are available from the National Institute for Biological Standards in the United Kingdom and are an essential component of many thyroid autoantibody assays. Such results can then be expressed in units per milliliter.

The actual standard serum preparation, however, cannot be included in every assay. Instead, a serum pool is usually compared and normalized to the original standard. Yet autoantibodies differ considerably in their affinity and epitope recognition of antigen. Hence, despite this attempt at standardization, assay results from different commercial assays may still vary considerably. When following antibody titers (e.g., after the treatment of thyroid cancer), it is important to use the same autoantibody assay.

### Pathogenic Role of Thyroglobulin and Thyroid Peroxidase Antibodies

Tg and TPO autoantibodies are a secondary response to thyroid injury and do not cause disease themselves. Both types of antibodies are polyclonal, and although they are of the IgG class, are not restricted to one particular IgG subclass. Polyclonality mitigates against a primary role in disease, but these antibodies may be important in determining the end-organ effects and may also be determinants of chronicity.

Whereas both TPOAb and TgAb levels correlate with lymphocytic infiltration of the thyroid gland, they do not transfer disease from mother to fetus or between animals. Thus, thyroid antibodies to Tg and TPO do not initiate disease. Both antibodies, however, may have complement-fixing cytotoxic activity, and TPOAb
autoantibodies, in particular, correlate with thyroidal damage and lymphocytic infiltration. Patients with AITDs have autoantibody "fingerprints," a characteristic spectrum of Tg and TPO autoantibodies belonging to IgG1, IgG2, IgG3, and IgG4 subclasses. This pattern may be inherited as an autosomal dominant trait within families.\[351\] [352]

Because IgG1 antibodies fix complement, whereas IgG4 antibodies do not, for example, this pattern of distribution may affect the disease phenotype.

**Thyroid Autoantibodies in Hashimoto’s Disease**

The disease most associated with TgAb and TPOAb is autoimmune thyroiditis, a term that embraces both goitrous Hashimoto's disease and atrophic thyroid failure. The titers of these antibodies correlate with the degree of thyroidal lymphocytic infiltration. Immunoassays show that both TgAb and TPOAb are found in almost 100% of such patients, but TPO antibodies are of higher affinity and in higher concentrations. This may be because TgAb is bound by circulating Tg, causing its concentration to be underestimated. In an unclear clinical situation, positive TgAb and TPOAb levels are diagnostic of primary autoimmune thyroid disease.

Antibody measurements may also be useful prognostically in mildly (subclinically) hypothyroid patients (i.e., with elevated TSH and normal T₄ levels), since the rate of overt hypothyroidism is about 3% to 5% per year in patients with a mildly increased TSH level and positive thyroid autoantibodies. Falling titers of TgAb indicate a good prognosis in treated thyroid cancer patients who show this autoantibody (20% of patients) and TPOAb has been shown to be an important predictor of postpartum thyroiditis, a transient form of autoimmune thyroiditis found in 8% to 10% of all women and in 33% or more of TPOAb-positive mothers.

**Thyroid Autoantibodies in Graves’ Disease**

Antibodies to Tg and TPO are also detectable in 50% to 90% of patients with Graves' disease, indicative of the associated thyroiditis that is evident histologically. Hence, Graves' disease may occur on a background of autoimmune thyroiditis. Although the presence of such autoantibodies favors a diagnosis of an autoimmune cause for the hyperthyroidism over other causes, the tests are neither sensitive nor specific in this setting and are interpretable only as part of the clinical scenario.

**Thyroid Autoantibodies in Nonautoimmune Thyroid Disorders**

Antibodies to Tg and TPO are more common in patients with sporadic goiter, multinodular goiter, and isolated thyroid nodules and cancer than in the general population. This finding usually represents an associated thyroiditis on histologic examination. Low levels of thyroid autoantibodies may occur transiently in patients with subacute (de Quervain's) thyroiditis but correlate poorly with disease course and are probably a nonspecific response to thyroid injury. There is also a higher prevalence of thyroid autoantibodies in other autoimmune diseases, such as insulin-dependent diabetes mellitus (IDDM), indicative of a common genetic susceptibility and etiology.

**The "Normal" Population**

Although the prevalence of thyroid autoantibodies depends on the technique used for detection, autoantibodies to Tg and TPO are common in the general population (see Table 10-18) and, at all ages, are almost five times more common in women than in men. The tendency to secrete thyroid autoantibodies is inherited in a mendelian-dominant manner and has been linked to polymorphisms in the CTLA-4 gene. Selected groups at risk include younger women and relatives of patients with anAITD, in whom the incidence is higher. Low levels of autoantibodies to TPO and Tg are of uncertain significance in the presence of normal thyroid function; however, within a family with anAITD, they remain a significant risk factor.

**Clinical Utility: Establishment of Disease Etiology**

Thyroid failure has a variety of causes, and autoimmune thyroid disease can be inferred by the presence of a family history of Graves’ disease or Hashimoto’s disease. However, the only simple way of confirming the autoimmune diathesis, other than biopsy, is the presence of significant levels of thyroid autoantibodies. The measurement of thyroid autoantibodies also allows the generation of data regarding the prevalence of AITDs within the patient's family.

**Prediction of Disease Onset**

Patients with increased TSH and normal T₄ levels progress to overt thyroid failure at a rate of about 5% per year if thyroid autoantibody levels are elevated. Hence, patients with mild (subclinical) hypothyroidism (increased TSH levels but apparently normal free T₄ values) and thyroid autoantibodies are at twice the risk for development of thyroid failure as patients without thyroid antibodies.

**Thyroid Cancer**

Approximately 20% to 40% of patients with thyroid cancer have thyroid autoantibodies, and their presence (indicating underlying immunoreactivity against thyroid cell antigens) may suggest a better prognosis for these patients. After total thyroidectomy and radiodine ablation, a sensitive assay should show a serum Tg level below 1 ng/mL. The presence of TgAb significantly interferes with this assessment, and attempts to correct for the presence of TgAb using Tg recovery from the serum are not always helpful. A falling titer or total loss of TgAb in such patients, however, is an important and reliable prognostic sign indicating the absence of thyroid cell antigens.

**Risk Analysis for Postpartum Thyroid Disease**

The prevalence of PPTD is 8% to 10% in the first 4 to 12 months after delivery. More than 33% of women who are TPOAb-positive early in pregnancy, particularly patients with high thyroid autoantibody levels, develop some form of PPTD. Hence, the measurement of TPOAb is important in pregnancy screening.

**Risk Analysis for Early Pregnancy Loss**

Thyroid autoantibodies are markers of an at-risk pregnancy but do not imply that thyroid dysfunction causes the increased risk. Many studies have shown double the rate of pregnancy loss in women with these antibodies. However, the presence of thyroid autoantibodies appears to be a signal of immune uncertainty. The role of placental megalin receptors, which may be activated by TgAb, is also of interest.

**Thyroid Disease Screening in Associated Autoimmune Conditions**

AITDs occur commonly with other forms of autoimmune disease. For example, patients with IDDM are at particular risk, and the presence of thyroid autoantibodies is helpful in selecting patients for monitoring of thyroid function.
Radioiodine Uptake

The only direct test of thyroid function employs a radioactive isotope of iodine as a tag for the body’s stable form of iodine. $^{127}$I. Most often the test involves the measurement of the fractional uptake by the thyroid of a tracer (i.e., a chemically inconsequential) dose of radioiodine. However, several factors have caused this test to be less frequently used and less valuable for diagnosis of thyroid disorders than in the past; specifically, (1) the improvement in indirect methods for assessing thyroid status and (2) the decrease in normal values for thyroid RAIU consequent to the widespread increase in daily dietary iodine intake. $^{366}$

Both $^{131}$I (half-life, 8.1 days) and $^{123}$I (half-life, 0.55 days) emit gamma radiation, which permits their external detection and quantitation at sites of accumulation, such as the thyroid gland. These isotopes ($I^*$) are physiologically indistinguishable, not only from one another but also from the naturally occurring $^{127}$I, which permits their use as valid tracers. The shorter half-life of $^{123}$I is preferable because the radiation delivered to the thyroid per amount of administered $^{123}$I is only about 1% of that delivered by $^{131}$I.

Physiologic Basis

When tracer quantities of inorganic radioiodine are administered either orally or intravenously, the isotope quickly mixes with the endogenous stable iodide in the extracellular fluid and begins to be removed by the two major sites of clearance, the thyroid gland and the kidneys. As this process continues, the plasma level of $I^*$ decreases exponentially. Normally, low values are reached by 24 hours, and inorganic $I^*$ is virtually undetectable in the plasma 72 hours after its administration. The thyroid content of $I^*$ increases rapidly during the early hours, then at a decreasing rate until a plateau is approached. The proportion of administered $I^*$ that is ultimately accumulated by the thyroid gland is a function of the clearance of iodide by the thyroid and kidneys. The relation is simply expressed as follows:

$$\text{RAIU at } \frac{C_T}{C_T + C_K}$$

where $C_T$ is the thyroid iodide clearance rate and $C_K$ is the renal iodide clearance rate.

The normal thyroid iodide clearance rate is approximately 0.4 L/hour, and the renal iodide clearance rate is 2.0 L/hour. Thus, the uptake of $I^*$ normally approximates 0.17 of the administered dose.

Measurements of the RAIU are generally made at 24 hours, both as a matter of convenience and because the value at 24 hours is usually near the plateau. The RAIU usually indicates the rate of thyroid hormone synthesis and, by inference, the rate of thyroid hormone release into the blood.

Radioactive Iodine Uptake

Little difference is noted if the uptake is measured at any time during the day following that on which the isotope was administered. For calculating therapeutic radioiodine doses in treating thyrotoxic Graves’ disease, an uptake at 3 to 6 hours may produce results comparable to those found at 20 to 28 hours. $^{367}$ With the use of this modified early RAIU measurement, diagnosis and treatment of thyrotoxic Graves’ disease can be accomplished on the same day.

In general, the range of normal values for the 24 hr radioiodine uptake in North America is approximately 5% to 25%. Higher values may indicate iodine deficiency or thyroid hyperfunction. As with other procedures, however, values in patients with mild hyperthyroidism may be at or just above the upper limit of the normal range (Table 10-19).

<table>
<thead>
<tr>
<th>TABLE 10-19 — Factors That Influence 24-Hour Thyroid Iodide Uptake</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Factors That Increase Uptake</strong></td>
</tr>
<tr>
<td>Increased hormone synthesis</td>
</tr>
<tr>
<td>Hyperthyroidism</td>
</tr>
<tr>
<td>Response to glandular hormone depletion</td>
</tr>
<tr>
<td>Recovery from thyroid suppression</td>
</tr>
<tr>
<td>Recovery from subacute thyroiditis</td>
</tr>
<tr>
<td>Antithyroid agents</td>
</tr>
<tr>
<td>Excessive hormone losses</td>
</tr>
<tr>
<td>Nephrotic syndrome</td>
</tr>
<tr>
<td>Chronic diarrheal states</td>
</tr>
<tr>
<td>Soybean ingestion</td>
</tr>
<tr>
<td>Normal hormone synthesis</td>
</tr>
<tr>
<td>Iodine deficiency</td>
</tr>
<tr>
<td>Dietary insufficiency</td>
</tr>
<tr>
<td>Excessive loss (dehalogenase defect, pregnancy)</td>
</tr>
<tr>
<td>Hormone biosynthetic defects</td>
</tr>
<tr>
<td><strong>Factors That Decrease Uptake</strong></td>
</tr>
<tr>
<td>Decreased hormone synthesis</td>
</tr>
<tr>
<td>Primary hypofunction</td>
</tr>
<tr>
<td>Primary hypothyroidism</td>
</tr>
<tr>
<td>Antithyroid agents</td>
</tr>
<tr>
<td>Hormone biosynthetic defects</td>
</tr>
<tr>
<td>Hashimoto’s disease</td>
</tr>
<tr>
<td>Subacute thyroiditis</td>
</tr>
<tr>
<td>Secondary hypofunction</td>
</tr>
</tbody>
</table>
Not reflecting decreased hormone synthesis
Increased availability of iodine
Diet or drugs
Cardiac or renal insufficiency
Increased hormone release
Very severe hyperthyroidism (rare)

Exogenous Thyroid Hormones

Increased availability of iodine
Cardiac or renal insufficiency
Increased hormone release

States Associated with Increased Radioactive Iodine Uptake

Hyperthyroidism

Hyperthyroidism causes an increased RAIU unless body iodide stores are increased. Such increases in uptake are always evident except in patients with severe thyrotoxicosis, in whom release of hormone can be so rapid that the thyroid content of I\(^{131}\) has decreased to the normal range by the time the measurement is made. This condition is rare and is usually associated with obvious thyrotoxicosis.

Aberrant Hormone Synthesis

The RAIU can be increased in the absence of hyperthyroidism in disorders in which iodine accumulation is normal but the secretion of hormone is impaired, such as in patients with abnormal thyroglobulin synthesis. The magnitude of the increase in uptake and the time at which the plateau is achieved vary with the nature and severity of the disorder. Differentiation of these states from hyperthyroidism is generally not difficult; in the former, clinical findings and laboratory evidence of hyperthyroidism are lacking, and indeed hypothyroidism may be present.

Iodine Deficiency

The RAIU is increased in acute or chronic iodine deficiency, as demonstrated by measurement of urinary iodine excretion, with urinary iodine values lower than 100 \(\mu g/day\) indicating deficiency. Chronic iodine deficiency is usually the result of an inadequate content of iodine in the food and water (endemic iodine deficiency). Patients with cardiac, renal, or hepatic disease may develop iodine deficiency if given diets severely restricted in salt, especially if diuretic agents are administered.

Response to Thyroid Hormone Depletion

Rebound increases in the RAIU are seen after withdrawal of antithyroid therapy, after subsidence of transient or subacute thyroiditis, and after recovery from prolonged suppression of thyroid function by exogenous administered. A striking increase in RAIU occurs in patients with iodide-induced myxedema after cessation of iodide administration. The duration of the rebound depends on the time required to replenish thyroid hormone stores.

Excessive Hormone Losses

In patients with nephrotic syndrome, excessive losses of hormone in the urine occurring in association with urinary loss of binding protein cause a compensatory increase in hormone synthesis and in the RAIU. A similar sequence may occur when losses of hormone via the gastrointestinal tract are abnormal, as in chronic diarrheal states or during ingestion of agents (e.g., soybean protein, cholestyramine) that bind T\(_4\) in the gut.

States Associated with Decreased Radioactive Iodine Uptake

A general increase in iodine intake has made RAIU values in hypothyroidism indistinguishable from those at the lower end of the normal range. Therefore, the major indication for measuring the RAIU is to establish the causes of thyrotoxicosis associated with decreased values of the RAIU.

Hypothyroidism

The problems involved in using the RAIU as an aid to the diagnosis of hypothyroidism have been discussed.

Exogenous Thyroid Hormone: Thyrotoxicosis Factitia

Except in disorders in which homeostatic control is disrupted or overridden (e.g., with Graves' disease or autonomously functioning thyroid nodules), administration of exogenous thyroid hormone suppresses TSH secretion and reduces the RAIU, usually to values below 5%.

Low values of the RAIU in a patient who is clinically thyrotoxic may indicate the presence of thyrotoxicosis factitia, the syndrome produced by the ingestion of excess thyroid hormone. The unmeasurably low level of T\(_g\) in serum differentiates thyrotoxicosis from other causes of thyrotoxicosis with a decreased RAIU.

Disorders of Hormone Storage

The RAIU is usually low in the early phase of subacute thyroiditis and in chronic thyroiditis with transient hyperthyroidism. In these instances, inflammatory follicular disruption leads to loss of the normal storage function of the gland and leakage of hormone into the blood. In the early stage of subacute thyroiditis, leakage of hormone is usually sufficient to suppress TSH secretion and the RAIU. Transient hyperthyroidism often occurs late in both diseases, when stores of preformed hormone are depleted; the RAIU may return to normal or increased values at that time.

Exposure to Excessive Iodine

Exposure to excessive iodine is the most common cause of a subnormal RAIU. Such decreases are spurious in the clinical sense because they do not indicate decreased absolute iodine uptake or decreased hormone production but can be produced by the introduction of excessive iodine in any form (organic, inorganic, or elemental). Special offenders are organic iodinated dyes used as x-ray contrast media and amiodarone (see Table 10-8). The duration of suppression of the uptake varies among individuals and with the compound administered. In general, dyes used for pyelography or computed tomography are cleared within weeks, whereas amiodarone may influence the uptake for up to 12 months because of its storage in fat.

A single large dose of inorganic iodide can decrease the uptake for several days, and chronic ingestion of iodide may depress the uptake for many weeks. Lugol's solution or saturated solution of potassium iodide (SSKI) in the usual dosage (2 to 5 drops three times/day) can deliver up to about 500 \(\mu g\) of iodine daily, as opposed to the customary intake of about 200 \(\mu g/day\) in the United States. Excessive quantities of iodine may also be present in vitamin and mineral preparations, vaginal or rectal suppositories, and iodinated antiseptics such as povidone (see Table 10-8). In patients with thyrotoxicosis, the RAIU may help to differentiate excessive hormone synthesis with an increased RAIU from destructive thyrotoxicosis with a subnormal RAIU. Inhibition of uptake by excess stable iodine is of shorter duration in hyperthyroid than in normal individuals.
The measurement of urinary iodine excretion is an invaluable means of establishing or excluding the existence of excessive body iodide stores; a random urine sample can be obtained, and the 24-hour iodine excretion can be extrapolated from the iodine/creatinine ratio. Values in excess of several milligrams per day can explain a low RAIU value, whereas values less than 1 mg/day suggest that a low RAIU value is due to one of the other disorders discussed in this section.

Figure 10-18 Examination of the thyroid gland. A. Sagittal section demonstrates relations of the isthmus of the normal thyroid gland. The superior border is inferior to the cricoid cartilage. The inferior thyroid border is essentially at the level of the superior surface of the manubrium. The inferior portions of the lateral lobes (not shown) extend more inferiorly than the isthmus. B. The cricoid cartilage is regarded as an important landmark. Especially when the thyroid gland is thought to be essentially normal or subnormal in size, the cricoid should be located. This is easily accomplished. The index fingers are then inserted so that their superior portion rests against the inferior portion of the cricoid while the inferior portion of these fingers is over the superior portion of the thyroid. The second and third fingers are rotated over other portions of the gland to evaluate its size, contour, consistency, possible adherence to surrounding structures, and other features. Because there is marked variation among different subjects in the length and thickness of the neck and in the length of the trachea superior to the level of the manubrium, the relative position of the thyroid may vary. In some cases, essentially all of the thyroid gland rests posterior to the sternum. In most instances, however, by having the patient moderately extend the neck (short of tightening the anterior neck muscles) and swallow repeatedly, it is possible to palpate most or all of the gland. Despite marked variations in neck-chest relations, thyroid tissue, when present, is found within 1 cm of the cricoid. By concentrating the palpation meticulously in the area where the thyroid is normally found, with rare exceptions the examiner can outline small as well as enlarged glands.
CLINICAL EVALUATION AND INITIAL LABORATORY TESTING

Manifestations of thyroid disease are usually due to (1) excessive or insufficient production of thyroid hormone, (2) local symptoms in the neck (principally goiter but occasionally pain or compression of adjacent structures), or, (3) in the case of Graves' disease, ophthalmopathy or dermopathy. Although attention is directed initially at the major features, it is crucial to define the metabolic state and to ascertain the nature of the underlying disorder. A functional diagnosis of thyroid disease is based on a carefully taken history, a thorough search for the physical signs of hypothyroidism or thyrotoxicosis, and an appraisal of the results of laboratory tests. Although conditioned by the functional diagnosis, the anatomic diagnosis depends largely on the examination of the thyroid gland itself.

Physical Examination

Local examination of the neck is best accomplished with the patient seated in a good light and with the neck moderately extended. The patient must be provided with a cup of water to facilitate swallowing.

The physician first inspects the neck from the front and on the sides, especially while the patient swallows, with the neck slightly extended. The presence of old surgical scars, distended veins, and redness or fixation of the overlying skin should be noted. If a mass is present, attention should be directed to its location and to whether it moves when the patient swallows.

The position of the trachea is noted. Movement on swallowing is a characteristic of the thyroid gland because it is enshrouded in the pretracheal fascia; this feature distinguishes a goiter from most other neck masses. If a goiter is large, however, that it occupies all the available space in the neck or if the thyroid gland is the seat of an invasive carcinoma or Riedel's thyroiditis that has caused fixation to adjacent structures, movement on swallowing may be lost. The physician should also inspect the dorsum of the tongue, which is the origin of the thyroglossal duct and rarely the seat of lingual thyroid tissue.

Standing behind the seated patient, the physician may examine the thyroid gland by palpating with the fingertips of both hands. The position of the cricoideal cartilage is determined first because the superior border of the isthmus lies just below it (see Fig. 10-18). The isthmus is a band of tissue crossing the front of the trachea joining the two lobes. The examiner then attempts to outline the thyroid gland and to determine the limits of the lower borders of the lateral lobes while the patient swallows sips of water at appropriate intervals. With practice, a normal thyroid gland can usually be palpated, particularly in women.

An alternative approach to the thyroid gland examination is for the physician to face the seated patient and use gentle pressure with the thumb to locate the thyroid isthmus. The right thumb is then moved laterally, without release of pressure, to compress the right lobe of the thyroid against the trachea as the patient again swallows sips of water. This strategy allows the palpat ing thumb to slide under and laterally displace the medial border of the sternocleidomastoid muscle. A similar strategy with the left thumb is employed for the left lobe. This technique is especially useful as an aid in detecting small nodules that may not be easily appreciated with the posterior approach.

The examiner notes the shape of the gland, its size in relation to normal, and its consistency, which is usually slightly more firm than adipose tissue. The normal thyroid lobe is approximately the same in size in frontal projection as the terminal phalanx of the patient's thumb. Whereas the diffuse colloid goiter and the hyperplastic gland in Graves' disease tend to be softer than normal, the gland in Hashimoto's disease is usually firm. In rare circumstances, the gland that is the seat of carcinoma or Riedel's thyroiditis may be "stony" hard. Irregularities of the surface, variations in consistency, and tender areas should be noted. If nodules are palpated, their shape, size, position, transluency, and consistency in relation to the surrounding tissue should be determined.

A search should be made for the pyramidal lobe, a thin band of tissue extending upward from the isthmus to the thyroid cartilage to the right or left of the midline. The pyramidal lobe may be mistaken for a pretracheal lymph node that sometimes accompanies thyroid carcinoma or thyroiditis. It is usually palpable in patients with generalized thyroid disease, such as Hashimoto's or Graves' disease. Thyroglossal cysts are midline masses that remain attached to the base of the tongue by the fibrotic thyroglossal duct and that move upward when the tongue is protruded.

During palpation, a vascular thrill may be felt that, in the absence of cardiac disease, is suggestive of hyperthyroidism. Finally, palpation should always include examination of the regional lymph nodes.

Auscultation of the neck can indicate the vascularity of an enlarged gland. A systolic or continuous bruit is sometimes heard over a hyperplastic gland. The physician should take care to distinguish a thyroid bruit from a murmur transmitted from the base of the heart or from a venous hum that can be obliterated by gentle compression of the external jugular vein or by turning the patient's head. A venous hum is generally found in younger patients with high cardiac output, such as in Graves' disease or severe anemia.

An arm-raising test is useful when a retrosternal goiter is suspected. The basis for this maneuver is that if the size of the thoracic inlet is already reduced by a retrosternal goiter, raising both arms until they touch the sides of the head further narrows the thoracic inlet and causes congestion and venous engorgement of the face and, sometimes, respiratory distress (Pemberton's sign) or even (rarely) syncope.

In addition to examination of the thyroid gland and the regional lymph nodes, the physician should seek evidence of compression or displacement of adjacent structures. Hoarseness may indicate compression of the recurrent laryngeal nerve, usually by a malignant thyroid neoplasm, and this possibility should be confirmed by laryngoscopy. Displacement of the trachea may be evident, and inspiratory stridor may indicate compression of the trachea.
Laboratory Evaluation

Initial laboratory tests for any patient with suspected thyroid disease should include an immunometric TSH assay. If the physician is reasonably confident that a functional disorder of the thyroid is present, an FT₄ or FT₃ should be included as an initial test to confirm the presence and assess the degree of the abnormality inferred from the TSH result. There is rarely reason to measure total T₃ in the initial evaluation unless the patient is receiving liothyronine or liotrix. More extensive guidelines for laboratory procedures relevant to specific thyroid conditions are discussed in Chapter 11, Chapter 12, and Chapter 13.
References


34. Larsen PR. Thyroidal triiodothyronine and thyroxine in Graves’ disease: correlation with presurgical treatment, thyroid status, and iodine content. J Clin Endocrinol Metab 1975; 41:10981104.


38. Lauberg P. Mechanisms governing the relative proportions of thyroxine and 3,3',5'-triiodothyronine in thyroid secretion. Metabolism 1984; 33:379392.
41. Lazarus JH. The effects of lithium therapy on thyroid and thyrotropin-releasing hormone. Thyroid 1998; 8:909913.
44. Glinser D. What happens to the normal thyroid during pregnancy? Thyroid 1999; 9:613635.
63. Ain KB, Mori Y, Relfoff S. Reduced clearance rate of thyroxine-binding globulin (TBG) with increased sialylation: a mechanism for estrogen-induced elevation of serum TBG concentration. J Clin Endocrinol Metab 1987; 65:989696.
70. Surks MI, Oppenheimer JH. Postoperative changes in the concentration of thyroxine-binding prealbumin and serum free thyroxine. J Clin Endocrinol Metab 1964; 24:794801.


Oppenheimer JH, Schwartz HL. Stereospecific transport to triiodothyronine from plasma to cytosol and from cytosol to nucleus in rat liver, kidney, brain, and heart. J Clin Invest 1985; 75:147154.


References


192. Davis PJ, Davis FB. Nongenomic actions of thyroid hormone. Thyroid 1996; 6:497504.


Mariotti S, Delange F. The disorders induced by iodine deficiency. Thyroid 1994; 4:107128.


References


7. Lauberg P. Persistent problems with the specificity of immunometric TSH assays. Thyroid 1993; 3:279283.


Chapter 11 - Thyrotoxicosis

Terry F. Davies
P. Reed Larsen

OVERVIEW

The term thyrotoxicosis refers to the biochemical and physiologic manifestations of excessive quantities of the thyroid hormones. We prefer the term thyrotoxicosis rather than hyperthyroidism to describe this disorder, because it need not originate in the thyroid gland. The term hyperthyroidism is reserved for disorders that result from overproduction of hormone by the thyroid gland itself, of which Graves’ disease is the most common (Table 11-1). The manifestations depend on the severity of the disease, the age of the patient, the presence or absence of extrathyroidal manifestations, and the specific disorder producing the thyrotoxicosis.

Peripheral Clinical Manifestations (Table 11-2)

Skin and Hair

The most characteristic change in thyrotoxicosis is the warm, moist feel of the skin that results from cutaneous vasodilatation and excessive sweating. Although the hands are usually warm and moist, the texture of the hands may be altered by occupational or environmental factors; hence, texture is best assessed on the inner aspect of the arm or over the chest. The elbows are smooth and pink, the complexion is rosy, and the patient blushes readily. Palmar erythema may resemble “liver palms,” and telangiectasia may be present. Increased diffuse pigmentation resembles that in adrenal insufficiency, but buccal pigmentation does not occur. The hair is fine and friable, and hair loss may be excessive. A history of early graying in the patient or in relatives is said to be common in Graves’ disease. The nails are often soft and friable. A characteristic finding is Plummer’s nails, a term applied to separation of the distal margin of the nail from the nail bed, with irregular recession of the junction (onycholysis).

Eyes

Retraction of the upper eyelid, evident as the presence of a rim of sclera between the lid and the limbus, is common in all forms of thyrotoxicosis, regardless of the underlying cause, and is responsible for the bright-eyed “stare” or “fish eyes” of the patient with thyrotoxicosis. Lid lag is the phenomenon in which the upper lid lags behind the globe when the patient is asked to gaze slowly downward, and globe lag occurs when the globe lags behind the upper lid when the patient gazes slowly upward. The movements of the lids may be jerky and spasmodic, and a fine tremor of the lightly closed lids can often be observed in severe cases. These ocular manifestations appear to be the result of increased adrenergic activity. It is important to differentiate these ocular manifestations, which occur in all forms of thyrotoxicosis, from those of infiltrative orbitopathy, which are characteristic of hyperthyroid Graves’ disease (see later).

Cardiovascular System

Alterations in cardiovascular function are due to increased circulatory demands that result from the hypermetabolism and the need to dissipate the excess heat produced. At rest, peripheral vascular resistance is decreased and cardiac output is increased as a result of an increase in stroke volume and heart rate. Thyroid hormones in excess have a direct inotropic effect mediated by alterations of contractile proteins. Tachycardia is almost always present, and tachycardia during sleep (pulse rate >90 beats/minute) serves to distinguish tachycardia of thyrotoxic origin from that of psychogenic causes. Widening of the pulse pressure is due to an increase in systolic pressure and a decrease in diastolic pressure.

<table>
<thead>
<tr>
<th>TABLE 11-1 – Varieties of Thyrotoxicosis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sustained Hormone Overproduction (Hyperthyroidism)</strong></td>
</tr>
<tr>
<td>Graves’ disease</td>
</tr>
<tr>
<td>Toxic multinodular goiter</td>
</tr>
<tr>
<td>Toxic adenoma</td>
</tr>
<tr>
<td>Iodine-induced (Jod-Basedow)</td>
</tr>
<tr>
<td>Trophoblastic tumor</td>
</tr>
<tr>
<td>Increased TSH secretion</td>
</tr>
<tr>
<td><strong>No Associated Hyperthyroidism</strong></td>
</tr>
<tr>
<td>Thyrotoxicosis factitia</td>
</tr>
<tr>
<td>Subacute thyroiditis</td>
</tr>
<tr>
<td>Thyroiditis with transient thyrotoxicosis (painless thyroiditis, silent thyroiditis, postpartum thyroiditis)</td>
</tr>
<tr>
<td>Ectopic thyroid tissue (struma ovarii, functioning metastatic thyroid cancer)</td>
</tr>
</tbody>
</table>

RAIU, radioactive iodine uptake; TSH, thyrotropin.

*Except for iodine-induced hyperthyroidism, associated with increased values of RAIU. Associated with decreased values of RAIU.

The increased force of cardiac contraction is often felt by the patient as palpitations and may be evident on inspection or palpation of the precordium. Because of the diffuse and forceful nature of the apex beat, the heart may seem enlarged but echocardiographic findings are usually normal. Heart sounds are loud and ringing, and a systolic or even a late diastolic or presystolic murmur (Means-Lerman scratch) may be present at the apex. A scratchy systolic sound along the left sternal border, resembling a pleuropericardial friction rub, may also be heard. These manifestations usually abate when a normal metabolic state is restored. Mitral valve prolapse occurs more frequently than in the normal population and may persist indefinitely.
Cardiac arrhythmias are almost invariably supraventricular. Approximately 10% of patients with thyrotoxicosis have atrial fibrillation, and a similar percentage of patients with otherwise unexplained atrial fibrillation are thyrotoxic. In a study of more than 2000 individuals 60 years of age or older, atrial fibrillation developed in 28% of those with a suppressed thyrotropin (TSH, or thyroid-stimulating hormone) level. 3 Paroxysmal supraventricular tachycardia may be demonstrable or may be suggested by the history. Systolic time intervals are altered in thyrotoxicosis, the pulse-ejection period is shortened, and the ratio of pre-ejection period to left ventricular ejection time is decreased.

The adequacy of the circulation is an important issue. The increased cardiovascular cost of a standard workload or metabolic challenge is adequately met if the patient is not or has not previously been in heart failure. Thus, in most patients without underlying heart disease, cardiac competence is maintained. Mild edema may occur in the absence of heart failure. Thyrotoxicosis may lead to congestive heart failure, but even so, the circulation time may remain short. Heart failure usually occurs in patients with preexisting heart disease, but it may not be possible to determine whether underlying heart disease is present until after thyrotoxicosis is relieved.

Atrial fibrillation decreases the efficiency of the cardiac response to any increased circulatory demand and may play a role in causing cardiac failure. Attempts to convert atrial fibrillation to sinus rhythm are usually of no avail while thyrotoxicosis is present. Regardless of the type of rhythm, the response to digitalis is decreased, possibly because of accelerated metabolism of the drug, and large quantities may be required to produce a clinical effect. Resistance to digitalis and failure of cardiac decompensation to respond to a usually adequate regimen should suggest the possibility of thyrotoxicosis. It is obviously important to deal definitively with thyrotoxicosis in a patient with concomitant heart disease.

Sympathetic Nervous System

Many of the manifestations of thyrotoxicosis and sympathetic nervous system activation are similar. As judged from the plasma concentrations of epinephrine and norepinephrine, as well as their urinary excretion and that of their metabolites, the activity of the sympathetic nervous system is not increased in patients or animals with thyrotoxicosis, and thyroid hormones may exert effects separate from, but similar and additive to, those of the catecholamines. Consideration of the relationship between catecholamines and thyroid hormone excess and deficiency reveals the futility of attempting a generalization in this area. 4

The reduction in heart rate and in some clinical manifestations of hyperthyroidism induced by -blockade in patients with this condition has led to the concept of an increased sympathetic tone or increased cardiac sensitivity to the sympathetic nervous system. Careful studies have shown that this is clearly not the case in terms of the heart. 4 Adequate -adrenergic blockade reduces the basal level of cardiac output, but the slope of the epinephrine dose-response curve is not altered in hyperthyroidism. In other tissues, the situation may be even more complex and species differences may exist.

In hyperthyroid patients, there are no alterations in -adrenergic receptor number on lymphocytes and no changes in lymphocyte -adrenergic responsiveness, suggesting that clinical effects are direct on the target cell. However, the 1-adrenergic receptor is down-regulated by thyroid hormone. 5

In normal subjects given thyroid hormone for 10 days, the -adrenergic receptor number in fat and skeletal muscle increased 60% and 30%, respectively, but metabolic and hemodynamic sensitivity to infused epinephrine in vivo were not altered. 6 7 There was no evidence of increased glycoemic, lipolytic, glycojenolytic, or ketogenic sensitivity to catecholamines,
when caloric intake exceeds the metabolic demand. Anorexia, rather than hyperphagia, occurs in about one third of elderly thyrotoxic patients and contributes to the picture of apathetic thyrotoxicosis.

Stools are frequently soft and the frequency of bowel movements is increased, but diarrhea is rare. When constipation has preceded the development of thyrotoxicosis, bowel function may become normal. Anorexia, nausea, and vomiting are uncommon but may occur with severe disease. These symptoms, as well as abdominal pain, may be forerunners of accelerated thyrotoxicosis. The increased gastric emptying and intestinal motility in thyrotoxicosis appear to be responsible for slight malabsorption of fat, and these functions return to normal when a normal metabolic state has been restored. Celiac disease and Graves’ disease may coexist, and a high proportion of patients have gastric achlorhydria. In most cases, acid secretion returns after relief of the thyrotoxicosis. Autoantibodies against gastric parietal cells are detectable in some patients with Graves’ disease, and approximately 3% have pernicious anemia. In the oral glucose tolerance test, the glycemic peak is frequently delayed.

Hepatic dysfunction occurs, particularly when thyrotoxicosis is severe; hypoproteinemia and increases in serum alanine aminotransferase and alkaline phosphatase levels may be present. Hepatomegaly and jaundice occasionally develop. Splanchnic oxygen consumption is increased, whereas splanchnic blood flow is essentially unchanged. As a result, the arteriovenous oxygen difference across the splanchnic bed is increased; hence, hypoxia may contribute to hepatic dysfunction. Hypoxia and the relative caloric deprivation may partly account for the depletion of hepatic glycogen that is evident both in the response to glycogenolytic agents and on direct analysis. In the absence of severe thyrotoxicosis or congestive heart failure, the liver usually appears normal on light microscopic examination. Ultramicrorscopic examination of the liver reveals enlarged mitochondria and hypertrophic smooth endoplasmic reticulum. Graves’ disease and autoimmune hepatitis may also coexist.

Nervous System

Alterations in nervous system function in patients with thyrotoxicosis are manifested by nervousness, emotional lability, and hypokinesia. The nervousness is not typical of the patient who is chronically anxious but, rather, is characterized by restlessness, shortness of attention span, and a compulsion to be moving around, sometimes referred to as almost “fidgeting.” Unlike the patient with neurocirculatory asthenia, the thyrotoxic patient wishes to be active but is hampered by fatigue and is tired from the neck down rather than from the top of the head down. Fatigue may be due both to muscle weakness and to the insomnia that is commonly present. In some patients, severe wasting and fatigue impair overall activity. Emotional lability causes patients to lose their temper easily and to have episodes of crying with only slight provocation. In rare cases, mental disturbance may be severe; manic-depressive, schizoid, or paranoid reactions may emerge.

The hypokinesia of the thyrotoxic patient is characteristic. During the interview, the patient cannot sit still, may drum on the table, may tap a foot, or may shift positions frequently. Movements are quick, jerky, exaggerated, and often purposeless. In children, in whom such manifestations tend to be more severe, Sydenham’s chorea may be suggested. Examination also reveals a fine, rhythmic tremor of the hands, tongue, or lightly closed eyelids. With the aid of a magnifying glass, a tremor of the eyeballs may be seen. The tremor may sometimes mimic that of parkinsonism, and coexisting parkinsonism should be considered. The electroencephalogram reveals an increase in fast-wave activity, and in patients with convulsive disorders, the frequency of seizures is increased.

The physiologic basis of these nervous system findings is not well understood; they may reflect increased adrenergic activity because some improvement occurs during treatment with adrenergic antagonists. The widespread distribution of thyroid hormone receptors in the brain makes it likely that alterations in cerebral metabolism are induced by thyroid hormone excess. Nevertheless, oxygen consumption by the brain is not altered.

Muscle

Weakness and fatigability are usually not accompanied by objective evidence of muscle disease except for the generalized wasting associated with loss of weight. Often the weakness is most prominent in the proximal muscles of the limbs, causing difficulty in climbing stairs or in maintaining the leg in an extended position. The latter maneuver can be employed to assess the degree of muscle weakness.

Occasionally, in severe untreated cases, muscle wasting that again tends to be proximal develops out of proportion to the overall loss of weight (thyrotoxic myopathy). In the extreme form, the patient may be unable to rise from a sitting or lying position and may be virtually unable to walk. This disorder may resemble progressive muscular atrophy or polymyositis; however, fasciculation is absent, and little if any inflammatory change is evident on biopsy. Instead, the muscle is atrophic and infiltrated with fat cells and lymphocytes. Electron microscopy reveals abnormal mitochondria and dilatations of the myotubular system. Electromyograms reveal a decreased duration of action potentials and an increased number of polyphasic potentials. The biochemical basis of the muscle weakness is uncertain but may be related to the impaired ability to phosphorylate creatine.

Myopathy affects men with thyrotoxicosis more commonly than women and may overshadow the other manifestations of the syndrome. In the most severe forms, the myopathy may involve the more distal muscles of the extremities and the muscles of the trunk and face. Although myopathy of ocular muscles is unusual, the disorder may mimic myasthenia gravis or the ophthalmic form of myasthenia. Muscular strength returns to normal when a normal metabolic state is restored, but muscle mass takes longer to recover.

Graves’ disease occurs in about 3% to 5% of patients with myasthenia gravis, and myasthenia gravis develops in about 1% of patients with Graves’ disease. Antibodies and T cells specific for receptors of the TSH receptor (TSHR) and the acetylcholine receptor are involved in the pathogenesis of the two diseases. Unlike thyrotoxic myopathy, the association of myasthenia gravis with Graves’ disease has a distinct female preponderance. Although the effect of both thyrotoxicosis and its alleviation on the course of myasthenia gravis is variable, in most instances myasthenia is accentuated during the thyrotoxic state and improves when a normal metabolic state is restored.

Periodic paralysis of the hypokalemic type may occur together with thyrotoxicosis, and its severity is accentuated by the latter disorder. The coincidence of the two disorders is particularly common in Asian and Latino men.

Skeletal System: Calcium and Phosphorus Metabolism

Thyrotoxicosis is generally associated with the following:

1. Increased excretion of calcium and phosphorus in urine and stool.
2. Demineralization of bone, as demonstrated by routine bone densitometry.
3. Occasionally, pathologic fractures, especially in older women.

In these instances, the pathologic changes are variable and may include ostestis fibrosa, osteomalacia, and osteoporosis.

Urinary excretion of collagen breakdown products is increased, indicating increased turnover of collagen. Kinetic studies indicate an increase in the exchangeable calcium pool and acceleration of both bone resorption and accretion, particularly the former. The changes lead to decreased bone density and a propensity to hip fractures in later years. As thyrotoxicosis is treated, bone density may improve. Postmenopausal women, however, may have a permanent reduction in bone density that may require treatment with agents that increase bone density (see Chapter 29). Much controversy has existed over the induction of decreased bone density by thyroid hormone replacement therapy in hypothyroidism and thyroid hormone suppression therapy in patients with thyroid cancer. Suffice it to say that postmenopausal women who receive excessive thyroid hormone are at certain risk of bone damage, but careful replacement therapy does not harm to anyone.

Hypercalcemia can occur in patients with thyrotoxicosis. The total serum calcium concentration is increased in as many as 27% of patients, and the ionized serum calcium level is elevated in 47%. The concentrations of heat-labile serum alkaline phosphatase and osteocalcin may also be elevated. These findings resemble those of primary hyperparathyroidism, but the concentration of immunoreactive parathyroid hormone in serum is decreased in most thyrotoxic patients with
hypercalcemia. True primary hyperparathyroidism and thyrotoxicosis may sometimes coexist. Hypercalcemia may be severe enough to induce anorexia, nausea, vomiting, polyuria, and occasionally impairment of renal function. The alterations in calcium metabolism in thyrotoxicosis may be due to a direct effect of thyroid hormones in stimulating bone resorption and are reversed when the eumetabolic state is restored. Plasma 25-hydroxyvitamin D, (25-hydroxycholecalciferol) levels are decreased in thyrotoxic patients, and this alteration may contribute to the decreased intestinal absorption of calcium and osteomalacia noted in some patients.

Renal Function: Water and Electrolyte Metabolism

Thyrotoxicosis produces no symptoms referable to the urinary tract other than mild polyuria. Nevertheless, renal blood flow, glomerular filtration, and tubular reabsorptive rates are increased. Total amounts of body water and exchangeable potassium are decreased, possibly because of a decrease in lean body mass, but the amount of exchangeable sodium tends to be increased. Serum sodium, potassium, and chloride concentrations are normal. In patients with thyrotoxicosis, the level of exchangeable magnesium is normal; the serum magnesium concentration is often decreased, and urinary magnesium excretion is increased.

Hematopoietic System

Red blood cells are usually normal, as judged by the usual indices, but red blood cell mass is increased. The increase in erythropoiesis appears to be due both to the direct effect of thyroid hormones on the erythroid marrow and to increased production of erythropoietin. A parallel increase in plasma volume also occurs, resulting in a normal hematocrit value. Other abnormalities in thyrotoxicosis include a reduced content of zinc and carbonic anhydrase and an increased content of sodium in red blood cells, probably because the activity of Na⁺-K⁺-ATPase is impaired (in contrast with the increased Na⁺-K⁺-ATPase activity sometimes seen in other tissues).

Approximately 3% of patients with Graves' disease have pernicious anemia, and an additional 3% have antibodies to intrinsic factor but normal absorption of vitamin B₁₂. Autoantibodies against gastric parietal cells may also be present in patients with Graves' disease, and the requirements for vitamin B₁₂ and folic acid appear to be increased. Rarely, thyrotoxicosis is associated with a mild hypocromic anemia characterized by adequate stores of iron in marrow and responsive to large doses of pyridoxine (vitamin B₆).

The total white blood cell count is often low because of a decrease in the number of neutrophils. The absolute lymphocyte count is normal or increased, leading to a relative lymphocytosis. The numbers of monocytes and eosinophils may also be increased. Splenic enlargement occurs in about 10% of the patients, and thymic and lymph node enlargement is common. These abnormalities are thought to be a reflection of the autoimmune aspects of Graves' disease because they do not occur in thyrotoxicosis stemming from other causes.

Platelet levels and the intrinsic clotting mechanism are normal, but factor VIII concentrations are often elevated and return to normal when the thyrotoxicosis is treated. Despite this elevation, there is an enhanced sensitivity to anticoagulants of the coumarin series because of accelerated clearance of the vitamin K-dependent clotting factors. Somewhat paradoxically, the dosage of such anticoagulants may have to be reduced in thyrotoxic patients and is increased in hypothyroid patients. Coincidental autoimmune thrombocytopenia may also occur.

Pituitary and Adrenocortical Function

The thyrotoxic state imposes several challenges on pituitary and adrenocortical function. The inactivation of cortisol is accelerated, including reduction of the A ring, which is rapidly followed by conjugation, and oxidation of the 11-hydroxy group to a keto group as a result of an increase in 11β-hydroxysteroid dehydrogenase (HSD) activity. As a result of these changes the disposal of cortisol is accelerated, but its rate of secretion is also increased, so the plasma cortisol concentration remains normal. The concentration of corticosteroid-binding globulin in plasma is normal. Urinary excretion of free cortisol and 17-hydroxycorticosteroids is normal or slightly increased, whereas urinary excretion of 17-ketosteroids may be reduced.

Basal pituitary-adrenal function is adequate, as indicated by normal plasma cortisol concentrations, and the response to an acute challenge, such as that imposed by insulin-induced hypoglycemia, is adequate. The rate of turnover of aldosterone is increased, but the plasma level is normal. Plasma renin activity is increased, and sensitivity to angiotensin II is reduced.

The response of plasma growth hormone concentration to insulin-induced hypoglycemia is subnormal, particularly in those with severe disease. This observation may not indicate deficient growth hormone production but, rather, may reflect depletion of pituitary stores from caloric inadequacy or accelerated removal of growth hormone from plasma. Incomplete suppression of plasma growth hormone concentration by induced hyperglycemia may also reflect prolonged caloric deprivation.

Reproductive Function

Thyrotoxicosis in early life may cause delayed sexual maturation, although physical development is normal and skeletal growth may be accelerated. Thyrotoxicosis after puberty influences reproductive function, especially in women. An increase in Iβ-hormone sometimes occurs in both men and women. The intermenstrual interval may be prolonged or shortened, and menstrual flow is initially diminished and ultimately ceases. Fertility may be reduced, and if conception takes place, the risk of miscarriage is increased. The association of thyroid autoantibodies and increased pregnancy loss is not related to changes in thyroid function; the thyroid autoantibodies are thought to represent a marker of immune instability predisposing to pregnancy interruption, as seen in an animal model.

In some patients, menstrual cycles are predominantly anovulatory with oligomenorrhea; in most patients, however, ovulation occurs, as indicated by a secretory endometrium. In the former situation, a subnormal midcycle surge of luteinizing hormone (LH) may be responsible. In premenopausal women with thyrotoxicosis, basal plasma concentrations of LH and follicle-stimulating hormone (FSH) are reportedly normal but may display enhanced responsiveness to gonadotropin-releasing hormone (GnRH).

Thyrotoxicosis, whether spontaneous or induced by triiodothyronine (T₃), is accompanied by an increase in the concentration of sex hormone-binding globulin in plasma. As a result, the plasma concentrations of total testosterone, dihydrotestosterone (DHT), and estradiol are increased, but their unbound fractions are normal or transiently decreased. The increased binding in plasma is responsible for the decreased metabolic clearance rate of testosterone and DHT. In the case of estradiol, however, the metabolic clearance rate is normal, suggesting that tissue metabolism of the hormone is increased.

Conversion rates of androstenedione to testosterone, estrone, estradiol, and of testosterone to DHT are increased. The increased rate of conversion of androgens to estrogens may be the mechanism of gynecomastia and erectile dysfunction in about 10% of thyrotoxic men and may be one mechanism of menstrual irregularities in women. Another more likely mechanism of menstrual changes is the disruption in amplitude and frequency of LH/FSH pulses due to thyroid hormone influences on GnRH signaling.

Catecholamines and Serotonin

Many effects induced by thyroid hormones are reminiscent of those induced by epinephrine, including tachycardia, increased cardiac output, and enhanced glycolysis, lipolysis, and calorigenesis (see earlier). Moreover, the fact that some of the manifestations of thyrotoxicosis among them eyelid retraction, tremor, excessive sweating, and tachycardia are at least partly alleviated by adrenergic antagonists has been interpreted as indicating that a state of increased adrenergic activity exists in the thyrotoxic organism. As discussed earlier, however, this apparent adrenergic hyperactivity appears to be a consequence of a direct effect of thyroid hormones on these tissues or due to decreased vagal tone.

Plasma levels of epinephrine and norepinephrine are normal but 24-hour catecholamine secretion may be increased.

Some manifestations of thyrotoxicosis, such as flushing, sweating, tachycardia, and gastrointestinal hypermotility, are also reminiscent of those of carcinoid syndrome. However, plasma serotonin levels, urinary 5-hydroxyindoleacetic acid excretion, and platelet monoamine oxidase activity are normal.
The stimulation of energy metabolism and heat production is reflected in the increased basal metabolic rate, increased appetite, and heat intolerance and in a sometimes slightly elevated basal body temperature. Despite an increased food intake, a state of chronic caloric and nutritional inadequacy often ensues, depending on the degree of increased metabolism. Both synthesis and degradation of protein are increased (the latter to a greater extent than the former), with the result that there is a net decrease in tissue protein, as indicated by negative nitrogen balance, loss of weight, muscle wasting, weakness, and mild hypoalbuminemia. The oral glucose tolerance curve is often abnormal and varies from one in which the peak glycemia is increased and somewhat delayed to one that is frankly diabetic. Plasma insulin concentrations, however, are increased, suggesting insulin resistance. The pathogenesis of these alterations remains to be defined. Preexisting diabetes mellitus is exacerbated by thyrotoxicosis, one cause being increased degradation of insulin.

Both synthesis and degradation of triglycerides and of cholesterol are increased, but the net effect is one of lipid degradation, as reflected by an increase in the plasma concentration of free fatty acids and glycerol and a decrease in the serum cholesterol level; serum triglyceride levels are usually slightly decreased. Postheparin lipolytic activity is reported to be decreased in some studies and increased in others. The enhanced mobilization and oxidation of free fatty acids in response to fasting, catecholamines, and growth hormone are probably due to activation of adenylate cyclase and result in a tendency to ketosis and to fatty infiltration of the liver, depending on the degree of caloric deprivation.
Composite Clinical Picture and Laboratory Tests in Thyrotoxic States

The effects of thyrotoxicosis on the major organ systems are the same, regardless of the underlying cause. Their frequency and intensity and other findings with which they are associated are influenced by the nature of the underlying disorder. To a large extent, the same is true of laboratory test results. Consequently the clinical picture, laboratory features, and differential diagnosis are considered in relation to the specific etiologic mechanisms (see Table 11-1).
ROBERT GRAVES’ DISEASE

Background

Robert Graves’ disease, although first described by Parry in 1825, is best known as Graves’ disease in the English-speaking world and as von Basedow’s disease on the continent of Europe because of the prominence of the disease reports by these eminent physicians. It is the most enigmatic and, in areas of iodine abundance, one of the most common of thyroid diseases.

Presentation

Graves’ disease is characterized by diffuse goiter, thyrotoxicosis, infiltrative orbitopathy and ophthalmopathy, and occasionally infiltrative dermopathy. In the individual patient, thyroid disease and the infiltrative phenomena may occur singly or together but run courses that are largely independent. The thyroid component is closely related to autoimmune thyroiditis (Hashimoto’s disease) in its pathogenesis and clinical course. In Graves’ disease, hyperthyroidism occurs in the presence of some degree of chronic thyroiditis and may ultimately be replaced, in the long term, by thyroid hypofunction. Conversely, hyperthyroidism may occasionally supervene in patients with preexisting Hashimoto’s thyroiditis. Both of these diseases may occur within the same family.

Autoimmune Characteristics

Autoimmune thyroid disease is characterized by the occurrence in the serum of antibodies against thyroid peroxidase (TPO) (the “microsomal” antigen), thyroglobulin (Tg), and the TSHR. T-cell mediated autoimmunity can also be demonstrated against the three primary thyroid antigens, as judged by a variety of criteria, including the ability of the T cells to elaborate various lymphokines and to exhibit a mitogenic response when exposed to thyroid antigens or to peptide sequences from the antigens. Autoimmune thyroid disease is also characterized by lymphocytic infiltration of the thyroid gland or remnant thyroid bed. In patients and their relatives, there is an increased frequency of other disorders of autoimmune origin, such as insulin-dependent diabetes mellitus, pernicious anemia, myasthenia gravis, adrenal atrophy, Sjögren’s syndrome, lupus erythematosus, rheumatoid arthritis, and idiopathic thrombocytopenic purpura [(see Chapter 37)].

Circulating autoantibodies specific to hyperthyroid Graves’ disease are directed against the TSHR (TSHR Abs) and behave as thyroid-stimulating antibodies. These antibodies can compete for the binding of TSH to its specific receptor site in the cell membrane (Fig. 11-1) and can activate adenylate cyclase as TSH agonists (Fig. 11-2). Simlar but distinct autoantibodies in the sera of some patients with autoimmune thyroiditis do not stimulate the thyroid cell and may block the ligand-binding site and act as TSH antagonists [(Fig. 11-3)].

The thyroid gland itself is a site of thyroid autoantibody secretion in autoimmune thyroid disease via the B cells that form part of the intrathyroidal infiltrate. Transplantation of Graves’ thyroid tissue into T-cell-deficient and B-cell-deficient mice with severe combined immunodeficiency (scid) mice results in the appearance of human thyroid autoantibodies, including TSHR Abs, in the serum. Additional evidence for a role of the thyroid itself in antibody production comes from animal models of thyroiditis and from the decline in thyroid autoantibody levels after antithyroid drug treatment, thyroidectomy, or radioiodine ablation. After thyroidectomy and radiiodine treatment, however, some patients show no decline in autoantibody secretion, which suggests extrathyroidal sources of continued production.

Pathology

In patients with Graves’ disease, the thyroid gland is characterized by a nonhomogeneous lymphocytic infiltration with an absence of follicular destruction (Fig. 11-4). Antithyroid drug treatment may reduce the degree of infiltration. Although the intrathyroidal lymphocyte population is mixed, most are T lymphocytes; B-cell germinal centers are less common than in autoimmune thyroiditis. However, both intraepithelial T cells and plasma cells can be seen in peripolesis within the thyroid follicles. Follicular epithelial cell size correlates with the intensity of the local infiltrate, suggesting local thyroid cell stimulation by TSHR Abs. Memory T cells may predominate within the T-cell population, but this finding can vary from patient to patient. Activated B-cell and T-cell markers are more frequent in intrathyroidal lymphocyte cultures than in peripheral blood cultures.

Prevalence

In the United States, the prevalence of Graves’ disease is uncertain but is assumed to be similar to that of the one well-designed epidemiologic survey published. This survey came from Whickham, a small town in the northeast of England (2800 in population), an area thought to be representative of the United Kingdom. The results indicated a prevalence of 2.7%, past and present, in women and a prevalence about one tenth as frequent in men. Overall, the incidence was estimated, in women, to be 1 case per 1000 per year over a 20-year follow-up. Graves’ disease is the most common cause of spontaneous hyperthyroidism in patients younger than 40 years of age, and the hazard rate does not change with age. The overall prevalence of autoimmune thyroid disease, comprising Graves’ disease and autoimmune thyroiditis, approaches or exceeds that of diabetes mellitus (Table 11-3).
Pathogenesis

The Major Antigen of Graves' Disease—the Thyrotropin Receptor

The TSH receptor (TSHR) is G protein-linked with seven transmembrane domains and employs cyclic adenosine monophosphate (cAMP) and the phosphoinositol pathways for signal transduction. The human TSHR (hTSHR) is the primary autoantigen of Graves' disease, as shown by the development of hyperthyroidism in mice and hamsters after immunization.

![Figure 11-2: Stimulation of adenylyl cyclase activity in human thyroid membranes by serum immunoglobulin G in normal control subjects and patients with thyroid disease.](Figure Not Available)

with hTSHR antigen (Fig. 11-5) (Figure Not Available) (see also Color Plate). Putative extrathyroidal TSHR messenger RNA (mRNA) and receptor protein have been reported in many other tissues, including extraorbital adipocytes, muscle cells, lymphocytes, and fibroblasts, but the physiologic and pathologic role of TSH receptors in these sites is still under investigation.

Molecular Biology of the Human Thyrotropin Receptor

Cloning of TSHR complementary DNA (cDNA) of animals and humans made it possible to define the structure of the hTSHR gene and its chromosomal location (14q31). The hTSHR gene spans more than 60 kb and is split into 10 exons. Seven hydrophobic transmembrane spanning regions in the hTSHR indicate that it is a member of the G protein-coupled receptor gene superfamily, and those receptors with large extracellular domains have been designated subgroup B. The hTSHR-specific mRNA of human thyroid consists of a major 4.3-kb transcript and additional smaller species, indicating that mRNA undergoes alternate splicing.

Protein Structure of the Human Thyrotropin Receptor

The hTSHR holoreceptor consists of a 100-kd, glycosylated, 744-amino acid sequence and a 20-amino acid signal peptide. ProTSHR is cleaved into two subunits, (or A) and (or B), which are linked by disulfide bonds to form the physiologic receptor. The 50-kd subunit is water-soluble and may contain the TSH-binding site, previously referred to as long-acting thyroid stimulator (LATS)-absorbing activity (LAA). The 30-kd subunit is water-insoluble, contains the membrane spanning domain, with its three extracellular loops and three cytoplasmic loops, and is 70% to 75% homologous with the LH/human chorionic gonadotropin (hCG) receptor. Shedding of the subunit has been suggested in vitro. The TSHR forms dimers and multimeric complexes on the thyroid cell surface, and these are of unclear physiologic significance.

Autoantibodies to the Thyrotropin Receptor

In Graves' disease, TSHR Abs bind to the TSH receptor, activate adenylate cyclase, induce thyroid growth, increase vascularity, and cause an increased rate of thyroid hormone production and secretion. TSHR Abs in patients with Graves' disease are referred to as stimulating or agonist types of TSHR Abs, as first described by Adams and Purves. Other varieties of TSHR Abs may also be present, namely a receptor antibody that acts as a TSH antagonist and referred to as blocking TSHR Abs or a neutral form of antibody with no functional effect on the receptor. Blocking TSHR Abs may be coincident with the stimulating type and may also predominate in certain patients after treatment with radioactive iodine, antithyroid drugs, or surgery. Blocking TSHR Abs can also be found in 15% of patients with autoimmune thyroiditis, particularly in patients without a goiter (the atrophic variety). TSHR Abs are not detectable in the normal population with the use of currently available methods.

Immunoreactivity of Thyrotropin Receptor Autoantibody

The self-infusion of sera from patients with Graves' disease caused thyroid stimulation and was the first demonstration of the role of TSHR Abs in the induction of human hyperthyroidism. Another example of the in vivo effects of TSHR Abs came from studies in neonates demonstrating the transplacental stimulation of the fetal thyroid in mothers with high titers of TSHR Abs. TSHR Abs show light chain restriction in many patients with Graves' disease, and TSHR Abs that exhibit TSH agonist bioactivity are in the immunoglobulin G1 (IgG1) subclass; both observations suggest oligoclonality.

As discussed earlier, autoantibodies that bind to the TSH may or may not activate adenylate cyclase and may thus be either TSH-stimulating or TSH-blocking (see Fig. 11-3). Further complicating this issue is the fact that many patients have both TSH-stimulating and TSH-blocking antibodies, the degree of thyroid...
stimulation depends on the relative concentration and bioactivity of the different autoantibodies.

Prevalence of Thyrotropin Receptor Autoantibodies in Graves' Disease

The fact that TSHRABs are detectable only in patients with autoimmune thyroid disease indicates that the autoantibodies are disease-specific, in contrast with the high prevalence of Tg antibodies and TPO antibodies in the normal population. Furthermore, TSHRABs are unique human autoantibodies and do not occur in natural animal disease. A total of 80% to 100% of untreated hyperthyroid patients with Graves’ disease have detectable TSHRABs with thyroid-stimulating activity. The levels of TSHRABs are decreased by treatment of the disease and, when they persist, are predictive of failure of response to antithyroid drug treatment. With time, TSHRAB-blocking auto-antibodies may become the prevalent type after treatment of Graves’ disease.

Figure 11-5 (Figure Not Available) Thyroid glands from hyperthyroid mice immunized with fibroblasts expressing the human thyrotropin-stimulating hormone receptor. Left, A thyroid gland from a control mouse. Right, An enlarged thyroid from a hyperthyroid mouse. Ruler is in centimeters. (From Kita ML, Ahmad RC, Marians H, et al. Regulation and transfer of a murine model of thyrotropin receptor antibodymediated Graves’ disease. Endocrinology 1999; 140:13921398.)

TABLE 11-3 – Incidence of Thyroid Dysfunction.

<table>
<thead>
<tr>
<th>Status</th>
<th>Females</th>
<th>Males</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothyroidism</td>
<td>4.1</td>
<td>0.6</td>
</tr>
<tr>
<td>Hyperthyroidism</td>
<td>0.8</td>
<td>&lt;0.1</td>
</tr>
</tbody>
</table>

*Per 1000 population per year over a 20-year period.

Antibodies interact with either linear (conformationally independent) or nonlinear (conformationally dependent) areas of a target molecule called the epitope. The strength of the binding of antibodies to target antigen (affinity of the antibody) depends largely on the number of binding sites the antibody has with the antigen. These sites contribute to the binding energy, which is therefore likely to be greater for large nonlinear epitopes than for small linear peptides. High-affinity antibodies are thus most likely to interact with nonlinear epitopes and to depend on the natural conformation of the target molecule.

The T-cell receptor is also a member of the Ig receptor superfamily but interacts only with a complex of antigen and HLA molecule. The CD8 T cells recognize target antigen complexed with HLA class I molecules (A, B, and C), and CD4 T cells recognize antigen complexed with HLA class II molecules (D, P, and Q). The antigen that interacts with T cells when complexed with autologous HLA molecules is a 15amino acid linear peptide derived from the whole antigen molecule.

Thyroid antigens are endocytosed by antigen-presenting cells (APCs), such as macrophages and dendritic cells, and secretory state.

Principles of Antigenic Recognition

Antibodies interact with either linear (conformationally independent) or nonlinear (conformationally dependent) areas of a target molecule called the epitope. The strength of the binding of antibodies to target antigen (affinity of the antibody) depends largely on the number of binding sites the antibody has with the antigen. These sites contribute to the binding energy, which is therefore likely to be greater for large nonlinear epitopes than for small linear peptides. High-affinity antibodies are thus most likely to interact with nonlinear epitopes and to depend on the natural conformation of the target molecule.

The T-cell receptor is also a member of the Ig receptor superfamily but interacts only with a complex of antigen and HLA molecule. The CD8 T cells recognize target antigen complexed with HLA class I molecules (A, B, and C), and CD4 T cells recognize antigen complexed with HLA class II molecules (D, P, and Q). The antigen that interacts with T cells when complexed with autologous HLA molecules is a 15amino acid linear peptide derived from the whole antigen molecule. Thyroid antigens are endocytosed by antigen-presenting cells (APCs), such as macrophages and dendritic cells; peptide breakdown products are then bound to HLA molecules, and the entire complex is transported to the cell surface where the peptide lies within a clearly defined binding groove within the HLA molecule.

In addition to the recognition of antigen (the first immune cell signal), immune cells (both T and B) also depend on secondary signals to enter an active proliferative and secretory state. Cytokines originating from T cells serve as secondary B-cell signals. Important second signals for the T cells include the CD80 (B7) family of molecules, and the entire complex is transported to the cell surface where the peptide lies within a clearly defined binding groove within the HLA molecule.

Figure 11-6 Human thyrotropin receptor (hTSH-R) exon structure. Outline structure of the TSH-R (A) in comparison with the porcine luteinizing hormone receptor (pLH-R) and rat follicle-stimulating hormone receptor (fFSH-R) genes and (B) the exon/intron organization of the TSH-R gene. A shows the similarity (%) between the hTSH-R and the pLH and rFSH receptors, respectively. (From Gross B, Mirahl M, Sar S, et al. Composite structure of the human TSH receptor gene. Biochem Biophys Res Comm 1991; 177:679687.)

Thyrotropin Receptor Autoantibody Epitopes

The extracellular domain of the TSHR is the major site of TSHRAB binding (see Fig. 11-7). The difference in functional activity of different TSHRABs also depends on receptor conformation and affinity. The use of recombinant chimeric TSH/LH receptors has suggested at least two major regions of TSHRAB binding that may convey thyroid-stimulating activity (the N-terminal end) and thyroid-blocking activity (the C-terminal end); however, the situation is probably more complex than this. Prokaryotic TSHR extracellular domain has also been used to identify immunogenic regions using antibodies from immunized mice and from patients with Graves’ disease. These data suggested that some TSHRABs may also recognize linear epitopes, but these may be neutral rather than pathologic antibodies.

Control of Thyrotropin Receptor Function in Graves’ Disease

Like TSH, TSHRABs cause cAMP-mediated generation of thyroid hormone and Tg, uptake and release of iodine, stimulation of protein synthesis, and cell growth.

Desensitization of the thyroidal cAMP response by prolonged exposure to TSHRABs can be observed in vitro but cannot be complete in vivo, or else patients would not remain hyperthyroid.

At lower levels of stimulation, the TSHR is positively regulated by TSH both in vivo and in vitro, and such resistance to desensitization by low levels of TSH may allow the hyperthyroid state to persist.

Principles of Antigenic Recognition

Antibodies interact with either linear (conformationally independent) or nonlinear (conformationally dependent) areas of a target molecule called the epitope. The strength of the binding of antibodies to target antigen (affinity of the antibody) depends largely on the number of binding sites the antibody has with the antigen. These sites contribute to the binding energy, which is therefore likely to be greater for large nonlinear epitopes than for small linear peptides. High-affinity antibodies are thus most likely to interact with nonlinear epitopes and to depend on the natural conformation of the target molecule.

The T-cell receptor is also a member of the Ig receptor superfamily but interacts only with a complex of antigen and HLA molecule. The CD8 T cells recognize target antigen complexed with HLA class I molecules (A, B, and C), and CD4 T cells recognize antigen complexed with HLA class II molecules (D, P, and Q). The antigen that interacts with T cells when complexed with autologous HLA molecules is a 15amino acid linear peptide derived from the whole antigen molecule. Thyroid antigens are endocytosed by antigen-presenting cells (APCs), such as macrophages and dendritic cells; peptide breakdown products are then bound to HLA molecules, and the entire complex is transported to the cell surface where the peptide lies within a clearly defined binding groove within the HLA molecule.

In addition to the recognition of antigen (the first immune cell signal), immune cells (both T and B) also depend on secondary signals to enter an active proliferative and secretory state. Cytokines originating from T cells serve as secondary B-cell signals. Important second signals for the T cells include the CD80 (B7) family of cell surface molecules found on APCs (e.g., macrophages and dendritic cells) that combine with CD28 on the T-cell surface and the CD40/CD40 ligand interaction, where CD40 is found more often on nonprofessional APCs (e.g., fibroblasts and thyroid cells). B cells and T cells that interact with specific antigen in the absence of a second signal may be deleted by apoptosis or may survive in a state of anergy (i.e., desensitized). Hence,

Data shown as 3H-thymidine uptake at 18 hours.
deletion and anergy combine to help define the repertoire of immune responses in an individual.

The Autoimmune Response

It is helpful to know whether autoimmune reactions are multireactive (representative of a secondary polyclonal immune response) or more focused, involving a restricted number of B cells and T cells. In autoimmune disease, a restricted immune response has been observed at the onset of the disease. The following discussion presents some of the evidence for this in Graves' disease.

Intrathyroidal T Cells

As described earlier, T cells in patients with autoimmune thyroid disease are reactive to thyroid antigens and to peptides derived from these antigens. About 10% of activated T cells infiltrating the thyroid gland in patients with autoimmune thyroid disease proliferate in response to thyroid cell antigens. Intrathyroidal T cells from patients with Graves' disease exhibit characteristics of helper T-cell subset 1 (Th1) (which secrete interleukin-2 [IL-2] and interferon) and Th2 (which secrete IL-4). Such T-cell populations enhance thyroid autoantibody secretion. Memory T cells enhance thyroid autoantibody secretion and may have both helper and cytotoxic (apoptosis-inducing) T-cell activity. The thyroid gland in patients with Graves' disease exhibits a small degree of apoptosis.

Bystander Activation

T-cell receptors (TCRs) consist of two noncovalently linked chains ( and ), each with variable (V), diversity (D, mainly), and junctional (J) regions. The V, D, and J genes code for the antigen recognition sites on the TCR that determine antigen specificity. In addition to the many V (>100) and J (>50) genes, random nucleotide (N) additions and deletions to the D region (the CDR3 region or DND region) provide additional complexity to the TCR repertoire. Studies of the V and V families expressed by intrathyroidal T cells from patients with Graves' disease have shown that human TCR V gene expression by T cells from within the thyroid gland differed from hTCR V gene expression found in peripheral blood from the same individuals. Additional evidence has been obtained in the thyroid gland for the presence of clonally expanded T cells based on the presence of multiple identical sequences within the generated fragments. Such information further documents that Graves' disease is an antigen-driven disorder that causes oligoclonal T-cell expansion. Similar data have been obtained in rheumatoid arthritis, multiple sclerosis, and other autoimmune diseases.

Regulatory T Cells

The presence of reduced levels of circulating CD8 (suppressor/cytotoxic) T cells in patients with Graves' disease suggested that a lack of suppressor/cytotoxic T cells might be responsible for the breakdown of tolerance in Graves' disease. However, thyroid hormone excess itself may give rise to changes in T-cell numbers. The immune system does exert some of its overall control via "regulatory" cells, including the secretion of T-cell cytokines and the suppressive influence of "anergized" T cells. In addition to these regulatory cells, other important mechanisms include both positive and negative selection of T cells and B cells in the thymus, and deletion of immature immune cells via antigen-initiated apoptosis, which takes place in the peripheral immune system. Deletion probably occurs when T cells and B cells see antigen in the absence of second signals (see earlier). When a mature immune cell sees antigen in the absence of second signal, it may become desensitized rather than deleted, a phenomenon known as anergy. Together, deletion, anergy, and apoptosis account for immunosuppression. However, certain HLA-DR haplotypes are associated with reduced suppressor T-cell function. For example, normal individuals with HLA-DR3 have reduced suppressor activity compared with non-DR3 individuals.

The Insult

Initiation is thought to occur with an insulin that leads to an immune response. This may take the form of a direct insult to the thyroid gland by a viral infection or another external influence, including trauma leading to activation of T cells or may be initiated elsewhere in the body. In the latter case, it would be the arrival of activated T cells in the thyroid gland that would start the process. Such an arrival may be nonspecific because the same T cells may arrive in many glands but the patient has a particular susceptibility to autoimmune thyroid disease (see "Risk Factors" later).

Bystander Activation

Evidence has mounted from a model of insulitis that bystander activation of local resident antigen-specific T cells may initiate autoimmunity. The presence of activated T cells within the thyroid gland following an insult may induce, via cytokine secretion, the activation of local thyroid-specific T cells. This series of events can occur only in a susceptible individual with the right immune repertoire. Bystander activation would arise from any activated T cells within the thyroid gland, which may be activated by many different infections and antigens unrelated to the thyroid gland itself. The attractiveness of this model is that many different types of infections would lead to the same clinical disease phenotype. There is much evidence for residual thyroid-resident T cells in the glands of patients with Graves' disease.

Molecular Memory (Specificity Crossover)

In addition to the effects of the direct release of cytokines from T cells activated elsewhere via the bystander effect, intrathyroidal T cells may become activated in another nonspecific way. Structural similarity between different antigens can lead to specificity crossover (or molecular mimicry). Antigenic similarity between bacteria...
and viruses and human proteins is common, and in one study 4% of monoclonal antibodies raised against a variety of viruses cross-reacted with antigens in tissues. Furthermore, mice infected with reovirus type 1 developed an autoimmune polyendocrinopathy with autoantibodies directed against normal pancreas, pituitary, thyroid, and gastric mucosa, suggesting molecular mimicry between a reoviral antigen and a common tissue antigen. Molecular mimicry has also been reported between Yersinia enterocolitica and the TSHR based on the observed cross-reaction between sera from patients with Yersinia infection and sera from patients with Graves’ disease and between retroviral sequences and the TSH receptor.

**Thyroid Cell Involvement by Aberrant Expression of Class II HLA Antigens**

Normal thyroid epithelial cells do not express HLA class II antigens, but they are expressed in thyroid glands from patients with autoimmune thyroid disease (Fig. 11.11) (see also Color Plate). As proposed by Hanafusa and colleagues, a local insult, such as a viral infection of the thyroid gland, can cause production of interferon or other cytokines in the thyroid gland, which in turn would induce HLA class II expression. This first-time expression would lead to enhanced presentation of thyroid autoantigens to the immune system and activation of local autoreactive thyroid-specific T cells in a susceptible individual. Support for this concept comes from the in vivo induction of MHC class II molecules on mouse thyrocytes by interferon that also induced autoimmune thyroiditis and the demonstration of the necessity for MHC class II antigen expression on TSH receptor expressing fibroblasts used in the induction of Graves’ disease in mice. A number of viruses may also induce MHC class II expression independent of immune cell cytokine secretion, including reovirus types 1 and 3 and cytomegalovirus.

**Cryptic Antigenic Epitopes**

T-cell tolerance depends on the visualization of self-antigens in sufficient amounts to initiate continuous T-cell deletion and anergy induction. However, many molecules are not seen in sufficient concentrations to cause the removal of T cells that may react to them. These antigens contain what are sometimes called cryptic epitopes. Hence, T cells specific for these cryptic epitopes may be present in normal immune repertoires. They may then induce autoaggressive T cells if such an epitope is uncovered or increased in concentration by a local insult. HLA class II antigen expression in a situation where it normally does not occur, such as the thyroid epithelial cell, would then allow the presentation of these normally cryptic thyroid antigens to local autoreactive T cells if they are present.
Risk Factors for Graves’ Disease

Genetic Susceptibility

The development and the subsequent course of Graves’ disease are greatly influenced by heredity. The role of heredity

factors is evidenced by the increased incidence in members of patients’ families of other autoimmune disorders, such as Hashimoto’s disease, insulin-dependent diabetes (type 1) or pernicious anemia, and of autoantibodies against endocrine tissues, gastric parietal cells, and intrinsic factor. Indeed, the propensity for development of thyroid autoantibodies appears to be an autosomal dominant trait linked to the CTLA4 gene that codes for a modulator of the second signal to T cells, as discussed earlier. In addition, monozygotic twins have a higher concordance rate of Graves’ disease than do dizygotic twins, and this is true despite the rearrangement of B-cell and T-cell V genes that cause the immune repertoires of identical twins to differ. (see also Chapter 37).

Hence, Graves’ disease appears to be an oligogenic disorder with a number of genetic loci that may contribute to disease susceptibility. However, no thyroid-specific genes have been found to be involved, and only chromosomal loci, not actual genes, have been linked to the disease. There is an increased frequency of the HLA DR3 and DQA10501 haplotypes in whites with Graves’ disease, although the HLA region provides less than 5% of the genetic susceptibility because it is associated but not linked. (133 134 135)

Infection

It is not known whether a specific infection initiates Graves’ disease. If infection were the cause of Graves’ disease, an identifiable agent should be present in most patients, and transfer of the agent to susceptible recipients should transfer the disease. It has been suggested that Graves’ disease is “associated” with infectious agents (e.g., Y. enterocolitica), but no studies meet the necessary criteria to prove this. Infections of the thyroid gland itself (e.g., subacute thyroiditis, congenital rubella) are associated with autoimmunity. (136) Nevertheless, a causative role of infectious agents has not been definitively demonstrated in Graves’ disease, although thyroid disease can be induced in experimental animals by certain viral infections. (137 138) Reports of retroviral sequences in the thyroid glands of patients with Graves’ disease have not been confirmed. (139 140 141)

Stress

Graves’ disease commonly appears to become evident either after severe emotional stress, such as the actual or threatened separation from a loved one, or after an acute fright, such as an automobile accident. There are, in fact, many clinical experiences and reports associating major stress with the onset of Graves’ disease, including data on the high incidence of thyrotoxicosis among refugees from Nazi prison camps. Some data suggest that stress induces an overall state of immune suppression by nonantigen-specific mechanisms, perhaps secondary to the effects of cortisol and corticotropin-releasing hormone action at the level of the immune cell. Furthermore, more patients with Graves’ disease are said to give a history of major stress in the 12 months before disease onset compared with control groups.

Following the acute immune suppression by stress, there is presumably an overcompensation by the immune system when the suppression is released. This would then precipitate autoimmune thyroid disease, as in the postpartum period during which Graves’ disease may occur 3 to 9 months after delivery. The rebound phenomenon would result in greater immune activity than normal and would precipitate disease if the individual were genetically susceptible.

Gender and Gonadal Steroids

Graves’ disease is more common in women than in men (7 to 10:1) and tends to become more prevalent after puberty. The female preponderance and the fact that the disorder is uncommon before puberty have suggested that gonadal steroids may be responsible for this difference. Indeed, androgens may suppress experimental autoimmune thyroiditis. Estrogen has been shown to influence the immune system, particularly the B-cell repertoire, and may be a reason for this susceptibility. However, Graves’ disease continues to occur after the menopause and is still seen in men. In fact, when the disease develops in men, it tends to occur at a later age, to be more severe, and to be accompanied more often by ophthalmopathy. Such observations have suggested that perhaps it is the X chromosome rather than sex steroids that is the responsible element in female susceptibility. Women have two X chromosomes and, therefore, would receive twice the gene dose. A locus on the X chromosome has indeed been linked to Graves’ disease, but the gene responsible has not yet been located. The phenomenon of X-gene inactivation has also been invoked in autoimmune disease. Female cells may inactivate different X chromosomes, leading to potentially differing gene expression between somatic cells and immune cells.

Pregnancy

Graves’ disease is uncommon during pregnancy because hyperthyroidism is associated with reduced fertility and increased pregnancy loss. In addition, pregnancy is a time of immunosuppression, so that the disease tends to improve as pregnancy progresses. Both T-cell and B-cell functions are diminished, and the rebound from this immunosuppression may contribute to the development of postpartum thyroid disease. As many as 30% of young women give a history of pregnancy in the 12 months before the onset of Graves’ disease, indicating that postpartum Graves’ disease is a surprisingly common presentation and that pregnancy is a major risk factor in susceptible women.

Iodine and Drugs

Iodine and iodine-containing drugs, such as amiodarone, may precipitate Graves’ disease or its recurrence in a susceptible individual. Iodine is most likely to precipitate thyrotoxicosis in an iodine-deficient population simply by allowing TSHRAs to be effective in stimulating more thyroid hormone to be formed. Whether there is any other precipitating event is unclear. Iodine may also damage thyroid cells directly and release thyroid antigens to the immune system.

Irradiation

There is no evidence that radiation exposure itself is a risk factor for Graves’ disease. However, there is evidence that thyroid autoantibodies are more prevalent in the radiation-exposed population and thus this fact should be considered. In addition, radioactive iodine (RAI) treatment may cause the onset or worsening of clinical ophthalmopathy, but often this is transient (see later).
Pathogenesis of Graves’ Orbitopathy and Dermopathy

The pathogenesis of the orbitopathy and dermopathy is now better understood than ever before. The extraocular muscle and adipose tissue are swollen by the accumulation in the extracellular matrix of glycosaminoglycans that are secreted by fibroblasts under the influence of cytokines such as interferon from local lymphocytes (Fig. 11-12 and Fig. 11-13) (see also Color Plate). This accumulation disrupts and impairs the function of muscle. As the disease runs its course and inflammation decreases, the damaged muscles become fibrosed. Hence, histologic examination shows a patchy muscle infiltrate predominantly of T cells, and some muscle cells exhibit HLA class II antigen as seen within the thyroid gland. Such T cells react in vitro with retro-orbital tissue. The antigen to which they react may be the TSH receptor itself.

The TSHR mRNA and protein are expressed in fibroblasts and adipocytes and in many other cells. However, retro-orbital tissues seem to express more TSHR than other sites, and/or the TSHR-specific T cells have a propensity for the retro-orbital tissues. Whatever the mechanism, the current working hypothesis is that the immune system recognizes an antigen common to the thyroid gland and retro-orbital tissues (and skin) and is likely to be the TSHR itself, the main antigen of Graves’ disease. Furthermore, patients with the most severe orbitopathy have the highest titers of TSHR Abs, and the level of TSHR Abs correlates with the severity of the eye disease. There is currently no convincing evidence that specific antibodies against orbital tissue contents play a primary pathogenic role. More likely, antigen-specific T cells have the major role in initiating the disorder. However, such antibodies may serve as markers of extraocular muscle inflammation.
Risk Factors in Ophthalmic Graves’ Disease

There is no evidence that a separate and distinct genetic risk can be ascribed to severe ophthalmic Graves’ disease, suggesting that it is mainly environmental factors that lead to the enhanced retro-orbital inflammation in some patients. All of the same risk factors (e.g., infection, stress, gender and gonadal steroids, pregnancy, and drugs) apply to the onset of both thyroid and eye involvement in Graves’ disease.

For example, it is well known that eye disease in men may be worse than expected. However, there are two distinct risk factors that deserve attention. The first is smoking, which may increase the risk for ophthalmic involvement in many studies, perhaps by causing anoxia or simply direct inflammation. The second is radioiodine, which in controlled clinical trials accentuates ophthalmic Graves’ disease. However, this worsening is typically mild and transient and can be ameliorated with corticosteroid treatment for the subsequent 3 to 4 months. For this reason, some physicians are reluctant to prescribe radioiodine to patients with severe eye disease unless the patients are receiving corticosteroids.
Natural History and Course of Graves' Disease

The course of the thyrotoxic component of Graves' disease is variable and often erratic. In some patients, thyrotoxicosis persists, although it may vary in severity. In others, the course may be cyclic, exhibiting remissions of varying frequency, intensity, and duration. This cyclic feature has an important bearing on treatment. With the passage of months or years, thyrotoxicosis tends to give way to euthyroidism. Approximately one third of patients become hypothyroid within 20 years of treatment with antithyroid agents.

The orbitopathy may or may not commence together with the thyrotoxic component. Thus, thyrotoxic patients may initially be free from eye disease but are affected by it months or years later or not at all. Conversely, Graves' disease may begin with orbitopathy and only later, if at all, be associated with thyrotoxicosis. In euthyroid patients with orbitopathy, so-called euthyroid Graves' disease, evidence of a thyroid abnormality, as judged from thyroid function tests and tests for TSHR Abs and other thyroid autoantibodies is common. Some such patients become hypothyroid within a few years, some become hyperthyroid, and a few remain euthyroid. Many euthyroid patients do have evidence of chronic thyroiditis. The course of thyroid function in many of these patients is therefore unpredictable.

Figure 11-14 Chronic pretibial myxedema in a patient with Graves' disease and orbitopathy. The lesions are firm and nonpitting, with a clear edge to feel. (Courtesy of Dr. Andrew Werner, New York, N.Y.)
Histopathology

Thyroid Gland

The older designation for Graves’ disease, diffuse toxic goiter, denoted that the gland was both enlarged and uniformly affected. The gland might vary in consistency from softer than normal to firm and rubbery. The outer surface is usually smooth but may be somewhat lobular; rarely, the gland is grossly nodular prior to treatment. The cut surface is red and glistening. Microscopically, the follicles are small and lined with hyperplastic columnar epithelium and contain scant colloid that displays much marginal scalloping and vacuolization (see Fig. 11-4). Nuclei are vesicular and basally located and exhibit occasional mitoses. Papillary projections of the hyperplastic epithelium extend into the lumina of the follicles. Vascularity is increased, and there is a varying infiltration by lymphocytes and plasma cells that collect in aggregates and may form germinal centers. In such regions, thyroid epithelial cells express HLA class II antigens not seen in normal thyroid glands and are large, perhaps due to local stimulation by TSHR Abs (see Fig. 11-11).

When the patient is given iodine or antithyroid drugs, the thyroid gland may undergo involution if TSHR Abs decrease. Then hyperplasia and vascularity regress, papillary projections recede, and follicles enlarge and become filled with colloid once again.

Eyes

In patients with infiltrative orbitopathy, the volume of orbital contents is increased because of an increase both in retrobulbar connective tissue and in extracellular muscle mass (see Fig. 11-11). Some of the increase in connective tissue is due to edema resulting from accumulation in the ground substance of hyaluronic acid and chondroitin sulfates, which are hydrophilic. The extraocular muscles are swollen, and some fibers exhibit loss of striation, fragmentation, and lymphocytic infiltration. The lacrimal glands may also be involved. Ultimately, the tissues fibrose.

Skin

Dermopathy (Fig. 11-14) (see also Color Plate) is usually a late manifestation, and 99% of patients with infiltrative dermopathy have Graves’ orbitopathy. The content of hyaluronic acid and chondroitin sulfates in the dermis is increased, presumably by lymphokine activation of fibroblasts, causing compression of the dermal lymphatics and nonpitting edema; the collagen fibers are separated and fragmented, and early lesions contain lymphocytic infiltrate. TSH receptor expression can be demonstrated in fibroblasts and adipocytes, and TSHR Abs are high. Nodule and plaque formation may occur in chronic lesions. The cause of the characteristic location of the dermopathy is unclear but, presumably, depends on trauma to the exposed areas.
Pathophysiology

In patients with Graves’ disease, all aspects of thyroid hormone economy are abnormal, including disruption of the regulatory control of thyroid function; alterations in thyroid function itself; changes in the concentration, binding, and metabolism of thyroid hormones; and manifestations of thyroid hormone excess in the peripheral tissues. Abnormalities in these parameters also occur in other forms of thyrotoxicosis but may differ in kind or amount.

An abnormality or override of normal regulatory control is inherent in all forms of thyrotoxicosis, as illustrated by the reemergence of TSH secretion when thyrotoxicosis is relieved. In Graves’ disease, regulatory mechanisms are overridden by the action of TSHR autoantibodies of the stimulating variety. The resulting hyperfunction of the thyroid gland leads to suppression of TSH secretion that is reflected in a suppressed or an undetectable serum TSH level. The basal TSH level may also be suppressed in patients with euthyroid Graves’ disease (indicating the presence of mild excess of thyroid hormone) and in patients in apparent remission, indicating that thyroid hormone excess is not necessarily associated with clinical thyrotoxicosis.

In this context, the term functional autonomy is often misused when the intent is to imply that thyroid function is independent of TSH stimulation. True functional autonomy occurs when the thyroid gland is capable of functioning at a normal or an increased pace in the absence of both TSH and any other circulating thyroid stimulator. Defined in this way, functional autonomy occurs with toxic multinodular goiter and toxic adenoma but not in Graves’ disease. In Graves’ disease, the thyroid gland is controlled by an abnormal stimulator, the TSHRAbs (as in molar pregnancy, in which hCG is responsible). When that stimulator is withdrawn (i.e., when the disease enters remission), hyperfunction subsides, and the nonautonomous nature of thyroid function becomes evident in the reemergence of normal TSH secretion and control of thyroid function.

The disturbance of thyroid function in Graves’ disease leads to hypersecretion of thyroid hormones. Thyroid avidity for iodine is enhanced, and thyroid iodide clearance is thus increased from its normal rate of 6 to 7 mL/min to 2 L/min in the most severe cases. The increase in iodide clearance rate reflects enhanced thyroid blood flow, even if extraction of iodine is assumed to be complete, and hypervascularity, most likely mediated by local angiogenic factors secreted by the thyroid cells. The enhanced thyroid iodide clearance rate is also due to an increase in the overall glandular mass and its unit functional activity, so that iodide transport and probably organic binding are enhanced. The increase in iodide transport is partly responsible for the enhanced susceptibility of the hyperthyroid gland to the inhibitory effects of iodide on organic binding reactions.

As judged from the normal ratio of iodotyrosines to iodothyronines, the rate of the coupling reaction must also be increased. The molar ratio of T3 to thyroxine (T4) in Tg is about twice normal, and this increase in T3 production cannot be ascribed to intrathyroidal iodine deficiency because the iodine content of Tg and the number of T3 residues per molecule are normal. It may simply reflect chronic hyperstimulation of the gland. The rates of turnover and release of the glandular iodine pool are increased, often greatly so. The major product of glandular secretion is T4 but the ratio of T3 to T4 in the thyroid secretion is increased in proportion to the overproduction of T3. In some instances, T3 appears to be the major secretory product, so that the serum T3 level is increased while serum T4 concentration is within the normal range (T3 toxicity).

The proportion of total plasma T4 and T3 in the free or unbound state is increased both because of a slight decrease in concentration of thyroxine-binding globulin (TBG) and because of the increase in the concentration of Tg. The fractional turnover rates of T3 and T4 are increased, leading to increased amounts of hormone in the peripheral pool and to an increase in total daily turnover of T3 and T4. In severe cases, this rate may increase from the normal of about 100 nmol of T3 and 50 nmol of T4 daily to more than 600 nmol daily for both hormones. The total daily disposal of T4 is disproportionately increased relative to that of T3, indicating that the production rate of T4 is disproportionately increased, probably owing to both a preferential increase in thyroid secretion of T3 and increased peripheral conversion of T4 to T3 by D1.
Clinical Picture

Graves' disease is most common in the third and fourth decades of life, is rare before age 10 years, and occurs in the elderly, sometimes in an apathetic form. The features include diffuse goiter, thyrotoxicosis, infiltrative orbitopathy, and occasionally infiltrative dermopathy. Because the orbitopathy and dermopathy may be independent of other manifestations, they are discussed separately. In other respects, the symptoms and signs of thyrotoxicosis are the same in patients with other causes of hyperthyroidism (see Table 11-3).

In most patients, the thyroid gland is enlarged; hyperthyroidism in Graves' disease occurs in a gland of normal size in only a small fraction of patients, though in the elderly goiter may be absent in 20%. The size of the thyroid gland is most often two or three times normal but may be massively enlarged. The consistency varies from soft to firm and rubbery. The enlargement is usually symmetrical. The surface is generally smooth but may feel lobular. In severe cases, a thrill may be felt, usually over the upper poles, and a thrill is always accompanied by an audible bruit. The thrills and bruits due to increased blood flow and are usually continuous but sometimes are present only in systole. The bruit is most easily detected at the upper or lower poles and should not be confused with a venous hum or murmur arising from the base of the heart. Mitral valve prolapse is more common than in the normal population and may account for a cardiac murmur. A thrill or bruit is suggestive of hyperthyroidism due to Graves' disease.

Manifestations of Infiltrative Orbitopathy and Dermopathy

Signs and Symptoms

Spasm and retraction of the eyelids lead to widening of the palpebral fissures so that the sclera are exposed above the superior margin of the limbus (Fig. 11-15). Lid retraction may be asymmetrical. When the patient looks downward, the upper lid lags behind the globe, exposing more sclera. When the patient gazes upward, the globe is displaced posteriorly by pressure from the thumb. The movements of the lids are jerky and spasmodic, and the tightly closed lids may show a tremor. Simple lid retraction and globe and lid lag are often a manifestation of the thyrotoxicosis per se. These manifestations will often abate when the thyrotoxicosis is relieved. On the other hand, significant swelling and inflammation of the eyelids and conjunctiva, and orbital contents, infiltrative orbitopathy, or ophthalmomonympathy may occur.

The disease symptoms and signs of infiltrative orbitopathy may appear in varying combinations. Early symptoms and signs include a sense of irritation in the eyes, resembling that caused by a foreign body, and excessive tearing that is often made worse by exposure to air or wind, especially if exophthalmos is present. The conjunctiva may be injected. Exophthalmos is frequently asymmetrical and may cause a feeling of pressure behind the globes. When exophthalmos is pronounced, the eyes may not close during sleep, a condition termed lagophthalmos. Exophthalmos may be masked by periorbital edema, which is a common accompaniment and source of complaint. Patients frequently describe blurred vision and easy tiring of the eyes. Double vision may occur in combination with the foregoing symptoms or alone. In severe cases, color vision, and then visual acuity, may be decreased or lost and the corneas may ulcerate or become infected (Fig. 11-16G).

Infiltrative orbitopathy may follow an independent course from the thyrotoxic manifestations and is often uninfluenced by their treatment (Fig. 11-15. Infiltrative orbitopathy is evident in about 50% of patients, but ultrasonography, computed tomography (CT), or magnetic resonance imaging (MRI) of the orbits reveals changes, such as swelling of extracocular muscles and increased retro-orbital fat, in virtually all patients with Graves' disease, including those in whom the clinical changes are minimal or absent (see Fig. 11-16). Occasionally, infiltrative orbitopathy occurs in the absence of hyperthyroidism (so-called euthyroid Graves' disease).

Physical Examination

The ocular findings are variable (see Fig. 11-16). Exophthalmos is usually bilateral and is often asymmetrical in degree. True unilateral exophthalmos is uncommon but can occur in the absence of thyrotoxicosis; the other eye usually becomes affected eventually. When one is following the course of the disease, the degree of exophthalmos must be measured objectively with the Hertel or Luedde exophthalmometer. This instrument permits measurement of the distance between the lateral angle of the bony orbit and an imaginary perpendicular tangent to the most anterior part of the cornea.

Generally, the upper limit of normal is 20 mm, although in African Americans 22 mm is normal. In severe exophthalmos, readings may be as high as 30 mm. To estimate the degree of exophthalmos, the physician can also stand behind the seated patient and look downward from above to assert the extent to which the eyes protrude beyond the plane of the forehead.

The lids are often reddened, and enlarged lacrimal glands may cause a bulging of the surface of the eyelids. The extent to which the upper and lower lids can be completely apposed should be determined because failure of apposition promotes drying and ulceration of the cornea. Injection of the bulbar conjunctiva may be accompanied by edema or frank chemosis, in which the edematous conjunctival bulges from under the lids and around the corneal limbus. Weakness of the extracocular muscles is evident in the patient's inability to achieve or maintain convergence. Upward gaze, and especially superolateral gaze, may be limited. Occasionally, upward gaze is paralyzed. In such cases, the neck is extended and the head is tilted backward to make possible a field of vision above the horizontal. Rarely, downward or inward gaze is impaired. Ophthalmoplegia usually occurs in association with other signs of infiltrative orbitopathy but may occur alone. In some cases, only a single muscle is affected (see Fig. 11-17).

An indication of the severity of the orbitopathy may be provided by an assessment of intracranial pressure, which can be measured by a specially devised instrument (an orbitonometer) or by clinical evaluation. The patient closes the eyes lightly, and the examiner determines the ease with which the globe can be displaced posteriorly by pressure from the thumb. The manifestations of extreme orbitopathy may be catastrophic and include subluxation of the globe. Blindness may result from ulceration or infection of the cornea secondary to incomplete apposition of the lids or to optic nerve ischemia due to reduced blood flow caused by increased intracranial and intraorbital pressure. In the most severe cases, which should now be unusual, ophthalmoscopic examination may reveal venous congestion and papilledema, these may be accompanied by visual field defects.

Classification and Objective Assessment of Graves' Orbitopathy

The American Thyroid Association has classified the eye changes of Graves' disease by using a mnemonic system in which the first letters of each category constitute the term NOSPECS (Table 11-4). NO connotes the absence or a mild degree of involvement. SPECS represents the more serious degrees of involvement. NOSPECS and the numerical indices derived from it are useful as a memory tool for physical examination. The system is less satisfactory for objective assessment of orbital changes.
Objective measurements for each eye separately should include the following:

1. Documentation of maximum lid fissure width.
2. Assessment of exposure keratitis with rose bengal or fluorescein.
3. Quantitation of extraocular muscle function (with the use of the Hess chart or Maddox rod test).
5. Measurements of visual acuity, fields, and color vision.

An overall activity score is helpful in the follow-up of such patients and can be determined by assigning 1 point each for the presence of spontaneous retrobulbar pain, pain on eye movement, eyelid erythema, conjunctival injection, chemosis, swelling of the caruncle, or eyelid edema or fullness. The range is thus 0 to 7. More sophisticated orbitopathy indices require careful assessment for their reproducibility (Table 11-5).

Infiltrative Dermopathy

Dermopathy occurs in 5% to 10% of patients with Graves' disease and is almost always accompanied by infiltrative orbitopathy, usually of severe degree. These lesions cause hyperpigmented, nonpitting induration of the skin of the legs, commonly over the pretibial area (pretibial myxedema) and the dorsa of the feet, usually in the form of individual nodules and plaques but occasionally becoming confluent with a smooth characteristic edge or shoulder (see Fig. 11-14). Rarely, lesions develop on the face, elbows, or dorsa of the hands. Clubbing of the digits is occasionally associated with long-standing thyrotoxicosis (thyroid acropachy) (see also Color Plate).
Laboratory Tests

In moderate or severe Graves’ disease, laboratory findings are consonant with the pathophysiology. The serum TSH level, when measured by a sensitive immunoassay, is almost totally suppressed, and serum T₄ and T₃ levels are elevated (see Table 10-13). The free T₄ and free T₃ indices are increased more than are the T₄ and T₃ levels. The serum T₃ concentration is proportionally more elevated than the serum T₄ level. The increase in thyroid iodide uptake and clearance rate is reflected in the increased radioactive iodine uptake (RAIU). In patients with severe accompanying illness, conversion of T₄ to T₃ may be impaired, permitting the return to normal of the serum T₃ concentration but usually not the free T₄ (T₄ toxicosis); a similar effect on the relation between serum T₄ and T₃ levels can be seen in patients with Graves’ disease who have been exposed to iodine. Occasionally, the discrepancy between T₄ and T₃ levels is exaggerated, the serum T₄ concentration being normal and the serum T₃ concentration alone being elevated (T₃ toxicosis).

The physiologic basis of these tests and the manner in which they are affected by factors other than thyroid disease have been discussed earlier (Chapter 10, Laboratory Assessment of Thyroid Status). Some practical aspects of the use of the tests in the diagnosis of Graves’ disease deserve emphasis. It is neither desirable nor feasible that all the major laboratory tests be used to make the diagnosis. Documentation that the serum TSH concentration is suppressed by an appropriate assay establishes the diagnosis of hyperthyroidism in most cases and excludes the possibility of TSH-induced hyperthyroidism (see later). Because there are other causes of suppressed serum TSH, such as depression and hypothalamic-pituitary disease.

### Table 11-4 – American Thyroid Association Classification of Eye Changes in Graves’ Disease

<table>
<thead>
<tr>
<th>Class</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No physical signs or symptoms</td>
</tr>
<tr>
<td>1</td>
<td>Only signs, no symptoms (signs limited to upper lid retraction, stare, lid lag, and proptosis to 22 mm)</td>
</tr>
<tr>
<td>2</td>
<td>Soft tissue involvement (symptoms and signs)</td>
</tr>
<tr>
<td>3</td>
<td>Proptosis 22 mm</td>
</tr>
<tr>
<td>4</td>
<td>Extraocular muscle involvement</td>
</tr>
<tr>
<td>5</td>
<td>Corneal involvement</td>
</tr>
<tr>
<td>6</td>
<td>Sight loss (optic nerve involvement)</td>
</tr>
</tbody>
</table>

### Table 11-5 – Assessment of Severity of Eye Disease by Orbitopathy Activity Score

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soft tissue inflammation</td>
<td></td>
</tr>
<tr>
<td>Slight</td>
<td>1</td>
</tr>
<tr>
<td>Moderate</td>
<td>2</td>
</tr>
<tr>
<td>Severe</td>
<td>3</td>
</tr>
<tr>
<td>Exophthalmos (mm)</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>0.2</td>
</tr>
<tr>
<td>17</td>
<td>0.4</td>
</tr>
<tr>
<td>18</td>
<td>0.6</td>
</tr>
<tr>
<td>19</td>
<td>0.8</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>21</td>
<td>2</td>
</tr>
<tr>
<td>22</td>
<td>3</td>
</tr>
<tr>
<td>&gt; 23</td>
<td>4</td>
</tr>
<tr>
<td>Palpebral aperture (mm)</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.15</td>
</tr>
<tr>
<td>9</td>
<td>0.45</td>
</tr>
<tr>
<td>10</td>
<td>0.75</td>
</tr>
<tr>
<td>11</td>
<td>1.05</td>
</tr>
<tr>
<td>12</td>
<td>1.35</td>
</tr>
<tr>
<td>13</td>
<td>1.65</td>
</tr>
<tr>
<td>14</td>
<td>1.95</td>
</tr>
<tr>
<td>15</td>
<td>2.25</td>
</tr>
<tr>
<td>16</td>
<td>2.55</td>
</tr>
<tr>
<td>17</td>
<td>2.85</td>
</tr>
<tr>
<td>18</td>
<td>3.00</td>
</tr>
</tbody>
</table>
### Differential IOP (mm Hg)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1</td>
</tr>
<tr>
<td>2</td>
<td>0.2</td>
</tr>
<tr>
<td>3</td>
<td>0.3</td>
</tr>
<tr>
<td>4</td>
<td>0.4</td>
</tr>
<tr>
<td>5</td>
<td>0.5</td>
</tr>
<tr>
<td>6</td>
<td>0.6</td>
</tr>
<tr>
<td>7</td>
<td>0.7</td>
</tr>
<tr>
<td>8</td>
<td>0.8</td>
</tr>
<tr>
<td>9</td>
<td>0.9</td>
</tr>
<tr>
<td>10</td>
<td>1.0</td>
</tr>
</tbody>
</table>

### Diplopia

- Intermittent: 1
- Inconstant: 2
- Constant: 3

### Cornea

- Initial lesions: 1
- Ulcers: 2
- Clouding/perforation: 3

### Optic neuropathy

- Abnormal VEP: 3
- VA = 0.50 - 0.9: 5
- VA = 0.10 - 0.4: 7
- VA < 0.1: 9

IOP, intraocular pressure; VA, visual acuity; VEP, visual evoked potential.


### Measuring Thyrotropin Receptor Autoantibodies

Two types of tests are usually employed for the detection of TSHRAbs. The first test assesses the capacity of patient serum or IgG to inhibit the binding of \(^1\)-labeled TSH to TSH receptors from thyroid membrane preparations (available commercially) or to Chinese hamster ovary (CHO) cells expressing recombinant human TSHR. This protein-binding inhibition assay is of low cost and good precision, and the frequency of positive results in patients with active and untreated disease is on the order of 80% to 90%. Enhanced assays of this type have recently been evaluated but not yet independently.  

The second test assesses the capacity of patient's serum or IgG to stimulate adenylate cyclase or to enhance thyroid hormone or Tg secretion or iodine uptake in isolated thyroid epithelial cells or CHO-TSHR cells. Tests of this type are more expensive, have relatively poor precision, and are positive in 80% to 90% of the patients with active untreated Graves' disease. Because of the proliferation of acronyms describing these antibodies, the authors encourage the designation of the specific assay used.

Both types of tests are now available commercially. New techniques are now available for measuring TSHRAbs using recombinant antigens and chemiluminescent tags, but whether they are more sensitive remains to be determined.

### Standardization

As with all autoantibody tests, it is important to use an internationally accepted standard to allow comparison of results from different laboratories. A TSHRAb standard from the Medical Research Council (MRC) in Britain is sometimes employed. Results may be reported in MRC units. Alternatively, results have often been reported in terms of equivalent TSH units. However, the hTSHRAbs from different patients may not give parallel results with the MRC standards or TSH standards when measured in different dilutions. This means that the conversion of hTSHRAb data into MRC units or TSH units can be erroneous.

### Indications for Measuring Thyrotropin Receptor Autoantibodies

Quantitations of TSHRAbs may be a useful indicator of the degree of disease activity in an individual patient and confirm the clinical diagnosis of Graves' disease. However, a bioassay is...
not needed in a hyperthyroid patient because the patient is already demonstrating antibody bioactivity. Demonstration of TSHRAs may also be of diagnostic value in the euthyroid patient with exophthalmos, especially when it is unilateral. High TSHRAs in a pregnant woman with Graves’ disease increase the likelihood that neonatal thyrotoxicosis will be present in her offspring, and in this situation a bioassay is preferred to a radioassay.

Another use of TSHRAb testing is in the prognosis of patients with Graves’ disease who are treated with antithyroid agents. A persisting high level of TSHRAs is a useful predictor of relapse on cessation of the drug. Unfortunately, in patients with low or negative titers, the test is much less helpful. Furthermore, the presence of iodine deficiency may also interrupt the development of hyperthyroidism despite the presence of TSHRAs, and this has raised unjustified criticism of the usefulness of this test.
Differential Diagnosis

The patient with major manifestations of Graves' disease (namely thyrotoxicosis, goiter, and infiltrative orbitopathy) does not pose a diagnostic problem. In some patients, however, one of the major manifestations either dominates the clinical picture or is present alone, and the disorder may mimic another disease. All of these issues can be resolved by appropriate laboratory testing (see Table 10-13).

The diffuse goiter of Graves' disease may rarely be confused with that of other thyroid diseases if thyrotoxicosis is present. In subacute thyroiditis, particularly the painless variant, asymmetry of the gland, tenderness, and systemic evidence of inflammation assist in the diagnosis. The very low RAIU distinguishes this disease from Graves' disease. When Graves' disease is in a latent or inactive phase and thyrotoxicosis is absent, the goiter may require differentiation from Hashimoto's thyroiditis or simple nontoxic goiter as possible diagnoses. The goiter of Hashimoto's disease is somewhat lobulated and firmer and rubbery compared with that of Graves' disease. Serum levels of thyroid antibodies are generally higher in Hashimoto's disease but may not be helpful in distinguishing individual patients. In the absence of thyrotoxicosis, the diffuse goiter of Graves' disease cannot be distinguished from nontoxic, or simple, goiter. An abnormal serum TSH concentration and the presence of TSHRAbs indicate underlying Graves' disease, but their absence does not exclude quiescent disease.

Eye Disease

The orbitopathy of Graves' disease, if bilateral and associated with thyrotoxicosis past or present, does not require differentiation from exophthalmos of any other origin. However, unilateral exophthalmos, even when associated with thyrotoxicosis, should alert the physician to the possibility of a local cause. Other diseases that may produce either unilateral or bilateral exophthalmos include orbital neoplasms, carotid-cavernous sinus fistulae, cavernous sinus thrombosis, infiltrative disorders affecting the orbit, and pseudotumor of the orbit. Mild bilateral exophthalmos, generally without infiltrative signs, is occasionally present on a familial basis and also sometimes occurs in patients with Cushing's syndrome, cirrhosis.
Treatment of Hyperthyroidism

It is not yet possible to treat the basic pathogenic factors in Graves’ disease. Existing therapies for both the thyrotoxic and the ophthalmic manifestations are only palliative.

The lack of general agreement as to which therapy is the best is due to the fact that none is ideal, as reflected in the treatment guidelines of the American Thyroid Association. Because the therapeutic problems posed by thyrotoxicosis and orbitopathy differ, and because they run independent courses, their treatments are discussed separately. Treatment of thyrotoxicosis is designed to impose restraint on hormone secretion either by means of chemical agents that inhibit hormone synthesis or release or by reducing the quantity of thyroid tissue.

Antithyroid Agents

The mechanisms of action of the various antithyroid drugs are discussed in the section on the formation and secretion of thyroid hormones in Chapter 10.

Iodide Transport Inhibitors

Both thiocyanate and perchlorate inhibit thyroid iodide transport. As discussed earlier, however, theoretical and practical disadvantages attend their use except in special circumstances.

Thionamides

The major agents for treating thyrotoxicosis are drugs of the thionamide class, most commonly propylthiouracil, methimazole, and carbimazole. These agents inhibit the oxidation and organic binding of thyroid iodide and, therefore, produce intrathyroidal iodine deficiency that further increases the ratio of T\(_4\) to T\(_3\) in the thyroid secretion, as reflected in the high T\(_4\)/T\(_3\) ratio in the serum. In addition large doses of propylthiouracil, but not methimazole, impair the conversion of T\(_4\) to T\(_3\) by deiodinase type 1 in the peripheral tissues. Because of this additional action, large doses of propylthiouracil may provide rapid alleviation of severe thyrotoxicosis.

The half-life of plasma of methimazole is about 6 hours, whereas that of propylthiouracil is about 1.5 hours, and both drugs are accumulated by the thyroid gland. A single dose of methimazole may exert an antithyroid effect for longer than 24 hours. This provides a rational basis for the single-daily-dose regimen of methimazole for mild or moderate thyrotoxicosis. The propylthiouracil concentration in serum correlates with the extent of blockade of organic binding of iodine within the thyroid gland. These drugs cross the placenta and can inhibit thyroid function in the fetus. Some evidence suggests that methimazole may cross the placenta more readily than propylthiouracil, but both drugs have been used highly effectively in pregnancy (see discussion hyperthyroidism and thyrotoxicosis in pregnancy later).

Immunosuppressive Action of Thionamides

Thionamide drugs may also directly influence the immune response in patients with autoimmune thyroid disease. This action occurs within the thyroid gland, where the drugs are concentrated. The action on the thyroid cells themselves decreases thyroid antigen expression and decreases prostaglandin and cytokine release from thyroid cells. Thionamides also inhibit the generation of oxygen radicals in T cells, B cells, and particularly the APCs and hence may cause a further decline in antigen presentation. More recently, it has been shown that methimazole induces the expression of Fas ligand on the thyroid epithelial cell, thus inducing apoptosis of infiltrating lymphocytes such as T cells that express Fas and decreasing the lymphocytic infiltration (see Fig. 12-7 in Chapter 12).

The clinical importance of immunosuppression and induction of apoptosis compared with inhibition of thyroid hormone formation is unclear. However, the decrease in the immune infiltration of patients on such drugs and the fall in autoantibody levels after their introduction to a patient is powerful evidence of their effect.

Use of Thionamides

An initial dose of methimazole commonly employed is 10 to 15 mg twice a day. An equivalent dose of propylthiouracil is 150 mg every 8 hours. Carbimazole, which is converted to methimazole in vivo and is equivalent in potency, is widely used in Europe but not in the United States. These doses are effective in most patients, but in some no therapeutic response is seen, and in some patients doses of up to 60 mg of methimazole or an equivalent amount of propylthiouracil daily may be required.

It is unlikely that a true state of complete resistance to these agents ever occurs. The higher doses are required in patients with severe thyrotoxicosis and large thyroid glands, or possibly because of more rapid degradation of the drug within the gland or extrathyroidally. When large amounts are required, doses of propylthiouracil should be administered at 4- to 6-hour intervals.

The therapeutic response to effective antithyroid therapy invariably occurs after a latent period because the agents inhibit the synthesis but not the release of hormone; hence reduction in the supply of hormone to the tissues does not occur until glandular hormone stores are depleted (see Fig. 11-4 and Fig. 11-19). Although propylthiouracil differs from methimazole in having the additional effect of inhibiting the peripheral conversion of T\(_4\) to T\(_3\), there appears to be little difference in the duration of the latent period when either of these agents is employed alone in the usual dosage because the extrathyroidal effect of propylthiouracil on conversion of T\(_4\) to T\(_3\) is more apparent at dosages greater than 600 mg/day. This effect may be an advantage in the acute treatment of severe hyperthyroidism.

Factors that influence the duration of the latent period include the quantity of hormone initially stored in the thyroid gland, its inherent rate of release, and the effectiveness of blockade of new hormone synthesis achieved. In an iodine-rich thyroid gland, as when the patient has received medications containing iodine, the clinical response to antithyroid agents may be delayed for months. As would be expected, the latent period is shortened by administration of large doses (more than 600 mg daily of propylthiouracil), and such doses should be given when a more rapid therapeutic response is required. Generally, improvement within the first 2 weeks includes decreased nervousness and palpitations, increased strength, and weight gain. Usually, the metabolic state becomes normal within about 6 weeks. At this time, the dosage can often be reduced substantially to maintain a normal metabolic state.

During treatment, the size of the thyroid gland decreases in one third to one half of the patients. In the remainder, it may remain unchanged or even enlarge. In the latter situation, the change signals either an intensification of the disease process, which often requires that the dosage of drug be increased, or the production of hypothryroidism and increased TSH secretion as a result of excessive dosage.

It is important to differentiate between these causes. Clinical criteria are the main guidelines by which the adequacy of treatment is judged, but confirmation may be sought in the serum T\(_3\) and T\(_4\) levels. Mild thyrotoxicosis may persist despite a serum T\(_4\) concentration in the normal range because the serum T\(_3\) concentration may still be elevated. The latter phenomenon may also account for maintenance of a normal metabolic state in the setting of a subnormal serum T\(_4\) level. The serum TSH concentration may remain subnormal for many months, presumably secondary to accelerated conversion of T\(_4\) to T\(_3\). An enlarging thyroid gland in a treated patient with Graves’ disease may indicate the presence of a neoplasm and should be investigated appropriately.
Antithyroid agents can cause hypothyroidism if given in excessive amounts over long periods. When this occurs, the patient often complains of gain in weight, sluggishness, and fatigue, and signs of mild hypothyroidism may be present, especially a delay in the relaxation phase of the deep tendon reflexes. One major sign of incipient hypothyroidism is enlargement of the thyroid gland secondary to increased TSH. The hypothyroidism can be reversed by reducing the dosage of the antithyroid drug or by administering supplemental thyroid hormone. To forestall this development, which may also have adverse effects on preexisting orbitopathy, some physicians employ supplemental thyroid hormone routinely, the "block-and-replace" approach.

**Block-and-Replace Regimen**

The logic behind prescribing a full dose of a thionamide drug and adding T₄ supplements to prevent the patient from becoming hypothyroid is twofold. First, a few patients are difficult to keep euthyroid with thionamide therapy alone, and a block-and-replace regimen can be helpful and requires fewer office visits. Second, the immunosuppressive action of the thionamides may be helpful in attenuating the natural history of the autoimmune thyroid diseases directly.

Although some investigators found the relapse rate after the block-and-replace approach to be much reduced, others have found no difference. One group has reported that continuing levothyroxine replacement after withdrawal of antithyroid drugs also increased the remission rate, possibly because suppression of pituitary TSH inhibited expression of thyroid antigens and reduced immune stimulation (an effect influenced by the level of TSHRAbs). Such studies have not been reproduced, and this approach is not recommended.

### Predicting the Response to Drug Withdrawal

A central question in the treatment of Graves' disease to which there is no simple answer is the appropriate duration of antithyroid drug treatment. As discussed earlier, antithyroid therapy may alter the course of the underlying autoimmune process, but remission after withdrawal of treatment will persist only if the disorder has entered a latent or inactive phase. This latter transition and the normal decline in the levels of TSHRAbs are more likely to occur the longer the course of treatment. This reasoning is the basis for the traditional practice of continuing antithyroid treatment for 6 to 12 months or longer. However, persistence of high levels of circulating TSHRAbs during treatment of Graves' disease portends recurrence after withdrawal of antithyroid drugs in iodine-replete areas.

Factors preventing a recurrence include (1) a change from stimulating antibody to blocking antibody, which occurs rarely, and (2) the progression of concomitant thyroiditis. Furthermore, iodine deficiency itself may prevent the recurrence of Graves' disease. These factors may explain why some authors have been unable to confirm the predictive value of TSHRAb measurements. The use of poorly validated assays for TSHRAbs may compound this problem. However, most patients do not have persisting high levels of TSHRAb, and predicting their outcome is more difficult; additional factors need to be taken into account.

Other features associated with the likelihood of long-term remission after withdrawal of therapy (Table 11-6) include (1) the initial presence of T₄ toxicity, (2) a small thyroid gland (less than twice normal), (3) a decrease in the size of the thyroid gland, and (4) in particular, return of the TSH concentration to normal during treatment. HLA typing is not helpful in such predictions.

Hence, treatment should generally be continued for about 6 to 12 months and then withdrawn if the TSHRAbs disappear. Alternatively, if the patient's condition can be easily controlled with low doses of antithyroid drugs and the serum TSH has returned to the normal range, antithyroid drugs may be withdrawn and the serum TSH concentration measured at monthly intervals. About 75% of relapses occur in the first 3 months after withdrawal of therapy, and most of the remainder occur during the subsequent 6 months. Suppression of the TSH

### TABLE 11-6 -- Factors Favoring Long-Term Remission after Antithyroid Therapy for Graves' Disease

| Factor                                      | Remission
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>T₄ toxicity</td>
<td>Favorable</td>
</tr>
<tr>
<td>Small goiter</td>
<td>Favorable</td>
</tr>
<tr>
<td>Decrease in goiter size during therapy</td>
<td>Favorable</td>
</tr>
<tr>
<td>Normal thyroid function tests and normal serum TSH</td>
<td>Favorable</td>
</tr>
<tr>
<td>Negative tests for TSH receptor autoantibody</td>
<td>Favorable</td>
</tr>
<tr>
<td>TSH, thyrotropin</td>
<td>Favorable</td>
</tr>
</tbody>
</table>

Concentration is the first signal of relapse even in the presence of a normal serum T₄ level.

### Long-Term Remission

The frequency with which long-term remission occurs after withdrawal of antithyroid therapy has decreased over the past 30 years. In part because of the increase in dietary iodine intake, but this decrease has also occurred in geographic regions where iodine intake has remained constant and low. Nevertheless, about one-third of patients experience a lasting remission. This fact alone indicates that antithyroid agents have a significant role as a sole therapy in the initial treatment of thyrotoxicosis.

### Adverse Reactions

Adverse reactions occur in a small number of patients taking thionamide drugs (Table 11-7). Agranulocytosis occurs in fewer than 1% of the patients, generally within the first few weeks or months of treatment. It is accompanied by fever and sore throat. When therapy is begun, the patient should be instructed to discontinue the drug and to notify the physician immediately should these symptoms develop. This precaution is more important than the frequent measurement of leukocyte counts. The frequency of lymphopenia in hyperthyroidism, a complete blood count with differential is recommended before antithyroid drug therapy is started. If the absolute neutrophil count falls below 1500 cells/µL, the drug should be withdrawn. If agranulocytosis occurs, the drug should be discontinued immediately and the patient treated with antibiotics as appropriate. Granulocyte colony-stimulating factor has been used to accelerate recovery that invariably takes place. Lymphocytes of patients who have developed agranulocytosis while taking propylthiouracil undergo blast transformation when exposed to propylthiouracil or methimazole; consequently, they should not be given a thionamide drug again. Granulocytopenia occurs during antithyroid therapy and is sometimes a forerunner of agranulocytosis, but, as already mentioned, it can also be a manifestation of thyrotoxicosis itself. For this reason, and as noted previously, a total white blood cell count with differential should be obtained before initiation of treatment with thionamide drugs.

Granulocytopenia that develops during the first few weeks of therapy may be difficult to interpret. In this circumstance, serial measurements of the leukocyte count should be made. If they display a downward trend, the antithyroid drugs should be discontinued. When serial measurements of the white blood cell count remain constant or return to normal, treatment need not be interrupted.

### TABLE 11-7 -- Incidence of Toxic Reactions with Antithyroid Drugs

<table>
<thead>
<tr>
<th>Drug</th>
<th>All Reactions (%)</th>
<th>Agranulocytosis (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methimazole</td>
<td>7.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Carbamazole</td>
<td>1.9</td>
<td>0.8</td>
</tr>
<tr>
<td>Propylthiouracil</td>
<td>3.3</td>
<td>0.4</td>
</tr>
</tbody>
</table>

A rash that can take many forms, including hives, occurs in as many as 10% of patients. Less frequent reactions include arthralgia, myalgia, neuritis, hepatitis (with propylthiouracil) or cholestasis (with methimazole) and rare liver necrosis necessitating transplantation, thrombocytopenia, loss of or abnormal pigmentation of the hair, loss of taste sensation, enlargement.
of lymph nodes or salivary glands, edema, a lupus-like syndrome, and toxic psychoses. The mechanisms underlying these reactions are not known, although some reactions disappear with discontinuance of treatment. It is obviously helpful to have a baseline complete blood count and liver function studies before initiation of antithyroid drugs to help interpret the presence of some of these side effects. We believe that the suspicion of any serious manifestation should be an indication for abandonment of antithyroid therapy and a recourse to surgery or 131 I.

Iodine and Iodine-Containing Agents

Iodine is now rarely used as a sole therapy. The mechanism of action of iodine in relieving thyrotoxicosis differs from that of the thionamides. Although quantities of iodine in excess of several milligrams can acutely inhibit organic binding (acute Wolff-Chaikoff effect), this transient phenomenon probably does not contribute to the therapeutic effect. Instead, the major action of iodine is to inhibit hormone release (see Chapter 10). Administration of iodine increases glandular stores of organic iodine, but the beneficial effect of iodine is evident more quickly than the effects of even large doses of agents that inhibit hormone synthesis (Fig. 11-19). In patients with Graves’ disease, iodine acutely retards the rate of secretion of T₄, an effect that is rapidly lost when iodine is withdrawn. These features of iodine action provide both disadvantages and advantages. The enrichment of glandular organic iodine stores that occurs when this agent is given alone may retard the clinical response to subsequently administered thionamide, and the decrease in RAU produced by iodine prevents the use of radioiodine as treatment for several weeks. Furthermore, if iodine is withdrawn, resumption of accelerated release of hormone from an enriched glandular hormone pool may exacerbate the disorder.

Another reason for not using iodine alone is that the therapeutic response on occasion is either incomplete or absent, and even if initially effective iodine may lose its effect with time. (This phenomenon, which has been termed iodine escape, should not be confused with the escape from the acute Wolff-Chaikoff effect [see Chapter 10].) Nevertheless, the rapid slowing of hormone release by iodine makes it more effective than the thionamide drugs when prompt relief of thyrotoxicosis is mandatory (see Fig. 11-19). Therefore, aside from its use in preparation for subtotal thyroidectomy, iodine is useful mainly in patients with acute or impending thyrotoxic crisis, severe thyrotoxic disease, or acute surgical emergencies.

If iodine is used in these circumstances, it should be administered with large doses of a thionamide, as the severity of the thyrotoxicosis itself indicates. The dose of iodine required for control of thyrotoxicosis is approximately 6 mg daily, a quantity much less than that usually given. Six milligrams of iodine is present in one eighth of a drop of saturated solution of potassium iodide (SSKI) or an 0.8 drop of Lugol’s solution; many physicians, however, prescribe 5 to 10 drops of one of these agents three times daily. Although it is advisable to administer amounts larger than the suggested minimal effective dose, huge quantities of iodine are more likely to produce adverse reactions, including iodide myxedema. We recommend the use of a maximum of 3 drops of SSKI three times daily.

In patients who are so ill that medications cannot be taken by mouth, antithyroid agents can be triturated and administered by stomach tube; iodine can be given by the same route or can be absorbed through the mucosa. When use of a stomach tube is contraindicated, thionamide drugs cannot be administered because no parental preparations are available. Here, the disadvantages attendant on administration of iodine may be accepted if the clinical situation is sufficiently serious. Iodine appears to be particularly effective after administration of a therapeutic dose of 131 I for the rapid alleviation of thyrotoxicosis.

Reactions to Iodine

Adverse reactions to iodine are unusual and are generally not serious but may include rash, which may be acneiform; drug fever; sialadenitis; conjunctivitis and rhinitis; vasculitis; and a leukemoid eosinophilic granulocytosis. Sialadenitis may respond to reduction of dosage; in the case of the other reactions, iodine should be stopped.

Ipodate

In doses of 1 g daily, the iodine-containing choleystographic contrast agent sodium ipodate (or iopanoate) causes a prompt decrease in serum T₄ and serum T₃ concentrations in patients with hyperthyroidism. These effects are the result of both the release of iodine and the ability of the agent to inhibit peripheral T₃ production from T₄, a combination that can be useful in the seriously ill patient. As with iodine itself, however, withdrawal of the drug carries the risk of an exacerbation. Hence, if the patient is sufficiently ill to warrant treatment with ipodate, large doses of antithyroid agents should be administered concomitantly.

Other Antithyroid Agents

Lithium

Lithium carbonate also inhibits thyroid hormone secretion, but, unlike iodine, it does not interfere with the accumulation of radioiodine. Lithium, 300 to 450 mg every 8 hours, is employed only to provide temporary control of thyrotoxicosis in patients who are allergic to both thionamide and iodide. This is because the blocking effect is often lost with time. The goal is to maintain a serum concentration of 1 mEq/L.

Dexamethasone

Dexamethasone, 2 mg every 6 hours, inhibits the glandular secretion of hormone, inhibits the peripheral conversion of T₄ to T₃, and has immunosuppressive effects. The inhibitory effect of dexamethasone on the conversion of T₄ to T₃ is additive to that of propylthiouracil, suggesting a different mechanism of action. Concurrent administration of propylthiouracil, SSKI, and dexamethasone to the patient with severe thyrotoxicosis effects a rapid reduction in serum T₃ concentration, often to within the normal range in 24 to 48 hours.

Beta-Blocking Agents

Agents that block the response to catecholamines at the receptor site (e.g., propranolol) ameliorate some of the manifestations of thyrotoxicosis and are often used as adjuncts in management. Tremulousness, palpitations, excessive sweating, eyelid retraction, and heart rate decrease; effects are rapidly manifested and appear to be mediated largely through the adrenergic nervous system, although propranolol may also impair the conversion of T₄ to T₃.

Adrenergic antagonists are most useful in the interval when a response to thionamide or radioiodine therapy is being awaited. They are of limited usefulness in patients with mild to moderate disease but are useful in patients with severe thyrotoxicosis, such as those with impending or actual thyrotoxic crisis (see special aspects of thyrotoxicosis later). Adrenergic antagonists are especially useful when tachycardia is contributing to cardiac insufficiency. However, the fact that -adrenergic blockers can reduce cardiac output without altering oxygen consumption can have adverse effects in some organs, such as the liver, where the arteriovenous oxygen difference is already elevated in the hyperthyroid state. Moreover, because thyroid hormone has a direct effect on the myocardium independent of the adrenergic nervous system, adrenergic antagonists reduce the heart rate by an independent mechanism (see earlier discussion of catecholamine-thyroid interrelationships).

Propranolol is the most widely used agent because it is relatively free from adverse effects and can be given orally in a dose of 20 to 80 mg every 6 or 8 hours. For intravenous use, a shorter-acting agent may be preferred (see treatment of “Thyrotoxic Crisis [Thyroid Storm]). Propranolol is contraindicated in patients with asthma or chronic obstructive pulmonary disease because it aggravates bronchospasm. Because of its myocardial depressant action, it is also contraindicated in patients with heart block and in patients with congestive failure, unless severe tachycardia is a contributory factor. Whether propranolol should be given chronically to preganant women with hyperthyroidism has been questioned, and we avoid it where possible. Some studies indicate that its use causes no significant complications, whereas others report an association with small size of the fetus, low Apgar scores, and postnatal bradycardia and hypoglycemia.

Another -blocking agent is metoprolol, a longer-acting drug that allows a once-a-day regimen when treatment is likely to be prolonged. Calcium channel-blocking
agents such as dillazem may also be used when -blocking agents are contraindicated.

Surgery

Both types of ablative therapy—surgery and radiodine—ameliorate thyrotoxicosis by permanent removal or destruction of thyroid tissue, impairing the capacity of the gland to synthesize hormone. Antithyroid therapy, aimed at preserving the thyroid gland, and ablative therapy are different, and their opposite properties may be advantageous or disadvantageous, depending on one's point of view. The impermanence of antithyroid therapy leads to a relatively frequent recurrence, whereas recurrence is uncommon with ablative therapy. However, antithyroid therapy probably does not cause permanent hypothyroidism, whereas the frequency of permanent hypothyroidism is very high with ablative therapy.

The surgical procedure of choice for the treatment of Graves’ disease is a bilateral subtotal thyroidectomy that avoids the dangers of hypoparathyroidism and laryngeal nerve injury. Surgery is effective in relieving hyperthyroidism, the frequency of recurrent hyperthyroidism after subtotal thyroidectomy in adults being less than 5% even when the procedure is performed by experienced surgeons. Nevertheless the high prevalence of postoperative hypothyroidism makes surgery an imperfect treatment.

Table 11-8 is taken from summaries of results of surgery that have not changed significantly in recent years except for the large decline in mortality to near zero in most reports. The incidence of permanent hypothyroidism ranged in frequency from 4% to approximately 30% and was highest in clinics in which internists did the follow-up examinations. In a study conducted by internists, a mean frequency of postoperative hypothyroidism of 28% was found in patients followed for 1 to 16 years, and the frequency in patients followed for 10 years was 43%.

Although it was previously assumed that hypothyroidism usually develops within 1 year after operation, long-term studies indicated a progressive increase in the cumulative incidence with time, similar to that produced by radioiodine but of lesser magnitude. It is likely that the frequency of partial impairment of thyroid function (as revealed by small increases in serum TSH) is even higher than that of hypothyroidism because the aim of subtotal thyroidectomy is to decrease thyroid reserve. The increasing frequency with time of hypothyroidism may result from progressive restriction of blood supply or from autoimmune destruction of the thyroid remnant.

If eventual thyroid failure is a frequent consequence of the Graves’ disease process itself, the increase in the cumulative frequency with time of hypothyroidism after either surgery or radiodine therapy is to be expected and is unavoidable. Treatment that destroys thyroid tissue would accelerate the emergence of hypothyroidism resulting from the disease process itself.

There is an inverse relationship between the frequency of recurrence and that of hypothyroidism, and both partly depend on the amount of thyroid tissue left in place. When one considers that thyroid glands vary in size and degree of hyperfunction and that the techniques of surgeons vary to a considerable extent, it is remarkable that a normal metabolic state is restored for long periods in most patients. The reason for this favorable outcome may be that the amount of tissue remaining after operation is insufficient to sustain a normal metabolic state and hence becomes stimulated by endogenous TSH. In this way, the patient's homeostatic mechanism provides the adjustment in thyroid function that surgery alone could not. This hypothesis is supported by the return of serum TSH levels to normal in patients restored to a normal metabolic state by surgery. However, this explanation would suffice only in the absence of TSHR Abs, which rapidly decrease and disappear in many patients after surgery. How the autoimmune disease is suppressed following surgery is unclear, but clearly the release of thyroid antigen during the procedure must induce apoptosis of many of the clones of TSHR-specific T cells and B cells.

Complications of Surgery

Because the hazards of subtotal thyroidectomy are inversely related to the experience and skill of the surgical team, it is impossible to generalize about the frequency of complications. Furthermore, data from the era in which surgery was common are probably no longer applicable (see Table 11-8). Unless circumstances are otherwise compelling, thyroidectomy should not be performed by surgeons who do the operation only occasionally. Bleeding into the operative site, the most serious postoperative complication, can rapidly produce death by asphyxia and requires immediate evacuation of the blood and ligation of the bleeding vessel. Even with subtotal surgery, the recurrent laryngeal nerve can be damaged. If such damage is unilateral, it causes dysphonia that usually improves in a few weeks but that may leave the patient slightly hoarse. If laryngeal nerve damage is bilateral, obstruction of the airway can cause stridor within hours; tracheostomy is then required, at which time the nature of the damage to the nerves should be explored.

Hypoparathyroidism can be either transient or permanent. Transient hypoparathyroidism results from inadvertent removal of some parathyroids and impairment of blood supply to those that remain. Depending on the severity of these insults, symptoms and signs of hypocalcemia appear, usually within 1 to 7 days after surgery. The earliest evidence of hypoparathyroidism may be anxiety and mental depression, followed by paresthesias and heightened neuromuscular excitability, such as Chvostek's and Trousseau's signs and carpopedal spasm. The serum calcium level is subnormal, and the serum inorganic phosphate level is increased.

Severe hypoparathyroidism should be treated with intravenous calcium gluconate. Milder cases can be treated with oral calcium carbonate in a dose of 1 g three times daily. It is impossible at the onset to predict whether hypoparathyroidism will be permanent or will regress within a few weeks, as usually occurs. Some surgeons insist on prophylactic calcium and vitamin D after every thyroidectomy. Of course, this approach may hide any developing deficiency. With the increasing use of ambulatory thyroid surgery, it is likely that this approach will grow.

However, the hypocalcemia that occurs immediately after surgery for thyrotoxicosis may not be due to transient hypoparathyroidism, because it occurs more frequently here than after surgery for other thyroid disorders. Instead, it may be due to retention of calcium by bone because of the demineralization of bone that occurs in hyperthyroidism, which begins to be reversed after cure of the hyperthyroid state and may contribute to the modest elevation in alkaline phosphatase during recovery. The frequency of permanent hypoparathyroidism correlates with the proportion of the thyroid gland removed and with the frequency of postoperative hypothyroidism. The incidence of mild hypoparathyroidism (or diminished parathyroid reserve) detectable years after surgery is probably greater than is generally supposed. The treatment of hypoparathyroidism is discussed in Chapter 26.

Preparation for Surgery

Preoperative use of antithyroid agents has greatly decreased the morbidity and mortality rates of surgery for Graves’ disease because these drugs deplete glandular hormone stores and restore the metabolic state to normal. However, these agents do not improve the hyperplasia and hypervascularity of the gland unless TSHRAbs levels fall. Iodine, however, is reported to cause a decrease in height of the follicular cells, enlargement of follicles with retention of colloid, and reduction of hypervascularity. Hence, the aim of preoperative management is to restore the metabolic state to normal with antithyroid agents and then to induce involution of the gland with iodine.

Patients who are to undergo subtotal thyroidectomy are first given antithyroid therapy in the manner described earlier. Often, relatively large doses are given in order to hasten the clinical response and because surgical candidates are often patients with severe disease or large goiters. After the metabolic state is restored to normal, SSRI is added (3 drops three times daily) for a further 7 to 10 days. During this period, a preexisting bruit or thrill may decrease in intensity or disappear entirely and

Table 11-8 -- Effects of Surgery for Hyperthyroidism

<table>
<thead>
<tr>
<th>Result</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recurrent hyperthyroidism</td>
<td>0.698</td>
</tr>
<tr>
<td>Vocal cord paralysis</td>
<td>0.034</td>
</tr>
<tr>
<td>Permanent hypoparathyroidism</td>
<td>0.036</td>
</tr>
<tr>
<td>Permanent hypothyroidism</td>
<td>5.875</td>
</tr>
</tbody>
</table>
the gland usually becomes firm.

Several cautions should be observed:

1. No date for surgery should be set until a normal metabolic state has been restored. Much too often, the operation is planned well in advance and the patient is given a standardized regimen independent of the clinical progress.

2. Therapy with iodine should not be started until a normal metabolic state has been restored; iodine should not be relied on to complete an as yet incomplete response to antithyroid therapy because iodine will enrich glandular hormone stores if the antithyroid drug is not entirely effective.

3. Antithyroid agents should not be withdrawn when iodine therapy is begun.

**Thyroid Surgery in the Hyperthyroid Patient**

Propranolol may be a useful adjunct in controlling signs and symptoms (see earlier) while the patient is being prepared for surgery. Propranolol has been used alone in preoperative preparation of the patient in whom surgery is to be undertaken, and although this mode of therapy is probably safe and effective in many patients with mild disease, thyroid crises can occur in patients receiving propranolol alone. Therefore, we believe that unless there is some compelling indication for the use of propranolol alone, restoration of the patient to a eumetabolic state, as outlined earlier, is appropriate before subjecting the patient to the stress of surgery.

**Radioiodine**

Radioiodine produces thyroid ablation without the complications of surgery. The principal disadvantages of radioiodine are the influence of radiation on Graves' ophthalmopathy and the high frequency of late hypothyroidism. Previously, there was concern that this form of therapy might also produce thyroid carcinoma, leukemia, or an increase in mutation rates. However, during the half-century in which radioiodine has been in use, no increased prevalence of thyroid or other carcinoma in treated patients has been noted. This phenomenon is to be contrasted with the increased prevalence of thyroid carcinoma in patients treated with low amounts of radiation in childhood or adolescence as exemplified by the results of the Chernobyl radiation leak. The prevalence of leukemia is also no greater in adults treated with radioiodine, and the frequency of genetic damage in the offspring of patients treated earlier with radioiodine does not appear to be increased.

Indeed, the conventional dose of radioiodine employed in the treatment of thyrotoxicosis delivers to the gonads a radiation dose about equivalent to that delivered by a barium enema examination or intravenous urogram.

In view of the lack of evidence of serious toxicity from radioiodine in doses generally employed for treating hyperthyroidism, the age limit for the use of radioiodine has been lowered progressively from the initial limit of 40 years, and in some clinics it is now employed in children and adolescents. Experience from the Chernobyl nuclear accident, which caused a large increase in the number of childhood thyroid cancers, may alter this trend, particularly for adolescents, and many physicians think that the use of any radioactivity in children should be avoided if possible. Hence, there is regional and international variation in the use of radioiodine therapy.

Attempts have been made to standardize the radiation delivered to the thyroid gland by varying the dose of radioiodine according to the size of the gland, the uptake of \(^{131}I\), and its subsequent rate of release (dosimetry). However, such calculations do not provide uniform results, probably because of variations in individual sensitivity. Hence, many physicians have settled on an arbitrary dose calculated to result in the delivery of 300 MBq (8 mCi) of \(^{131}I\) to the thyroid gland 24 hours after administration. Others aim to deliver 50 to 100 GY (5000 to 10,000 rad) to the gland.

**Hypothyroidism after Radioiodine**

Many reports have documented that the incidence of hypothyroidism is significant during the first year or two after treatment with RAI and continues to increase at a rate of approximately 5% per year thereafter. The incidence of postradioiodine hypothyroidism at 5 years is approximately 30% and at 10 years is approximately 40%, although values as high as 70% have been reported. Such values, of course, also depend on the dose delivered in the different centers.

The beneficial effect of radioiodine and the early induction of hypothyroidism are both consequences of radiation-induced destruction of thyroid parenchyma. Radiation thyroiditis may develop within the first few weeks of treatment, as evidenced by epithelial swelling and necrosis, disruption of follicular architecture, edema, and infiltration with mononuclear cells. Resolution of the acute phase is followed by fibrosis, vascular narrowing, and further lymphocytic infiltration. These changes account for the early response to radioiodine, be it favorable or excessive, but do not appear sufficient to account for the continuing development of hypothyroidism with time.

In some studies, the likelihood of hypothyroidism is increased by the presence of high levels of thyroid antibodies and, presumably, of thyroid specific T cells, at the time of treatment and with increasing age of the patient. The two predisposing factors may be related to one another. If this is true, it is unlikely that the early ablative effects can be obtained free from subsequent late effects, and doses of radioiodine sufficient to exert an early therapeutic action will inevitably be associated with a high frequency of delayed hypothyroidism.

This therapeutic dilemma with respect to radioiodine therapy is handled differently in different practices. Some continue to administer the conventional dose because of its effectiveness and because hypothyroidism, when it eventually occurs, can be easily treated. A disadvantage of such an approach is that the onset and progression of hypothyroidism may still be insidious, that prolonged follow-up of patients may not be possible, and that patients may not associate symptoms arising long after therapy with a complication. The advantage of this approach is that it minimizes the dangers of persistent or recurrent thyrotoxicosis, which may be hazardous, especially in the elderly.

One way to minimize the frequency of hypothyroidism is to administer a dose per gram of estimated weight that is larger the greater the gland size. In this way, many patients with large thyroid glands are given large doses, and the converse is true for thyroid glands that are small. However, regimens of this type do not appear to improve the treatment of hyperthyroidism in the short run, and it is not known whether the rate of later hypothyroidism is changed.

In a controlled, prospective study, the effects of a single conventional dose of approximately 5.2 MBq/g (140 µCi/g) of estimated glandular weight were compared with the effects of half the dose. Although the therapeutic effect of radioiodine developed more slowly in patients receiving the half-dose, and although a greater proportion required antithyroid drug therapy until this effect became apparent, the frequency of remission after 2 years was the same as that in patients receiving the conventional dose, and recurrence of thyrotoxicosis was no more common. In the full-dose group, the incidence of hypothyroidism was 6% at 1 year and 29% at 5 years, whereas in the half-dose group the corresponding values were 4% and 7%. However, thereafter the cumulative frequency with time with the low dose was...
similar to that observed with conventional doses. For these reasons, some physicians advocate an ablative approach to treatment.

The use of antithyroid drugs before RAI treatment is widely used to theoretically decrease a post-RAI increase in thyroid hormone release. This is considered especially dangerous in older age groups with ischemic heart disease in which cardiac deaths have been reported. Antithyroid drugs may also prevent the post-RAI increase in thyroid autoantibodies that may affect ophthalmopathy. It is important to monitor free T4 and free T3 levels in at-risk patients and to consider -adrenergic blockade whether or not antithyroid drugs are used before RAI treatment.

Orbitopathy and Radioiodine

As discussed earlier, Graves' orbitopathy is probably the result of a specific cytokine between retro-orbital and thyroid antigens, perhaps the TSH receptor itself. Any worsening of the autoimmune thyroid response might therefore worsen the orbital immune response. Following radioiodine therapy, the levels of circulating TSH-RAs are elevated strikingly, perhaps secondary to impairment of immune restraint caused by the intrathyroidal irradiation where regulatory cells may be more sensitive. This change is in keeping with exacerbation of preexisting myxedema after radioiodine administration. Similarly, carefully conducted studies indicate that eye disease worsens in about 10% of patients with Graves' orbitopathy who are treated with radioiodine (Fig. 11-22), although others disagree. Such changes, if any, are usually mild and temporary but on occasion can involve dramatic deterioration.

Some physicians advocate the use of glucocorticoids at the time of radioiodine treatment to prevent such effects. One regimen involves prednisone, 0.4 to 0.5 mg/kg 1 month before 131I treatment, with a gradual tapering over 3 to 4 months. Others suggest that radioiodine may not be the treatment of choice in patients with significant orbitopathy. However, maneuvers such as careful control of thyroid function before and after therapy and cessation of smoking by the patient may minimize ocular changes.

Other Side Effects of Radioiodine

Additional hazards may attend the use of radioiodine, particularly large doses. The parathyroid glands are exposed to radiation in patients treated with radioiodine. Although parathyroid reserve may be diminished in some patients, development of overt hypoparathyroidism is rare. The effect of radioiodine on other tissues that concentrate iodide (e.g., the salivary glands, the gastric glands, and the breasts) has received little attention.

Radiation thyroiditis itself may lead to an exacerbation of thyrotoxicosis 10 to 14 days after radioiodine is administered, with occasionally serious consequences, including precipitation of thyrotoxic crisis and aggravation of patients with severe thyrotoxicosis or cardiac insufficiency. In thyrocardiac disease, therefore, antithyroid drugs should be given for several weeks before radioiodine is given to deplete glandular hormone stores. This prevents an outpouring of hormone if severe radiation thyroiditis occurs. The antithyroid agent is withdrawn about 3 to 5 days before administration of the radioiodine; if the clinical condition warrants, the agent can be begun again 7 days later.

Because 131I administration is contraindicated during pregnancy, a pregnancy test should be carried out in women of childbearing age before 131I therapy is initiated if there is any possibility of pregnancy.

General Measures

Several general measures may contribute to the well-being of the patient with severe thyrotoxicity. Help with family responsibilities and avoidance of physical exertion are also important. In addition, a diet rich in protein, calories, and vitamins may repair the nutritional deficiencies that are common in thyrotoxic patients.

Choice of Therapy

The choice of therapy for thyrotoxicosis is influenced by emotional attitudes, economic considerations, and family and personal issues. Our choice of therapy takes into account the natural history of the disease, the advantages and disadvantages of the available therapies, and the features of the population group in which the patient falls. Apart from patients directly requesting surgery, this procedure is recommended only when the shortcomings of other modes of therapy are of particular importance (e.g., patients with antithyroid drug allergy, a cold nodule, patients with very large goiters, and patients with the need for a rapid return to normal).

Occasionally, in young adults, it is necessary to remove a diffuse toxic goiter because of obstructive symptoms or cosmetic disfigurement. Nevertheless, only a small percentage of patients with Graves' disease are recommended for surgery. The choice, therefore, is among antithyroid drugs, RAI, or a mixture of both.

In one common approach to therapy in adults, the physician initiates treatment with antithyroid drugs in all patients to produce a euthyroid state before reaching a final decision regarding a definitive therapeutic strategy. This allows the patient to return to a euthyroid status as rapidly as possible and provides an estimate of the antithyroid drug dose requirement. The magnitude of the drug requirement and the size of the thyroid gland are two of a number of factors considered in the evaluation of the patient with regard to the likelihood of a remission. The options for treatment are explained to the patient during these first months of contact, and individual recommendations are then formulated. This approach allows the establishment of a workable physician-patient relationship, which is especially important in addressing anxieties about the use of radioiodine. Such concerns lead many patients, especially those younger than 50 years of age, to elect a trial of antithyroid drugs before definitive therapy with 131I.

Patients with a large thyroid gland, a maintenance thionamide dose requirement of more than 400 mg/day of propylthiouracil, and high titers of TSH-RAs require prolonged antithyroid treatment and are advised that the chance of spontaneous remission is less than 30%. A therapeutic trial is generally pursued for 6 to 12 months if long-term thionamide therapy is selected. One can, in theory, treat forever unless side effects become a problem. When a decision in favor of radioiodine is made, I may be prescribed at a dose designed to result in the retention of about 300 MBq (8 mCi) in the thyroid gland at 24 hours. This estimate is based on a 123I uptake test performed immediately before treatment and at least 5 days after stopping thionamides. Patients with a larger goiter (more than four times normal) or those who have received large doses of propylthiouracil may require more radioiodine. Because 131I is given when the patient is euthyroid and this is a relatively large dose, no additional therapy is required immediately after treatment except for patients in whom a recurrence of hyperthyroidism poses a medical risk (e.g., patients with coronary artery disease or congestive heart failure).

Patients are seen at 4-week intervals after 131I administration, and hypothyroidism is treated when it appears with an elevated TSH level, generally within 3 months. Women planning to become pregnant are advised to wait for an arbitrary period of 6 months after 131I therapy to allow for resolution of any transient effects of gonadal radiation. If, after a period of 6 months, hyperthyroidism is still present and the patient is symptomatic, the treatment is repeated, generally with about 1.5 times the initial dose of 131I.
Although the foregoing reflects our approach to therapy, the opinions of thyroïdologists differ widely. \[151\] In view of the several approaches to treatment available, each with its advantages and disadvantages, it is incumbent on the physician to explain these factors thoroughly to patients, to indicate a preference and the reasons for it, and to allow the final choice to rest with the patient when appropriate.

### Hypothyroidism in the Recently Hypothyroid Patient

The early onset of hypothyroidism may cause distinct symptoms in the previously thyrotoxic patient after \[151\] or surgical treatment or even with high doses of thionamide drugs. Such patients may develop severe muscle cramps, often in large muscle groups such as the trapezius or latissimus dorsi or the proximal muscles of the extremities. Such symptoms can develop even when the serum hormone levels are only low-normal or slightly decreased and before the TSH concentration has risen. It is possible to mistake a symptom such as back pain for an unrelated illness unless the patient is warned in advance. It is also not unusual for patients to complain of hypothyroid symptoms when thyroid function test results return to within the normal range. Such patients appear to have trouble adjusting to the normal thyroid hormone levels after being exposed to excessive amounts for long periods.

### Treatment of Infiltrative Orbitopathy or Infiltrative Dermopathy

Infiltrative orbitopathy varies in severity from the common mild form to a severe form that threatens vision. The latter type is rare but remains difficult to treat. Indeed, the most effective therapies are merely palliative. The natural course of the disorder, which is variable and characterized by exacerbations and remissions, makes conclusions about the efficacy of any treatment difficult. A further source of confusion is the variable terminology for describing the manifestations of orbitopathy and the lack of rigid criteria for defining their severity. Use of the American Thyroid Association classification and its expanded indices, described earlier, is strongly recommended (see Table 11-4 and Table 11-5).

### Effect of Treatment of the Thyroid Gland on Orbitopathy

The first question that arises is whether different treatments for thyrotoxicosis affect the course of the eye disease differently. Subtotal thyroidectomy and thionamide drug therapy do not influence ophthalmopathy unless they lead to the development of hypothyroidism. Hypothyroidism has an adverse effect on the disorder and should be treated fully when it occurs. However, exogenous thyroid hormone in the absence of hypothyroidism does not improve the ophthalmopathy.

As discussed earlier, controlled studies suggest that radioiodine treatment may lead to a slight but significant worsening of orbitopathy (see earlier discussion), \[160\] and it may be best to avoid radioiodine in patients with severe eye disease. Alternatively, as mentioned earlier, coincidental glucocorticoid therapy may prevent deterioration of orbitopathy after radioiodine but may itself cause significant side effects. \[165\] Controlled, prospective studies of the influence of antithyroid drug treatment prior to radioiodine on oculare changes are needed.

### Symptomatic Treatment

Treatment modalities can be those that are largely symptomatic (useful mainly in the mild form) and those that attempt to arrest or reverse the progression of the disorder. With milder forms, little treatment is required. The patient who experiences photophobia and sensitivity to wind or cold air can benefit by wearing dark glasses, which also afford protection from foreign bodies. Elevation of the head of the bed at night and instillation of lubricants, such as 1% methylcellulose, may help when the eyelids do not appose completely during sleep. Artificial tears can be used during the day. Because the opthalmic manifestations tend to be self-limited and the progression to a more severe form is uncommon, such measures usually suffice to tide the patient over until the disorder regresses spontaneously.

### Glucocorticoids

The appearance of increasing proptosis with inability to appose the eyelids or of severe infiltrative manifestations such as chemosis warrants the use of more vigorous therapeutic measures. Such changes, even when severe, may respond favorably and rapidly to glucocorticoids. Some physicians use massive doses of prednisone (120 to 140 mg/day). If improvement occurs, the dose is decreased to the lowest level at which improvement is maintained. The latter dose is still likely to be large, but it is hoped that a halt to the progression or actual regression of the disease will occur before untoward effects make withdrawal of the drug necessary. Other physicians find that much smaller doses of prednisone (20 to 30 mg/day) can be highly effective with rapid reduction to a longer-term maintenance dose (10 to 15 mg/day). Intravenous hydrocortisone pulse therapy is said to have the advantage of fewer side effects than high doses of prednisone. \[169\] \[170\]

To circumvent the inevitable side effects of large doses of glucocorticoids, periodic injection of depot preparations of glucocorticoids subconjunctivally or into the retro-orbital space has been tried but is not recommended. Such treatment may have a dramatic effect on irritative symptoms as well as on diplopia, but the efficacy varies, and systemic effects of the glucocorticoids are sometimes seen. Moreover, this treatment entails the risk of puncture of the globe or a retro-orbital hematoma. It is important to protect the patient's bones during corticosteroid treatment, especially with a postmenopausal woman, and preferably with a bisphosphonate drug such as alendronate, given at 70 mg once a week (see Chapter 27).

### External Radiation

The value of external radiation to the orbits has been established by some, but not all, controlled trials. \[165\] \[171\] In fact, this treatment is steroid-sparing rather than steroid-replacing therapy and is said by some to work best in combination. \[172\] Whether it is more effective than prednisone therapy is unclear; as a result, the combined therapy has long been advocated. \[172\] The safe administration of highly collimated supervoltage radiation to the retro-orbital space requires experienced personnel. Exophthalmos and ophthalmoparesis are usually affected minimally. \[173\] There is a clear need for a reliable disease marker to monitor the effects of such treatment. A recent trial in mild eye disease failed to show any advantage to this approach. \[174\]

### Orbital Decompression

If glucocorticoid therapy and external radiation do not halt progression of the disease and if loss of vision is threatened either by ulceration or infection of the cornea or by changes in the retina or optic nerve, orbital decompression can be performed by a variety of techniques. \[175\] In some patients, a desire for a nearly complete cosmetic correction may be such that decompression surgery is the only therapy. This procedure usually involves removal of either the lateral wall or the roof of the orbit or resection of the lateral wall of the ethmoid sinus and the roof of the maxillary sinus. \[176\] This operation may cause diplopia, and even in the best of hands corrective muscle surgery may be necessary later.

### An Approach to the Treatment of Orbitopathy

There are no controlled trials to support the suggestion that infiltrative orbitopathy is benefited or that its progression is retarded by total ablation of the thyroid gland, whether surgery, radioiodine, or a combination of the two is used. Hence, we recommend a trial of oral glucocorticoid therapy for patients with severe or progressive orbitopathy. If effective doses cannot be tolerated, a course of external radiation may be attempted if edema predominates.

Along with these major forms of treatment local measures should be employed. Ulceration and infection of the cornea should be treated with antibiotics, lubricants, and protective shields. An attempt to appose the eyelids by means of sutures (tarsorrhaphy) should be performed only by an experienced ophthalmologist because sutures may be torn out and cause scarring.

The management of severe orbitopathy should never be undertaken by the endocrinologist or by the ophthalmologist acting alone. Close, coordinated observation of the effects of medical therapy and the progress of the disease is necessary to determine whether and when surgery is appropriate. Surgery almost invariably halts the
progress of the disease and preserves vision if it is performed in time. This decision is influenced by the ability of the available surgical team because the degree of success of such procedures is proportional to experience.

Treatment of Infiltrative Dermopathy

Treatment of infiltrative dermopathy is necessary as soon as the condition is recognized. The application of a topical, high-potency glucocorticoid preparation with an occlusive dressing may cause regression or disappearance of the lesion. Long-standing untreated dermopathy is more resistant to treatment.
Hyperthyroidism and Thyrotoxicosis in Pregnancy

As discussed earlier, postpartum thyroiditis with transient thyrotoxicosis may occur with some frequency (5% to 10%) during the postpartum period. An overactive thyroid gland, however, is much less common in pregnancy itself. When thyrotoxicosis is present during pregnancy, it is usually more severe and is usually due to Graves’ disease. Difficulty in conception and fetal wastage are increased in women with Graves’ disease, but occasional patients become pregnant despite antecedent untreated hyperthyroidism. More commonly, a woman under treatment for hyperthyroidism becomes pregnant, or hyperthyroidism develops after pregnancy is under way. Whatever the sequence, pregnancy complicates the diagnosis and treatment of hyperthyroidism in Graves’ disease and influences its severity and course.

Diagnosis

Pregnancy and hyperthyroidism are both accompanied by thyroid enlargement, a hyperdynamic circulation, and hypometabolism. Amenorrhea may occur in thyrotoxicosis not associated with pregnancy. In pregnancy, serum T4 and T3 levels are increased by changes in glycoprotein, and thus in both conditions, the total serum T4 and T3 levels are elevated. The most useful laboratory tests in their differentiation are measurement of the serum TSH and free T4 levels. Serum TSH is suppressed in hyperthyroidism during pregnancy, just as it is in nonpregnant patients. However, there is sometimes a modest suppression of TSH (between 0.1 and 0.4 mU/L) during the 8th to 14th weeks of normal pregnancy because of stimulation of the thyroid gland by hCG during this interval. A serum TSH below 0.1 mU/L and an elevated free T4 or free T3 level strongly suggests coexistent hyperthyroidism.

Treatment During Pregnancy

The management of hyperthyroidism during pregnancy can be an even greater problem than the diagnosis; however, pregnancy has an attenuating influence on the hyperthyroid state because of the immunosuppression associated with pregnancy, manifested here by a decrease in the level of thyroid autoantibodies (including levels of TSHRAbs).

Iodine and Beta-Blockers

One of the few clinical situations in which the biologic activity of the TSHRAbs is helpful in predicting its effect on the newborn.

Surgery

Surgery during the last trimester, and probably during the first trimester as well, is not desirable because of the possible induction of premature labor. Although surgery may be successful during the middle trimester, it is best to avoid major surgery during pregnancy if possible.

Antithyroid Drugs

Because antithyroid drug treatment poses no greater risk to the mother or fetus than does surgery and possibly involves less risk, medical therapy is the method of choice. Yet because of the usual improvement in the disease, the dosage of antithyroid drug required to control the disease in the latter phases of pregnancy is generally much less than that required in the same patient when she is not pregnant.

Certain aspects of placental physiology are relevant to the use of antithyroid drugs. Propylthiouracil and methimazole readily cross the placenta. They are concentrated in the fetal thyroid, and in excess quantity can cause goitrous hypothyroidism in the fetus. The administration of as little as 100 to 300 mg/day of propylthiouracil to the mother causes a slight decrease in serum T4 concentration and an elevated TSH level in neonates. The long-term complication of this mild hypothyroidism is unknown but should be kept in mind in view of the observations of reduced childhood intelligence when mothers have increased TSH levels. Although maternal T4 crosses the placenta (as obviously evidenced by infants born normal with congenital hypothyroidism), placental transfer is not efficient and varies from patient to patient. For these reasons, the flux of antithyroid agent to the fetus should be limited by giving the mother the smallest dosage of antithyroid agent that induces a physiologic state consistent with normal pregnancy. The serum free T4 level should be maintained in the upper normal range. However, the concentration of hormone is not as critical as the clinical status of the patient. A modest tachycardia is a physiologic response to the increased metabolic demands of pregnancy; and pulse rates of 90 to 100 beats/minute are well tolerated without evidence of myocardial decompensation during delivery.

In most cases, the daily maintenance dose of propylthiouracil should be 200 mg or less, although maintenance doses up to 600 mg may occasionally be required. Propylthiouracil has been generally preferred to methimazole because of the greater transplacental passage of the latter drug, but both drugs have proven equally safe in millions of pregnancies throughout the world and accumulate equally in breast.
Changes in the Immune Response

Pregnancy induces a variety of immune changes that are responses to the paternal foreign antigens that must not be rejected. These include a T-cell shift from Th1 to Th2 autoimmune responses and an overall decrease in all autoimmune responses as evidenced by marked decreases in thyroid autoantibodies. Following delivery, these immune changes are slowly lost and a return to normal is observed but only after a period of exacerbated autoimmune reactivity in which large increases in T-cell and autoantibody activity occur. It is at this time3 to 9 months post partum that autoimmune thyroid disease recurrence or new onset is seen. The mechanisms behind these changes are not fully understood, but maternal microchimerism has been invoked, among other theories, and appears to be associated with Graves' disease-susceptible HLA haplotypes.

Presentation

A high percentage of women in the 20- to 35-year age group give a history of pregnancy in the 12 months before the onset of Graves' disease. Pregnancy and the postpartum state also apparently influence the course of hyperthyroidism in Graves' disease. Patients in clinical remission during pregnancy are prone to postpartum relapse. In 41 pregnancies in 35 patients in remission, 78% were followed by development of thyrotoxicosis during the postpartum period. The patients with Graves' disease and postpartum thyrotoxicosis were classified into three categories:

1. Some patients had persistent recurrent hyperthyroidism with an elevated RAIU (classic Graves' disease).
2. Some had a transient disorder associated with a normal or an elevated RAIU (transient Graves' disease).
3. Some patients, especially those with the highest titers of TPOAb, experienced a transient thyrotoxicosis with a decreased RAIU that is the thyrotoxic phase of postpartum thyroiditis. This phase, in turn, may be followed by a hypothyroid phase (see later).

The Desire for Pregnancy

A special problem related to hyperthyroidism and pregnancy is presented by the patient who is in early remission after a course of antithyroid drug treatment or is being treated with antithyroid agents and wants to become pregnant in the near future. Management with antithyroid agents can be continued through pregnancy or reinstituted should hyperthyroidism recur, but in such instances definitive therapy (radioiodine or surgery) should be considered to forestall the complexities of managing hyperthyroidism during pregnancy. As with the therapy of Graves' disease in general, such decisions must involve education of the patient so that the risks and benefits of the various alternatives are clearly appreciated.

Nursing and Antithyroid Drugs

Older studies suggested that relatively more methimazole than propylthiouracil appeared in breast milk of women receiving these drugs, but more recent evidence shows little difference between them (see earlier). However, it is generally recommended that women who take antithyroid drugs should be advised not to nurse their infants because of the difficulty in monitoring young babies. No serious drug side effects have been reported in neonates whose mothers were taking antithyroid drugs, although periodic thyroid function tests would seem appropriate if the mother continues to breast-feed.
Treatment of Graves' Disease in Children and Adolescents

Thyrotoxicosis in childhood and adolescence is almost always the result of Graves' disease. Thyrotoxicosis in this age group is worthy of special consideration because treatment is less satisfactory than in adults. Hence, there is more uncertainty concerning its management, probably because the disease tends to be more severe in children.

Radioiodine

For several reasons, we do not often use radioiodine in the treatment of childhood thyrotoxicosis. At least three factors weigh against such use:

1. The enhanced carcinogenic potential of radiation in the thyroid gland of the infant or child is evidenced by the correlation between childhood thyroid carcinoma and a history of radiation therapy to the head, neck, or chest in childhood and the increased incidence of thyroid cancer in children exposed to radiation from the Chernobyl nuclear accident.
2. Among patients with thyrotoxicosis, those treated in childhood or adolescence are thought to be at greatest risk for transmitting genetic damage, although available data suggest that this may not be likely.
3. Postradioiodine hypothyroidism is a particularly undesirable complication in young children because inadequate or interrupted therapy can impair growth, development, and scholastic performance.

Antithyroid Drugs

The choice between destructive and antithyroid drug therapy may be a difficult one. The data indicate that children have a lower incidence of long-term remission after antithyroid therapy than adults, although some believe that thyrotoxicosis often undergoes remission after adolescence. A course of 1 to 2 years of antithyroid therapy seems reasonable, and a second course of antithyroid therapy is regularly employed if recrudescence or relapse occurs after the first course. Measuring TSHRAs may be helpful in assessing the child's progress. If sustained remission does not follow a second course of therapy and particularly if the patient has passed through adolescence during this period, radioiodine therapy or surgery may be considered.

Surgery

Most surgical series reveal a relatively high frequency of postoperative hypothyroidism in children. Recurrences are also more frequent, presumably as a result of attempts to avoid hypothyroidism. Complications such as hypoparathyroidism and recurrent laryngeal nerve damage must be borne over a long life span. All this leads to surgery's being less appropriate than antithyroid drugs for most children.
OTHER CAUSES OF THYROTOXICOSIS

Toxic Multinodular Goiter

Toxic multinodular goiter is a disorder in which hyperthyroidism arises in a multinodular goiter, usually of long standing, and is the result of one of several pathogenetic factors. It is important to avoid the term toxic nodular goiter because this confuses toxic multinodular goiter, as here described, with a toxic adenoma of the thyroid gland (see later).

Pathogenesis

The pathogenesis of toxic multinodular goiter cannot be considered apart from that of its invariable forerunner, nontoxic multinodular goiter, from which it emerges slowly and surreptitiously. Two hallmarks of the disorderstructural and functional heterogeneity and functional autonomyrevolve over time; the increase in the extent of autonomous function causes the disease to move from the nontoxic to the toxic phase, but the mechanisms of this change in all cases are uncertain. The somatic mutations in the TSHR gene demonstrated in toxic adenomas have been demonstrated in some cases of toxic multinodular goiter and appeared to differ from nodules to nodules. However, only about 60% of toxic nodules have TSHR mutations, and only a few have G protein mutations. Hence, there are many nodules with undetermined causes of their autonomy.

Studies show that radioiodine becomes localized in one or more discrete nodules, whereas iodine accumulation in the remainder of the gland is usually suppressed. Radioiodine is of little help because this confuses toxic multinodular goiter, as here described, with a toxic adenoma of the thyroid gland (see later).

The overproduction of thyroid hormone in toxic multinodular goiter is usually less than that in Graves' disease. First, the clinical manifestations of thyrotoxicosis are rarely flagrant. Second, the serum 

\[ T_4 \] and 

\[ T_3 \] concentrations may be only marginally increased, and a suppressed serum TSH level may be the only abnormality.

Finally, the total RAU is only slightly increased or within the normal range.

The mildness of the hyperthyroidism is consistent with either of its presumed pathogenetic origins. The effectiveness of any stimulus to hyperfunction may be blunted in a thyroid gland that is the seat of a preexisting nontoxic multinodular goiter because of the associated impairment in the efficiency of hormone synthesis. Toxic multinodular goiter is a common complication of nontoxic multinodular goiter, but its precise incidence is unknown. It usually occurs after the age of 50 years in patients who have had nontoxic multinodular goiter for many years. Like its forerunner, toxic multinodular goiter is many times more common in women than in men. Sometimes hyperthyroidism develops abruptly, usually after exposure to increased quantities of iodine, which permits autonomous foci to increase hormone secretion to excessive levels and which may simply exacerbate already established mild hyperthyroidism (iodine-induced hyperthyroidism, von Basedow's disease). In addition, Graves' disease may either present or develop in a multinodular gland, as confirmed by the presence of TSHRAbS of the stimulating variety.

Toxic multinodular goiter is almost never accompanied by infiltrative ophthalmopathy, and when the two coexist, it represents the emergence of Graves' disease. The clinical manifestations of toxic multinodular goiter also differ from those in Graves' disease. Cardiovascular manifestations tend to predominate, possibly because of the age of the patients, and include atrial fibrillation or tachycardia, with or without heart failure. Surveys have indicated that TSH was suppressed in 26% of elderly patients with atrial fibrillation (see also Table 11-2).

A decreased response to digitalis may alert the physician to the presence of thyrotoxicosis. Weakness and wasting of muscles are common. The nervous manifestations are less prominent than in younger patients with thyrotoxicosis, but emotional lability may be pronounced. Because of the physical characteristics of the thyroid gland and its frequent retrosternal extension, obstructive symptoms are more common than in Graves' disease. On palpation, the characteristics of the goiter are the same as those of the more common nontoxic multinodular goiter (see later). In as many as 20% of elderly patients with thyrotoxicosis, the thyroid gland is firm and irregular but not distinctly enlarged. Ultrasonographic examination confirms the diagnosis as toxic multinodular goiter rather than a single toxic adenoma or Graves' disease.

### Laboratory Tests and Differential Diagnosis

The challenge to determine whether the patient with a multinodular goiter is thyrotoxic can be resolved only with laboratory tests. If the free 

\[ T_4 \] index or free 

\[ T_3 \] index is elevated and the TSH level is suppressed, the diagnosis of hyperthyroidism is established. TSH levels intermediate between 0.1 and the 0.5 mIU/L lower limit of normal are not usually associated with significant symptoms. Such patients have thyroid autonomy but are not thyrotoxic. The pituitary-hypothalamic axis provides the most sensitive indicator of the level of thyroid hormone that is specifically relevant to the individual patient. Monitoring the concentration of serum TSH takes advantage of this sensitivity and is one of the most useful ways of establishing the existence of autonomous thyroid function. The RAU is of little help because thyrotoxicosis may exist in association with values that are normal or only slightly increased.

### Treatment

Radioiodine

Radioiodine is the treatment of choice for patients with toxic multinodular goiter despite disagreement about the size and number of doses required to achieve a therapeutic response. Along the eastern seaboard of the United States, the responsiveness of toxic multinodular goiter to radioiodine may differ little from that of the diffuse toxic goiter of Graves' disease. However, in areas where goiter was formerly endemic, such as the Great Lakes area of the United States, toxic multinodular goiter is said to be more resistant to radioiodine. The more resistant variety of toxic multinodular goiter may be a reflection of low radioiodine uptake for a

| Table 11-9 – Atrial Fibrillation and Thyrotoxicosis |
|---------------------------------|-----------|
| Total no. of patients examined | 443       |
| Euthyroid patients             | 303       |
| Hypothyroid patients           | 23        |
| Hyperthyroid patients          | 117 (26.4%) |

variety of reasons ranging from increased iodine consumption to low sodium-iodide symporter (NIS) activity.

Because of the age of the patient and variations in sensitivity to radioiodine, increased doses of RAI are often administered. These doses are likely to be larger than those used in Graves’ disease because the uptake of $^{131}$I tends to be lower and the gland larger. Many patients with this disorder have underlying heart disease. Therefore, the administration of radioiodine should be preceded by a course of antithyroid therapy until a eumetabolic state is achieved. Medication is then discontinued for at least 3 days before radioiodine is administered. Seven days thereafter, the antithyroid drug is reinstituted so that the thyrotoxicosis is controlled until radioiodine takes effect. After 6 to 8 weeks, the antithyroid drug is gradually withdrawn; if thyrotoxicosis recurs, a second course of therapy is given. This entire treatment sequence should be accompanied by adequate $\beta$-blockade if the cardiac status permits.

**Surgery**

Surgical therapy is often recommended after adequate preoperative preparation in patients with obstructive manifestations or when it is feared that such manifestations may result from the temporary thyroid enlargement that radioiodine sometimes produces, particularly in patients with retrosternal extensions of the goiter. In these patients, MRI is recommended to define the extent of the goiter and the adequacy of the tracheal walls. Respiratory function studies may also be helpful in assessing the need for surgery. When surgery is contraindicated, even significant obstructive symptoms can be relieved by adequate radioiodine therapy.
Toxic Adenoma (Plummer's Disease)

A third, less common form of hyperthyroidism is caused by one or more autonomous adenomas of the thyroid gland. As herein employed, the term toxic adenoma refers to a tumor in a thyroid that is otherwise intrinsically normal. The disorder is usually caused by a single adenoma that is palpable as a solitary nodule and therefore is sometimes referred to as hyperfunctioning solitary nodule or toxic nodule. Occasionally, two or three adenomas of similar character are present.

Pathogenesis

Toxic adenomas are true follicular adenomas (for histopathologic characteristics, see Chapter 13). The basic pathogenesis

Figure 11-23 (Figure Not Available) This diagram shows the thyrotropin (TSH) receptor and its ectodomain, transmembrane loops, and intracellular segment and illustrates activating and inactivating mutations of the TSH receptor. The amino acids are indicated by the single-letter code and numbered consecutively, starting with the transcription initiation codon. The vertical lines indicate exon boundaries. (From Sunthornthepvarakul T, Gottschalk ME, Hayashi Y, et al. Resistance to thyrotropin caused by mutations in the thyrotropin-receptor gene. N Engl J Med 1995; 332:1551-1557.)

of a large fraction of them is one of several somatic point mutations in the TSHR gene, commonly in the third transmembrane loop. These single nucleotide substitutions cause amino acid changes that lead to constitutive activation of the TSH receptor in the absence of TSH (Fig. 11-23) (Figure Not Available). It appears therefore that the TSHR is “tripped” from an off state to an on state. Similarly, loss of function rather than gain of function mutations may also occur in the TSHR gene and may cause hypothyroidism (see later). A small number of autonomous adenomas have mutations in the G protein genes that lead to a similar state of constitutive activation.

The course is one of progressive growth and increasing function over many years. At first, the adenoma may be present as a small nodule or may be palpable; in either case, it can be detected in a radioidine thyroid scan as a localized area of increased radioidine accumulation (Fig. 11-24) and much of the remainder of the gland may be suppressed. With further growth, a progressively increasing share of glandular function is assumed by the adenoma, with the result that the remaining tissue is increasingly suppressed. Ultimately the remainder of the gland is completely suppressed and atrophic.

![Image 11-24](https://example.com)

**Figure 11-24** Scanning of a hyperfunctioning hot nodule corresponding to physical examination with a faint outline of the remaining suppressed gland. In this unusual case, Graves' disease developed a few months later after an oral contrast agent load. (From Soule J, Mayfield R. Graves' disease after 131I therapy for toxic nodule. Thyroid 2001; 11:919-2.)

and the thyroid scan reveals function only in the adenoma (hot nodule).

Although continued growth of the adenoma causes secretion of excessive quantities of hormone, some time may pass before thyrotoxicosis becomes overt. The extranodular tissue usually retains its capacity to function if TSH is provided, either by exogenous administration or by ablation of the nodule.

Clinical Picture

Toxic adenoma occurs in a younger age group than does toxic multinodular goiter, often in patients in their 30s or 40s. Frequently there is a history of a long-standing, slowly growing lump in the neck. It is unusual for adenomas to produce thyrotoxicosis until they have achieved a diameter of 2.5 to 3 cm. The adenoma can undergo central necrosis and hemorrhage; as a result, the thyrotoxicosis may be relieved, the remainder of the thyroid gland may resume its function, and the adenoma may appear on a scan as a cold area, suggesting a thyroid carcinoma. Calcification in the area of hemorrhage may take place and may be evident on sonogram examination. Such calcification is usually gross and irregular and does not resemble the finely stippled calcification of the psammoma bodies seen in papillary cancers. If the TSH level is suppressed and the single nodule is seen by sonography, a thyroid radioidine uptake and scan would be necessary only as a prelude to ablative treatment.

The peripheral manifestations of toxic adenoma are generally milder than those of Graves' disease and are notable for the absence of infiltrative orbitopathy and myopathy: cardiovascular manifestations may be prominent. The nodule is usually felt as a smooth, well-defined, round or ovoid mass that is firm and moves freely on swallowing. Often the remainder of the gland is not palpable. A bruit is never present.

Laboratory Tests

The results of laboratory tests depend on the stage of the disorder. At first, serum thyroid hormone concentrations may be normal except for borderline suppression of the serum TSH. This, together with ultrasonographic examination to exclude multiple nodules, confirms the diagnosis. Later a thyroid scan may show localization of radioisotope in the palpated nodule, but this does not occur until TSH secretion is suppressed. If the nodule continues to grow, frank hyperthyroidism is accompanied by elevation of serum thyroid hormone levels and metabolic indices. When the nodule is small, the RAIU is normal but cannot be suppressed completely by exogenous thyroid hormone. However, TSH and, therefore, function in the extranodular tissue are suppressed by exogenous hormone, allowing identification of the autonomous nature of the nodular lesion by scanning even before the lesion has become sufficiently large to suppress serum TSH.

Occasionally, values for serum T3 concentration are normal, and only the serum T4 level is increased (T3 toxicity). Relative to its overall rate of occurrence, toxic adenoma is the most frequent cause of T3 toxicity. If there is any question about the presence of the suppressed lobe, exogenous recombinant human TSH may be administered before scanning to demonstrate uptake in this tissue. Incidental thyroid carcinoma may rarely coexist within a gland exhibiting a hyperfunctioning adenoma, and malignant nodules that cause functional hyperthyroidism are rare.

Treatment

Although many hyperfunctioning adenomas eventually cause clinical hyperthyroidism, some do so slowly and others not at all. Therefore, treatment of asymptomatic patients with functional adenomas is decided on an individual basis. The degree of TSH suppression is an index of the progression of thyroid hormone production by the adenoma. Suppression below the lower limits of normal indicates that hyperthyroidism is present and that therapy should be given except in unusual situations.

Two therapies are available: radioidine and surgery.

Radioidine

In terms of the specificity of treatment, functioning thyroid nodules should be ideal candidates for radioidine therapy. The radiation should, in theory, be directed almost exclusively to the diseased tissue. This is because TSH levels are suppressed and the normal thyroid tissue surrounding the nodule does not take up
radioiodine. However, this suppression may be incomplete, resulting in uptake of radioiodine, and thyroid failure may develop in many patients. For patients older than 20 years of age with a nodule 3 cm in diameter or smaller, $^{131}$I is an appropriate treatment if the risk of eventual hypothyroidism is acceptable.

In general, higher doses of radioiodine are required than in Graves' disease, namely 185 to 370 MBq (5 to 10 mCi) deposited at 24 hours. Because of the potential for hypothyroidism with higher $^{131}$I doses, prolonged follow-up is mandatory. Suppression of TSH by exogenous $T_3$, 25 µg/day for 7 days, may be used in appropriate patients to reduce $^{131}$I uptake by the normal thyroid tissue during therapy.

Surgery

Patients with large nodules accompanied by physical symptoms are most readily treated with surgical excision. Surgical excision is also often used in young patients. The toxic adenoma is not diffusely hypervascular, and consequently preoperative preparation with iodine is not required. In patients with overt thyrotoxicosis, however, a normal metabolic state should be restored with an antithyroid drug or -blockers before surgery.
Inherited Nonautoimmune Autosomal Dominant Hyperthyroidism

Toxic diffuse thyroid hyperplasia without the pathologic characteristics of autoimmune disease has been reported in a few families and appears to be inherited as an autosomal dominant condition. Polymorphic genomic mutations in the TSHR gene have been reported to cause constitutively activated TSHRs differing from family to family. Recessive mutations on both chromosomes have also been described as causing hyperthyroidism while the parents remained euthyroid. These gain of function mutations, mostly in the transmembrane regions of the TSHR, are similar to those somatic mutations seen in toxic adenomas but are in the germline (see Fig. 11-23) (Figure Not Available). Treatment is by radioiodine ablation or thyroidectomy, depending on the age of the patient.
Transient Hyperthyroidism

Subacute Thyroiditis

Subacute thyroiditis has been termed granulomatous giant cell or de Quervain’s thyroiditis. It is thought to be caused by a viral infection of the thyroid gland and often follows an upper respiratory illness. A tendency to a seasonal and geographic aggregation of cases has been noted. The mumps virus has been implicated in some cases, and coxsackievirus, influenza virus, echoviruses, and adenoviruses may also be etiologic agents. Evidence of thyroid autoimmunity is often present during the active phase of the disease. This autoimmunity is usually transitory, although some patients may retain evidence of thyroid autoimmunity for many years.

A small number of patients eventually develop autoimmune thyroid disease. It is likely that subacute thyroiditis is the clinical phenotype for a wide variety of virus infections that affect the thyroid gland. Subacute thyroiditis is uncommon, but mild cases may be mistakenly diagnosed as pharyngitis. Women are more frequently affected than men, and the peak incidence is in the fourth and fifth decades.

Histopathology

The histopathologic changes (Fig. 11-25) differ from those in Hashimoto’s disease. The lesions are patchy in distribution and vary in their stage of development from area to area. Affected follicles are infiltrated predominantly with mononuclear cells and show disruption of epithelium, partial or complete loss of colloid, and fragmentation and duplication of the basement membrane. To this extent, the histopathologic appearance may resemble that in Hashimoto’s disease.

A characteristic feature is the well-developed follicular lesion that consists of a central core of colloid surrounded by the multinucleate giant cells, from which stems the designation giant cell thyroiditis. Colloid may be found in the interstitium or within the giant cells (colloidophagy). The follicular changes progress to form granulomas. Interfollicular fibrosis and an interstitial inflammatory reaction are present to varying degrees. When the disease subsides, an essentially normal histologic appearance is restored.

Pathophysiology

Apoptosis of follicular epithelium and loss of follicular integrity are the primary events in the pathophysiology. Tg, preformed hormone, and abnormal iodinated materials are released into the circulation, often in quantities sufficient to elevate not just the serum Tg level but also the serum T4 and T3 concentrations, produce clinical thyrotoxicosis, and suppress TSH secretion. As a result of the last effect, all thyroid function is suppressed, the RAIU decreases to low levels, and hormone synthesis ceases. Destruction of the follicular epithelium is the primary contributor to lowering of the RAIU and disruption of hormone synthesis, because external TSH may fail to increase the RAIU normally. Later in the disease, when stores of preformed hormone are depleted, serum T4 and T3 concentrations decline, sometimes into the hypothyroid range, and the serum TSH level rises, often to elevated values. As the disease becomes inactive, the RAIU may be greater than normal for a time as hormone stores are repleted. Ultimately, when hormone secretion resumes, serum T4 and T3 concentrations rise, and serum TSH concentration decreases to normal values.

Clinical Picture

The characteristic feature is the gradual or sudden appearance of pain in the region of the thyroid gland with or without fever. The pain, which is aggravated by turning the head or swallowing, characteristically radiates to the ear, jaw, or occiput and may mimic disorders arising in these areas. The absence of pain does not exclude the diagnosis, because biopsy-proven painless, subacute thyroiditis occurs, but it must be distinguished from autoimmune thyroiditis. Hoarseness and dysphagia may be present, and patients may complain of palpitation, nervousness, and lassitude. Lassitude can be extreme, considering the local nature of the disease, and indicates a systemic component. Although acute manifestations are present in severe cases, in milder disease, which is often wrongly diagnosed, symptoms may be present for months.

On palpation, at least part of the thyroid gland is slightly to moderately enlarged, firm, even nodular, and usually exquisitely tender, one lobe frequently being more severely affected than the other. Indeed, the symptoms may be truly unilateral. The overlying skin may be warm and red. Occasionally the locus of maximal involvement migrates over the course of a few weeks to other parts of the gland.

The disease usually subsides within a few months, leaving no residual deficiency of thyroid function, but often passes through a transient phase of hypothyroidism, resembling the syndrome of transient silent autoimmune thyroiditis preceded by transient thyrotoxicosis. In rare cases, the disease may smolder, with repeated exacerbations over many months and with hypothyroidism sometimes being the final result.

Laboratory Tests

The laboratory findings vary with the phase of the disease. During the active phase, the erythrocyte sedimentation rate (ESR) can be increased to a remarkable extent. Indeed, a diagnosis...
of active subacute thyroiditis is hardly tenable when the ESR is normal. The leucocyte count is normal or, at most, moderately increased. The serum Tg level is characteristically high, in keeping with the degree of thyroid destruction.

Subacute thyroiditis is one of several causes of low-uptake thyrotoxicosis; the others are so-called silent thyroiditis (the early phase of Hashimoto’s disease [described earlier], thyrotoxicosis factitia, and iodine-induced hyperthyroidism. For reasons described earlier, the RAU is subnormal, despite the presence of normal, or often elevated, values of serum T₄ and T₃ concentrations. At this point in the course, basal serum TSH levels are suppressed. In the typical patient, TPO and Tg autoantibodies either are not detectable or are present in low levels. In milder cases, some uptake of radioiodine may persist in unaffected portions of the gland, as revealed by a radioiodine scan; however, this is unusual, and a diagnosis of active, subacute thyroiditis should be viewed with suspicion if the RAU is normal.

In the hypothyroid phase, serum T₄ and T₃ concentrations are low and the serum TSH concentration is appropriately elevated (see Fig. 11-26). With recovery, the RAU returns to normal or high levels, and values for serum T₄ and T₃ concentrations are restored to normal.

Differential Diagnosis

Subacute thyroiditis must be differentiated from (1) acute hemorrhagic degeneration in a preexisting thyroid nodule, (2) Hashimoto’s disease of acute onset, and (3) acute pyogenic thyroiditis.

Differentiation from hemorrhage into a nodule presents no difficulty when this occurs in a multinodular goiter, because other nontender nodules can be felt. Detection is more difficult when there is hemorrhage into a solitary nodule, but this should be easily seen on ultrasonography. In both varieties of hemorrhage, function in the remainder of the gland persists, and the ESR is rarely elevated.

Hashimoto’s disease of acute onset may be accompanied by pain and tenderness in the thyroid gland, but the gland usually is diffusely affected and may present as a truly painless thyroiditis with thyrotoxicosis and a decreased RAU but with a histologic picture of autoimmune thyroiditis and no giant cells, often termed hashitoxicosis. This may be difficult to distinguish from painless, subacute thyroiditis. Lack of elevation of the ESR and high titers of thyroid autoantibodies strongly suggest the former.

Acute pyogenic thyroiditis is distinguished by the presence of a septic focus elsewhere, by a greater inflammatory reaction in the tissues adjacent to the thyroid gland, and by much greater leukocytic and febrile responses. The RAU is usually preserved in acute pyogenic thyroiditis. Rarely, widespread infiltrating thyroid cancer can present with a clinical and laboratory picture almost indistinguishable from that of subacute thyroiditis. 

Treatment

In mild cases, aspirin, nonsteroidal anti-inflammatory drugs, or cyclooxygenase-2 inhibitors generally control the symptoms. In more severe cases, glucocorticoids (e.g., prednisone up to 40 mg/day) alleviate the manifestations but do not influence the underlying disease process. Hence, the symptoms may be exacerbated if treatment is withdrawn too early but do again respond if treatment is reinstituted. The chance of relapse may be minimized if glucocorticoid therapy is continued at a dose that maintains the patient in an asymptomatic state until the RAU has returned to normal. Thyroid hormone replacement therapy may decrease the size of the gland by suppressing TSH and by relieving the pressure on the thyroid capsule. Because TSH is needed for thyroid cell regeneration, however, such therapy should be decreased as the symptoms subside.

Thyrotoxicosis in Silent Thyroiditis

Thyrotoxicosis is associated with the early phase of subacute thyroiditis in both painful and painless variants. In addition, thyrotoxicosis can also occur without pain in early autoimmune thyroiditis (Hashimoto’s disease), in which biopsy of the thyroid gland reveals the histopathologic changes of Hashimoto’s disease rather than those of subacute thyroiditis (see Fig. 11-27). This syndrome has variously been alluded to as silent thyroiditis with thyrotoxicosis, hyperthyroiditis, or hashitoxicosis, and cannot be distinguished from the early phase of postpartum thyroiditis (see later).

The cardinal features are thyrotoxicosis associated with depressed values of the RAU in the absence of excess body iodide stores, lack of pain or tenderness in the thyroid area, and spontaneous resolution of the thyrotoxic phase of the disease. There is a tendency to pass through a transient euthyroid phase and then a hypothyroid phase before a long-term return to euthyroidism and a tendency for the syndrome to recur. The thyroid gland is enlarged in only about 50% of cases, and enlargement is usually mild and unaccompanied by nodularity. Thyrotoxicosis is usually mild, and this is reflected in the extent of elevation of serum T₄ and T₃ levels. High levels of TPO autoantibodies can be detected in most patients by sensitive assays. Systemic manifestations of inflammation are lacking, and the ESR is normal or nearly normal.

Several aspects of the pathophysiology of this disorder are instructive. As in subacute thyroiditis, because of widespread apoptosis and TSH suppression the rate of ongoing synthesis of thyroid hormones is negligible, justifying the classification of this disorder among those that lead to “thyrotoxicosis without hyperthyroidism.” Decreased values of the RAU are due partly to suppression of TSH secretion by the excess of circulating hormones because the serum TSH level is suppressed; function of the thyroid follicular cell is also impaired, however, because the RAU does not increase after administration of TSH, presumably secondary to T cellmediated and antibody-mediated thyroid cell death.

The tendency of the disorder to pass through a hypothyroid phase is not surprising, in view of the extensive depletion of glandular hormone stores that occur while hormone is leaking from the gland and new hormone synthesis is impaired.

The duration of the thyrotoxic phase averages about 1 to 2 months. About 50% of patients return to a euthyroid phase and remain well, at least for some time. In the remaining 50%, a hypothyroid phase may follow and last from 2 to 9 months. In most instances, euthyroidism is eventually restored, but permanent hypothyroidism may develop in some patients years later. About one-third of patients retain a goiter, usually with persistence of thyroid autoantibodies in the serum. The opposite sequelae of thyrotoxicosis may also occur months or years after restoration of a euthyroid state, and some patients experience multiple recurrences.

Treatment of the thyrotoxic phase consists of alleviation of the peripheral manifestations through the use of β-blockers or sedatives. Reportedly, prednisone (30 to 50 mg/day) decreases the duration of the thyrotoxic phase without the risk of relapse on its withdrawal but is not needed. If the hypothyroid phase is mild and brief, patients may not require treatment. When treatment with levothyroxine is needed, it should be undertaken with the understanding that it will be withdrawn approximately 6 months later, because the hypothyroidism is unlikely to be permanent.

The underlying nature of the disorder is an autoimmune
dysregulation. Extensive lymphocytic infiltration and the presence of plasma cells within the thyroid are similar to those seen in more classic Hashimoto's thyroiditis, as are the circulating thyroid autoantibodies. The latter, however, may merely reflect a response to the inflammatory release of antigens. The occurrence of the syndrome in patients known to also have Graves' disease, which we and others have observed, and the later emergence of hypothyroidism (Hashimoto's disease) are also consonant with an autoimmune etiology. On the other hand, the absence of high levels of circulating antithyroid antibodies in many patients and the permanent resolution indicate that the immune system regains its equilibrium in these patients. This provides a great opportunity to further our understanding of these control mechanisms.

Thyrotoxicosis in Postpartum Thyroiditis

The postpartum thyroiditis syndrome is similar in presentation, course, and pathophysiology to silent thyroiditis. Transient thyrotoxicosis with low RAIU may develop within 3 to 6 months after delivery and is often followed by a period of hypothyroidism of several months' duration and an eventual return to a euthyroid state. In some patients, only a hypothyroid phase is evident. The incidence of postpartum thyroiditis varies geographically but may occur in as many as 10% of women and in more than 30% of those with positive TPO autoantibodies. This argues for prenatal assessment for the presence of TPO antibodies. In women found to be positive for TPO antibodies, postpartum assessment of thyroid function is recommended at 2, 4, 6, and 12 months.

As in the similar syndrome of silent thyroiditis, not temporally related to pregnancy, recurrences are common after subsequent pregnancies. Most patients have a small goiter and positive tests for TPO antibodies, although levels may be low. The syndrome has also been observed post partum in patients known to have prepartum Graves' disease. There is a strong association with the HLADR3 and HLADR5 haplotypes, which are also associated with autoimmune thyroid disease. The occurrence of the disorder after delivery is probably due to a rebound of immune activity after its suppression during pregnancy. However, the role of fetal microchimerism has also been implicated in the initiation of disease.
Hyperthyroidism Caused by Thyrotropin or Thyrotropin Receptor Agonists

Rarely hyperthyroidism results from hypersecretion of TSH or TSH-like activity because of three causative factors:

1. A TSH-secreting pituitary adenoma.
2. Inappropriate hypersecretion of TSH secondary to localized pituitary resistance to thyroid hormones or increased secretion of TRH.
3. Excessive secretion of hCG from trophoblastic tumors.

All varieties are associated with a diffuse, hyperfunctioning goiter. Features of autoimmune thyroid disease are absent in these patients and in the families of patients.

When TSH is the cause, the serum TSH concentrations are not suppressed at a time when serum free \( T_4 \) or \( T_3 \) concentrations are elevated. \(^{234,235}\)

Thyrotropin-Secreting Pituitary Tumors

In the adenomatous variety, a mass lesion is present in the pituitary gland. (see Chapter 8). The concentration of free subunits of TSH in serum is elevated, and serum TSH concentrations may fall to increase after TRH administration. In patients with nonadenomatous TSH hypersecretion, in contrast, subunits are not present in the blood in high concentrations and the response to TRH is usually normal. \(^{236,237}\) In addition, patients with thyroid hormone resistance may respond to oral \( T_3 \) rather than \( T_4 \).

Patients with excess TSH in the absence of resistance present a difficult therapeutic problem. In some cases, TSH secretion can be suppressed if somewhat large doses of thyroid hormone are administered, but this results in worsening of the thyrotoxicosis. Hyperthyroidism can be controlled, of course, by thyroid ablation, but serum TSH levels then increase still further, raising the question as to whether a TSH-producing adenoma may ultimately develop. Bromocriptine, a dopamine agonist, may suppress TSH secretion and alleviate the hyperthyroidism in this disorder. \(^{238}\) Somatostatin analogues have also been used. However, TSH-producing tumors usually require surgical resection. \(^{239}\) \(^{240}\) \(^{241}\) Treatment with 3,5,3'-triiodothyroacetic acid has also been successful. \(^{242}\)

The occurrence of TSH-induced hyperthyroidism further supports the argument that a serum TSH concentration should be measured as part of the initial work-up of every patient who is hyperthyroid and has a diffuse goiter. The remote possibility that a patient with Graves’ disease might have an artifactual elevation of TSH concentration because of a heterophilic antibody cross-reacting with mouse immunoglobulin (see Chapter 6) must be kept in mind. \(^{243}\) For some sera, the use of a different assay kit may confirm the elevated level. Alternatively, mouse serum may be added to the assay tube to absorb the heterophile antibody.

A Note on Pituitary Resistance to Thyroid Hormone

In some patients with inherited thyroid hormone resistance due to mutations in the thyroid hormone receptor, the hypothalamic pituitary feedback mechanism may be more resistant to the effects of thyroid hormone compared with peripheral tissues, such as the heart. \(^{244}\) \(^{245}\) \(^{246}\) These patients may have a hyperthyroid appearance with tachycardia, nervousness, and goiter associated with an elevated free \( T_4 \) index. Because the thyroid hormone hyperproduction is TSH-driven, however, serum TSH concentrations are detectable (>0.1 mU/L) or even elevated inappropriately for the circulating thyroid hormone levels (see Chapter 12).

In general, the manifestations are due not to excessive but to inadequate thyroid hormone action, and these individuals may require treatment with thyroid hormone or thyroid hormone analogues and \(-\)adrenergic receptor-blocking agents rather than antithyroid drugs. \(^{247}\) \(^{248}\) This argues again for the appropriateness of at least one serum TSH measurement in every hyperthyroid patient because this is the only way that an accurate diagnosis can be achieved (see discussion of thyroid hormone resistance in topic of hypothyroidism). Rarely, families with just pituitary resistance to thyroid hormone (PRTH) may respond to treatment with liothyronine. \(^{249}\)

Hyperthyroidism in Trophoblastic Disease

Thyroid hyperfunction often accompanies hydatidiform mole, choriocarcinoma, or metastatic embryonal carcinoma of the testis. Such neoplasms, particularly hydatidiform mole, elaborate differentially glycosylated hCG molecules that exhibit crossover specificity for binding to the TSH receptor and can induce thyroid overactivity. \(^{250}\) \(^{251}\) \(^{252}\) \(^{253}\) Some patients have clinically overt thyrotoxicosis; however, clinical manifestations are usually not prominent, and goiter is absent or minimal despite laboratory evidence of a hyperthyroid state. Free \( T_4 \) or free \( T_3 \) levels are increased, and TSH values are suppressed. The reason for this discordance between the clinical and the laboratory indices is not known but the discordance may be due to the relatively short duration of thyroid hormone excess. The possibility of a molar pregnancy should always be considered in a young woman with thyrotoxicosis, because appropriate therapy is evacuation of the uterus.
Iodine-Induced Hyperthyroidism

Administration of supplemental iodine to subjects with endemic iodine deficiency goiter can result in iodine-induced hyperthyroidism and even Graves’ disease. This response, termed iodine-induced hyperthyroidism or the Jod-Basedow effect, occurs in only a small fraction of individuals at risk. The best-studied experience has been in Tasmania, where a temporary increase in thyrotoxicosis occurred shortly after the addition of small quantities of iodine to bread as a means of correcting iodine deficiency.

Studies revealed two major patterns of underlying thyroid disorder. In the first, especially common in older individuals, nodular goiter with areas of autonomous function were present and TSHRAs of the type found in Graves’ disease were not detectable in the blood. In the second pattern, which occurred in younger individuals with diffuse goiter, stimulating TSHRAs were often present. These findings suggest that the Jod-Basedow effect occurs only in thyroid glands in which function is independent of TSH stimulation. The occurrence of the Jod-Basedow effect should not be construed as a reason for failing to treat endemic iodine deficiency. Apart from the many other benefits that accrue from iodine treatment and prophylaxis, over the long run the frequency of spontaneous hyperthyroidism associated with the development of autonomous nodules is diminished.

Iodine-induced hyperthyroidism is an important disorder in areas of the world where dietary iodine intake is high. In regions where iodine intake is marginal but overt iodine deficiency is absent, moderate increments in iodine intake may induce hyperthyroidism in patients with autonomous thyroid nodules, and large pharmacologic doses of iodine, can do so in geographic areas where the iodine intake is more than adequate. Consequently, the physician must be alert to the possibility of inducing hyperthyroidism when administering large quantities of iodine in expectorants, radiographic contrast media, medications containing iodine (e.g., amiodarone), or any other form to patients with nodular goiter. Because nodular goiter is generally a disease of older people, induction of the Jod-Basedow phenomenon can have serious consequences, because enrichment of the thyroid gland with iodine forestalls administration of $^{131}$I and delays the response to antithyroid agents. In these patients, serum $T_3$ concentration is sometimes normal, although total and free $T_4$ concentrations are increased and TSH is suppressed. Confirmation that the patient has been exposed to large quantities of iodine can be obtained by demonstrating that the RAIU is low and urinary iodine excretion increased (more than several milligrams per day).

The treatment of these individuals may be difficult. Even after discontinuation of exogenous iodide, uptake of $^{131}$I by the thyroid gland may remain low, not adequate for conventional doses of radiiodine. The elevated thyroid hormone content also makes thionamide drugs less effective. In some cases, it may be necessary to treat such individuals for prolonged periods (6 to 9 months) before administering radiiodine therapy; if uptake is detectable, however, larger doses of radiiodine may be given to destroy thyroid tissue.

**Amiodarone**

Amiodarone is an iodine-rich drug that has become increasingly popular because of its effectiveness in combating severe cardiac arrhythmias. The drug has complex effects on the thyroid gland, although most patients (80%) remain euthyroid. Structurally, the drug resembles $T_3$ (Fig. 11-30) and contains 37% iodine. It has been estimated that a huge amount (200 mg/day) of iodide is made available in a daily dose of 600 mg. It has a half-life of 50 to 60 days and therefore remains available for a long period even after drug withdrawal.

Amiodarone inhibits types 1 and 2 5’-deiodinases and increases $T_3$ levels at the same time as it increases TSH. The drug may also compete for the $T_3$ receptor. Amiodarone also has a direct cytotoxic effect on thyroid cells via induction of apoptosis. In addition, an iodine load may precipitate thyroid autoimmune disease, as discussed earlier, in susceptible individuals and induces most commonly autoimmune hyperthyroidism, but Graves’ disease may also be precipitated. Most commonly, however, it is the influence of the iodine load that causes the clinical effects, particularly in areas of iodine deficiency where thyrotoxicosis is most commonly seen and in areas of iodine deficiency where hypothyroidism is most commonly seen.

Amiodarone-induced hyperthyroidism may develop either rapidly in a patient or even after a few years of treatment. The different presentations have been divided as follows:

- **Type I** occurs in abnormal thyroid glands secondary to iodine excess.
- **Type II** is secondary to drug-induced destructive thyroiditis, sometimes resulting in high serum IL-6 levels that can be checked in the serum.

Both types may occur together, and treatment is often a major challenge. Antithyroid drugs and radiiodine may be ineffective because of the large intrathyroidal iodine load. Thyroidectomy is therefore the treatment of choice if the patient’s clinical condition permits. This allows the continuation of the drug. The use of combined potassium perchlorate (1 g daily) and methimazole (40 mg daily) may be tried if surgery is not possible. A complete blood count must be checked regularly with such treatment, which is fraught with side effects. If Type II predominates, steroids may control the destructive thyroiditis (2 to 6 mg of dexamethasone daily). In contrast, when hyperthyroidism is the presentation, due to thyroid destruction at the onset of Hashimoto’s disease or the failure to escape from iodine blockade, levothyroxine replacement is effective. TSH should be monitored appropriately. If the drug is discontinued, there may be spontaneous remission. Potassium perchlorate may provide increased speed of recovery by discharging intrathyroidal iodide.

**Hamburger Thyrotoxicosis**

An unusual form of exogenous thyrotoxicosis occurred in the midwestern portion of the United States in 1984 and 1985. The source was the inclusion of large quantities of bovine thyroid in ground beef preparations. When the slaughtering practices were changed, this condition, called hamburger thyrotoxicosis, disappeared. Such a possibility, however remote, should be considered, especially if one is confronted with epidemic exogenous thyrotoxicosis. Here the cause was probably the direct ingestion of large amounts of thyroid hormone, but the additional iodine load and its consequences should also be considered.
Thyrotoxicosis Caused by Nonthyroid Sources of Thyroid Hormone

Thyrotoxicosis Factitia

Thyrotoxicosis that arises from the (usually chronic) ingestion of excessive quantities of thyroid hormone usually occurs in individuals with a background of underlying psychiatric disease, especially in paramedical personnel who have access to thyroid hormone or in patients for whom thyroid hormone medication has been prescribed in the past. Generally, the patient is aware of taking thyroid hormone but may adamantly deny it. In other instances, large doses of thyroid hormone or other thyroactive material, such as iodocasein, may be taken without the knowledge of the patient, usually as part of a regimen for weight reduction. Symptoms are typical of thyrotoxicosis and may be severe.

In the absence of preexisting thyroid disease, the diagnosis is made from the combination of typical thyrotoxic manifestations, together with thyroid atrophy and hypofunction. Infiltrative ophthalmopathy never occurs, but lid lag, stare, and other "thyrotoxic" eye signs may be present. Hypofunction of the thyroid gland is evidenced by suppressed serum Tg levels and subnormal values of RAIU, which can be increased by administration of TSH. Serum \( T_4 \) concentrations are increased unless the patient is taking \( T_3 \), in which case they will be subnormal. Serum \( T_3 \) concentrations are increased in either case. TSH levels are suppressed. The presence of low, rather than elevated, values of serum Tg concentration is a clear indication that the thyrotoxicosis results from exogenous hormone rather than thyroid hyperfunction.

This disorder may be confused with other varieties of thyrotoxicosis associated with a subnormal RAIU and absence of goiter, including the syndrome of Hashimoto's disease preceded by transient thyrotoxicosis (silent thyroiditis), ectopic thyroid tissue, and hyperfunctioning meta-static follicular carcinoma. Evidence for the latter two disorders can be obtained by demonstration of the ectopic focus or foci by external radioiodine scanning and the fact that serum thyroglobulin is increased in both conditions.

Differentiation from silent and painless thyroiditis may be difficult. The presence of circulating TPO and Tg antibodies points to painless chronic autoimmune thyroiditis, whereas a firm thyroid gland and a brief history suggest the painless variant of subacute thyroiditis.

Treatment consists of withdrawing the offending medication. Psychiatric help may be necessary.

Ectopic Thyroid Tissue

Thyroid tissue may be present in teratomas, especially in the ovary (struma ovarii), and such foci may produce thyrotoxicosis. Rarely, hyperfunctioning metastases of follicular carcinoma can produce thyrotoxicosis. The distinguishing features of such lesions have been discussed earlier.
Special Aspects of Thyrotoxicosis

T₃ Toxicosis

Concurrent measurements of T₃ and T₄ production rates have revealed a disproportionate increase in T₃ production in most patients with hyperthyroidism. Whether this phenomenon results solely from the preferential increase in thyroid synthesis of T₃, preferential intrathyroid T₃ to T₄ conversion or from a disproportionate increase in peripheral conversion of T₃ to T₄ is uncertain, but the preferential synthesis is probably responsible in most instances. In the extreme case, the production rate of T₃ alone is increased, resulting in the thyrotoxic state designated T₃ toxicity. In some patients, T₃ toxicity may be the forerunner of the usual form of thyrotoxicosis in which production of both T₃ and T₄ is increased, and in other patients it persists as such. T₃ toxicity may occur with Graves’ disease, toxic multinodular goiter, or toxic adenoma. (See Table 10-14.)

The prevalence is not known, but T₃ toxicity appears to be more common than the conventional types of hyperthyroidism in areas of iodine deficiency. In our experience, it tends to be more frequent in older people.

The diagnosis should be suspected in a patient with clinical manifestations of thyrotoxicosis in whom the serum T₄ level and free T₄ concentration or index are normal or decreased while the serum TSH concentration is suppressed. Documentation of an elevated free T₃ level confirms the diagnosis. The presence of a palpable goiter and a normal or increased RAIU excludes the diagnosis of thyrotoxicosis factitia induced by ingestion of T₃. Experience suggests that patients with T₃ toxicity are more likely to have a long-term remission after withdrawal of antithyroid drug therapy than patients with the usual form of thyrotoxicosis.

T₄ Toxicosis

T₄ toxicity refers to thyrotoxicosis with an increased serum T₄ concentration and free T₄ concentration or index but a normal or decreased serum T₃ concentration. This phenomenon occurs in two circumstances. One is iodine-induced thyrotoxicosis, in which about one third of the patients have a normal serum T₄ concentration and the remainder display proportionate elevations of serum T₃ and T₄ concentrations. The second is thyrotoxicosis accompanied by severe intercurrent illness. Here, that component of the serum T₄ usually contributed by peripheral T₃ monodeiodination. With is decreased or lacking, so the serum T₃ concentration, sustained mainly or entirely by direct thyroid secretion, is normal or low, although the serum T₄ concentration is high.

Concomitantly, the serum rT₃ concentration is increased, often markedly, owing to inhibition of its 5’monodeiodination. With recovery from the intercurrent illness, serum rT₃ concentration declines and serum T₄ concentration increases into the thyrotoxic range. T₄ toxicity of this type is so be differentiated from the low serum T₄ level and elevated serum T₄ concentration that are occasionally found in the euthyroid sick syndrome. A reduced serum TSH level may not distinguish patients who are hyperthyroid from those who are not. Some of these changes are thought to be cytokine-mediated.

Thyrotoxic Crisis (Thyroid Storm)

Thyrotoxic crisis, also called accelerated hyperthyroidism or thyroid storm, is an extreme accentuation of thyrotoxicosis. The crisis, however, may be exaggerated by an inexperienced physician. It is an uncommon but serious complication, usually occurring in association with Graves’ disease but sometimes with toxic multinodular goiter. Before the availability of adequate means for achieving full preoperative control, crisis frequently followed subtotal thyroidectomy (surgical crisis).

Thyrotoxic crisis is usually of abrupt onset and occurs in patients in whom preexisting thyrotoxicosis has been treated incompletely or has not been treated at all. Crisis is usually precipitated by infection, trauma, surgical emergencies, or operations and, less commonly, by radiation thyroiditis, diabetic ketoacidosis, toxemia of pregnancy, or parturition. The mechanism by which such factors worsen thyrotoxicosis may be related to cytokine release and acute immunologic disturbance caused by the precipitating condition. The serum thyroid hormone levels in crisis are not appreciably greater than those in uncomplicated thyrotoxicosis.

The clinical picture is one of severe hypermetabolism. Fever is almost invariable and may be severe; sweating is profuse. Marked tachycardia of sinus or ectopic origin and arrhythmias may be accompanied by pulmonary edema or congestive heart failure. Tremulousness and restlessness are present; delirium or frank psychosis may supervene. Nausea, vomiting, and abdominal pain may occur early in the course. As the disorder progresses, apathy, stupor, and coma may supervene, and hypotension can develop. If unrecognized, the condition may be fatal. This clinical picture in a patient with a history of preexisting thyrotoxicosis or with goiter or exophthalmos or both is sufficient to establish the diagnosis, and emergency treatment should not wait laboratory confirmation.

There are no foolproof criteria by which severe thyrotoxicosis complicated by some other serious disease can be distinguished from thyrotoxic crisis induced by that disease. In any event, the differentiation between these alternatives is of no great significance because treatment of the two is the same.

Treatment aims to correct both the severe thyrotoxicosis and the precipitating illness and to provide general support. The patient thought to have thyroid crisis should be monitored in a medical intensive care unit during the initial phases of therapy. The therapy itself is designed to inhibit hormone synthesis and release and to antagonize the adrenergically mediated aspects of peripheral thyroid hormone action.

Large doses of an antithyroid agent (300 to 400 mg of propylthiouracil every 4 hours) are given by mouth, by stomach tube, or, if necessary, per rectum. Propylthiouracil may be preferable to methimazole because it has the additional action of inhibiting the peripheral generation of T₃ from T₄ by Type I iodothyronine deiodinase. Administration of propylthiouracil initiates therapy for the postcrisis period and prevents enrichment of glandular hormone stores by the iodine, which administration is of more immediate importance. The latter agent, administered either as SSKI (5 drops every 6 hours), ipodate (0.5 g twice daily) or, if available, sodium iodide intravenously (0.250 g every 6 hours), acutely retards the release of hormone from the thyroid gland.

Propylthiouracil should be administered before iodine to inhibit the synthesis of additional thyroid hormones from the administered iodine. Nonetheless, because iodine blocks release of preformed thyroid hormones from the thyroid gland, its administration

Thyrotoxic Crisis (Thyroid Storm)

Thyrotoxic crisis, also called accelerated hyperthyroidism or thyroid storm, is an extreme accentuation of thyrotoxicosis. The crisis, however, may be exaggerated by an inexperienced physician. It is an uncommon but serious complication, usually occurring in association with Graves’ disease but sometimes with toxic multinodular goiter. Before the availability of adequate means for achieving full preoperative control, crisis frequently followed subtotal thyroidectomy (surgical crisis).

Thyrotoxic crisis is usually of abrupt onset and occurs in patients in whom preexisting thyrotoxicosis has been treated incompletely or has not been treated at all. Crisis is usually precipitated by infection, trauma, surgical emergencies, or operations and, less commonly, by radiation thyroiditis, diabetic ketoacidosis, toxemia of pregnancy, or parturition. The mechanism by which such factors worsen thyrotoxicosis may be related to cytokine release and acute immunologic disturbance caused by the precipitating condition. The serum thyroid hormone levels in crisis are not appreciably greater than those in uncomplicated thyrotoxicosis.

The clinical picture is one of severe hypermetabolism. Fever is almost invariable and may be severe; sweating is profuse. Marked tachycardia of sinus or ectopic origin and arrhythmias may be accompanied by pulmonary edema or congestive heart failure. Tremulousness and restlessness are present; delirium or frank psychosis may supervene. Nausea, vomiting, and abdominal pain may occur early in the course. As the disorder progresses, apathy, stupor, and coma may supervene, and hypotension can develop. If unrecognized, the condition may be fatal. This clinical picture in a patient with a history of preexisting thyrotoxicosis or with goiter or exophthalmos or both is sufficient to establish the diagnosis, and emergency treatment should not wait laboratory confirmation.

There are no foolproof criteria by which severe thyrotoxicosis complicated by some other serious disease can be distinguished from thyrotoxic crisis induced by that disease. In any event, the differentiation between these alternatives is of no great significance because treatment of the two is the same.

Treatment aims to correct both the severe thyrotoxicosis and the precipitating illness and to provide general support. The patient thought to have thyroid crisis should be monitored in a medical intensive care unit during the initial phases of therapy. The therapy itself is designed to inhibit hormone synthesis and release and to antagonize the adrenergically mediated aspects of peripheral thyroid hormone action.

Large doses of an antithyroid agent (300 to 400 mg of propylthiouracil every 4 hours) are given by mouth, by stomach tube, or, if necessary, per rectum. Propylthiouracil may be preferable to methimazole because it has the additional action of inhibiting the peripheral generation of T₃ from T₄ by Type I iodothyronine deiodinase. Administration of propylthiouracil initiates therapy for the postcrisis period and prevents enrichment of glandular hormone stores by the iodine, which administration is of more immediate importance. The latter agent, administered either as SSKI (5 drops every 6 hours), ipodate (0.5 g twice daily) or, if available, sodium iodide intravenously (0.250 g every 6 hours), acutely retards the release of hormone from the thyroid gland.

Propylthiouracil should be administered before iodine to inhibit the synthesis of additional thyroid hormones from the administered iodine. Nonetheless, because iodine blocks release of preformed thyroid hormones from the thyroid gland, its administration
blockade of atrio-ven-tricular node conduction.

When treatment is successful, improvement is usually manifested within 1 or 2 days and recovery occurs within a week. At this time, iodide and dexamethasone are gradually withdrawn and plans for long-term management are made.
Subclinical Hyperthyroidism

With the advent of thyroid function screening and the presence of highly sensitive TSH assays, many patients with mild thyroid disease now consult physicians for advice. The presence of a chronically suppressed serum level TSH with peripheral free thyroid hormone levels within the normal range has been defined as mild or "subclinical" hyperthyroidism. In a study of more than 25,000 people attending a health fair, 0.9% were found to have this condition.

Subclinical hyperthyroidism may occur with clinical symptoms and signs that include weight loss, anxiety, atrial fibrillation, and osteoporosis or with no symptoms at all. All of the causes of thyrotoxicosis may be associated with such a mild presentation, most often Graves’ disease and toxic multinodular goiter. Indeed, it is these patients who are the ones with precipitated severe thyrotoxicosis after an increase in their iodine intake.

Making the correct diagnosis is helpful in deciding treatment and excluding transient thyroiditis. When there are symptoms or obvious risks, treatment is required. If there are no signs or symptoms and the diagnosis has been made biochemically, watchful observation may be appropriate. In patients with mild Graves’ disease, a short course of low-dose antithyroid drugs may cure the condition. In the presence of toxic nodules, radioiodine may be more appropriate, particularly in elderly populations, whose risk of cardiac arrhythmias is increased.
References

35. Silva JE. Thyroid hormone control of thermogenesis and energy balance. Thyroid 1996; 5:481492.


References


195. Volpe R. Evidence that the immunosuppressive effects of antithyroid drugs are mediated through actions on the thyroid cell, modulating thyrocyte-immunocyte signaling: a review. Thyroid 1994; 4:1217-223.


419


241. Tuttle RM, Patience T, Budd S. Treatment with propylthiouracil before radioactive iodine therapy is associated with a higher treatment failure rate than therapy with radioactive iodine alone in Graves' disease. Thyroid 1999; 5:243247.


Glinoer D. Thyroid hyperfunction during pregnancy. Thyroid 1998; 8:859864.


References


52. Beck-Peccoz P, Chatterjee V. The variable clinical phenotype in thyroid hormone resistance syndrome. Thyroid 1994; 4:225231.


344. Danese MD, Powe NR, Sawin CT, Ladenson PW. Screening for mild thyroid failure at the periodic health examination: a decision and cost-effectiveness analysis. JAMA 1996; 276:28592.

422
Chapter 12 - Hypothyroidism and Thyroiditis

P. Reed Larsen
Terry F. Davies

HYPOTHYROIDISM

Many structural or functional abnormalities can impair the production of thyroid hormones and cause the clinical state termed hypothyroidism. The causes can be divided into six main categories:

1. Hypothyroidism with compensatory thyroid enlargement due to transient or progressive impairment of hormone biosynthesis (goitrous hypothyroidism).
2. Permanent loss or atrophy of thyroid tissue (iatrogenic hypothyroidism).
3. Transient hypothyroidism.
4. Consumptive hypothyroidism.
5. Central hypothyroidism, that is, hypothyroidism due to insufficient stimulation of a normal gland as a result of hypothalamic or pituitary disease or defects in the thyroid-stimulating hormone (TSH) molecule itself.
6. Resistance to thyroid hormone (RTH).

Primary hypothyroidism accounts for approximately 99% of cases, with fewer than 1% being due to TSH deficiency. Central hypothyroidism is discussed in Chapter 7 and Chapter 8.

Clinically apparent acquired impairment of thyroid function affects about 2% of adult women and about 0.1 to 0.2% of adult men. Subclinical hypothyroidism, an elevated TSH level in an asymptomatic patient, was recently found in 8.9% of a self-selected U.S. population of approximately 20,000. Neonatal screening programs for congenital hypothyroidism discover hypothyroidism in almost 1 in 3500 newborns.

Clinical Presentation

Hypothyroidism can affect all organ systems, and these manifestations are largely independent of the underlying disorder but are a function of the degree of hormone deficiency. The following sections discuss the pathophysiology of each organ system at various levels of thyroid hormone deficiency, from mild to severe. The term myxedema, formerly used as a synonym for hypothyroidism, refers to the appearance of the skin and subcutaneous tissues in the patient in a severely hypothyroid state (Fig. 12-1). Hypothyroidism of this severity is rarely seen today, and the term should be reserved to describe the physical signs.

Skin and Appendages

Hypothyroidism causes an accumulation of hyaluronic acid that alters the composition of the ground substance in the dermis and other tissues. This material is hygroscopic, producing the mucinous edema that is responsible for the thickened features and puffy appearance (myxedema) with full-blown hypothyroidism. Myxedematous tissue is characteristically boggy and nonpitting and is apparent around the eyes, on the dorsa of the hands and feet, and in the supracavicular fossae (Fig. 12-1). It causes enlargement of the tongue and thickening of the pharyngeal and laryngeal mucous membranes.

A histologically similar deposit may occur in patients with Graves’ disease, usually over the pretibial area (infiltrative dermopathy or pretibial myxedema). In addition to having a puffy appearance, the skin is pale and cool as a result of cutaneous vasoconstriction. Anemia may contribute to the pallor; hypercarotenemia gives the skin a yellow tint but does not cause scleral icterus. It causes enlargement of the tongue and thickening of the pharyngeal and laryngeal mucous membranes.

Wounds of the skin tend to heal slowly. Easy bruising is due to an increase in capillary fragility. Head and body hair is dry and brittle, lacks luster, and tends to fall out. Hair may be lost from the temporal aspects of the eyebrows, which is not specific for hypothyroidism. The secretion of the sweat glands and sebaceous glands are reduced, leading to dryness and coarseness of the skin, which in extreme cases may resemble ichthyosis.

In secondary hypothyroidism, the degree of hypothyroidism is less severe and the changes in the skin and its appendages may be less striking than in primary hypothyroidism. The skin is pale and cool and tends to be thinner and finely wrinkled, and infiltration of the tissues is less prominent. Depigmentation of areas that are normally pigmented, such as the areolae, occurs in a central but not primary hypothyroidism.

Histopathologic examination of the skin reveals hyperkeratosis with plugging of hair follicles and sweat glands. The dermis is edematous, and the connective tissue fibers are separated by an increased amount of metachromatically staining, periodic acid-Schiff positive mucinous material. This material consists of protein complexed with two mucopolysaccharides: hyaluronic acid and chondroitin sulfate B. The glycosaminoglycans are mobilized early during treatment with thyroid hormone, leading to an increase in urinary excretion of nitrogen and hexosamine.

Cardiovascular System

The cardiac output at rest is decreased because of reduction in both stroke volume and heart rate, reflecting loss of the inotropic and chronotropic effects of thyroid hormones. Peripheral vascular resistance at rest is increased, and blood volume is reduced. These hemodynamic alterations cause narrowing of pulse pressure, prolongation of circulation time, and decrease in blood flow to the tissues. The decrease in cutaneous...
circulation is responsible for the coolness and pallor of the skin and the sensitivity to cold. In most tissues, the decrease in blood flow is proportional to the decrease in oxygen consumption, so the arteriovenous oxygen difference remains normal. The hemodynamic alterations at rest resemble those of congestive heart failure. However, in hypothyroidism, cardiac output increases and peripheral vascular resistance decreases normally in response to exercise unless the hypothyroid state is severe and of long standing.

In severe primary hypothyroidism the cardiac silhouette is enlarged (Fig. 12-2), and the heart sounds are diminished in intensity. These findings are the result largely of effusion into the pericardial sac of fluid rich in protein and glycosaminoglycans, but the "flabby" myocardium may also be dilated. Pericardial effusion is rarely of sufficient magnitude to cause tamponade. In central hypothyroidism, the heart is typically small.

Angina pectoris is uncommon, but it may appear or worsen during treatment of the hypothyroid state with thyroid hormone. There is considerable controversy over whether hypothyroidism is a risk factor for atherosclerosis. The Whickham Study showed no increase in cardiovascular mortality in patients with subclinical hypothyroidism over 20 years, whereas the Rotterdam Study suggested that there was a twofold increase in risk. Systemic vascular resistance is increased, and hypertension is more common.

Electrocardiographic changes include sinus bradycardia, prolongation of the PR interval, low amplitude of the P wave and QRS complex, alterations of the ST segment, and flattened or inverted T waves. Pericardial effusion is probably responsible for the low amplitude in severe hypothyroidism. Rarely, complete heart block may be present, but this disappears when hypothyroidism is treated. Systolic time intervals are altered; the pre-ejection period is prolonged, and the ratio of pre-ejection period to left ventricular ejection time is increased. Echocardiographic studies have revealed a high frequency of asymmetrical septal hypertrophy and apparent obstruction of the left ventricular outflow tract, suggesting idiopathic hypertrophic subaortic stenosis. These findings disappear when the hypothyroidism is treated, and their hemodynamic significance is uncertain.

Serum levels of homocysteine, creatine kinase, aspartate aminotransferase, and lactate dehydrogenase may be increased. Typically, the isoenzyme patterns suggest that the source of the increased creatine kinase and lactate dehydrogenase is skeletal, not cardiac, muscle. All levels return to normal with therapy.

The combination of large heart, hemodynamic and electrocardiographic alterations, and the serum enzyme changes has been termed myxedema heart. Myxedema heart rarely causes heart failure by itself because the usual hemodynamic response to exercise in hypothyroidism is typically normal, although exceptions have been reported. In the patient with hypothyroidism, as in the normal individual, theValsalva maneuver leads to a decrease in pulse pressure, whereas in the patient with heart failure the pulse pressure does not decrease but displays the so-called square-wave response. In the absence of coexisting organic heart disease, treatment with thyroid hormone corrects the hemodynamic, electrocardiographic, and serum enzyme alterations of myxedema heart and restores heart size to normal.

Figure 12-2

A

B

Alimentary System

Although most patients experience a modest gain in weight, appetite is usually reduced. The weight gain that occurs is caused partly by retention of fluid by the hydrophilic glycoprotein deposits in the tissues. Peristaltic activity is decreased and, together with the decreased food intake, is responsible for the frequent complaint of constipation. The latter may lead to fecal impaction (myxedema megacolon). Gaseous distention of the abdomen (myxedema ileus), if accompanied by colicky pain and vomiting, may mimic mechanical ileus.

Elevations in the serum levels of carcinoembryonic antigen, which may occur on the basis of hypothyroidism alone, add to the impression that an organic obstruction is present. Acutees in the absence of another cause is unusual in hypothyroidism, but it can occur, usually in association with pleural and pericardial effusions. Like pericardial and pleural effusions, the ascitic fluid is rich in protein and glycosaminoglycans.

Achlorhydria after maximal histamine stimulation may be present in patients with primary hypothyroidism. Circulating antibodies against gastric parietal cells have been found in about one third of patients with primary hypothyroidism and may be secondary to atrophy of the gastric mucosa. Overt pernicious anemia is reported in about 12% of patients with primary hypothyroidism. The coexistence of pernicious anemia and other autoimmune diseases, such as gluten enteropathy, with primary hypothyroidism reflects the fact that autoimmunity plays the central role in the pathogenesis of these diseases (see Chapter 37).

Hypothyroidism has complex effects on intestinal absorption. Although the rates of absorption for many substances are decreased, the total amount absorbed may be normal or even increased because the decreased bowel motility may allow more time for absorption. Malabsorption is occasionally overt.

Liver function test results are usually normal, but levels of aminotransferases may be elevated, probably because of impaired clearance. The gallbladder contracts sluggishly and may be distended, but whether these changes predispose to the development of gallstones is unknown.

Atrophy of the gastric and intestinal mucosa and myxedematous infiltration of the bowel wall may be demonstrated on histologic examination. The colon may be greatly distended, and the volume of fluid in the rectal cavity is unusually increased. The liver and pancreas are normal.

Central and Peripheral Nervous System

Thyroid hormone is essential for the development of the central nervous system. Deficiency in fetal life or at birth causes retention of the infantile characteristics of the brain, hypoplasia of cortical neurons with poor development of cellular processes, retarded myelination, and reduced vasculature. If the deficiency is not corrected in early postnatal life, the damage is irreversible. Deficiency of thyroid hormone beginning in adult life causes less severe manifestations that usually respond to treatment with the hormone. Cerebral blood flow is reduced, but cerebral oxygen consumption is usually normal; this finding is in accord with the conclusion that the oxygen consumption of isolated brain tissue in vitro, unlike that of most other tissues, is not stimulated by administration of thyroid hormones. In severe cases, decreased cerebral blood flow may lead to cerebral hypoxia.

All intellectual functions, including speech, are slowed. Loss of initiative is present, slow-wittedness and memory defects are common, lethargy and somnolence are prominent, and dementia in elderly patients may be mistaken for senile dementia. Psychiatric disorders are common and are usually of the paranoid or depressive type and may induce agitation (myxedema madness). Headaches are frequent. Cerebral hypoxia due to circulatory alterations may predispose to confusional attacks and syncope, which may be prolonged and lead to stupor or coma. Other factors predisposing to coma in hypothyroidism include exposure to severe cold, infection, trauma, hyperventilation with carbon dioxide retention, and depressant drugs.
of the pigment required for dark adaptation. Hearing loss of the perceptive type is frequent due to myxedema of the eighth cranial nerve and serous otitis media. Perceptive deafness may also occur in association with a defect in the organic binding of thyroidal iodide (Pendred's syndrome) (see discussion of iodine metabolism), but in these instances it is not due to hypothyroidism per se.

Thick, slurred speech and hoarseness are due to myxedematosus infiltration of the tongue and larynx, respectively. Body movements are slow and clumsy, and cerebellar ataxia may occur. Numbness and tingling of the extremities are frequent; in the fingers these symptoms may be due to compression by glycosaminoglycan deposits in and around the median nerve in the carpal tunnel (carpal tunnel syndrome). The tendon reflexes are slow, especially during the relaxation phase, producing the characteristic "hang-up reflexes"; this phenomenon is due to a decrease in the rate of muscle contraction and relaxation rather than a delay in nerve conduction.

The presence of extensor planter responses or diminished vibration sense should alert the physician to the possibility of coexisting pernicious anemia with combined system disease. Electroneurographic changes include slow alpha wave activity and general loss of amplitude. The concentration of protein in the cerebrospinal fluid is often increased, but cerebrospinal pressure is normal.

Histopathologic examination of the brain in patients with untreated hypothyroidism reveals that the nervous system is edematous with mucinous deposits in and around nerve fibers. In patients with cerebellar ataxia, neural myxedematous infiltrates of glycosgen and mucinuous material are present in the cerebellum. There may be foci of degeneration and an increase in glial tissue. The cerebral vessels show atherosclerosis, but this is much more common if the patient has had coexistent hypertension.

Muscular System

Stiffness and aching of muscles are common and are worsened by cold temperatures. Delayed muscle contraction and relaxation cause the slowness of movement and delayed tendon jerks. Muscle mass may be reduced or enlarged due to interstitial myxedema. Muscle mass may be slightly increased, and the muscles tend to be firm. Rarely, a profound increase in muscle mass with slowness of muscular activity may be the predominant manifestation (the Kocher-Debri-Semelaigne, or Hoffmann) syndrome. Myoclonus may be present. The electromyogram may be normal or may exhibit disordered discharge, hyperirritability, and polyphasic action potentials.

On histopathologic examination, the muscles appear pale and swollen. The muscle fibers may show swelling, loss of normal striations, and separation by mucinous deposits. Type I muscle fibers tend to predominate.

Skeletal System: Calcium and Phosphorus Metabolism

Thyroid hormone is essential for normal growth and maturation of the skeleton, and growth failure is due both to impaired general protein synthesis and to a reduction in growth hormone, but especially of insulin-like growth factor I. Before puberty, thyroid hormone plays a major role in the maturation of bone. Deficiency of thyroid hormone in early life leads to both a delay in the development of growth plates and of an abnormal, stippled appearance of the epiphyseal centers of ossification (epiphyseal dysgenesis). Impairment of linear growth leads to dwarfism in which the limbs are disproportionately short in relation to the trunk but cartilage growth is unaffected. Urinary excretion of calcium is decreased, as is the glomerular filtration rate, whereas fecal excretion of phosphorus is variable. Calcium balance is also variable, and any changes are slight. The exchangeable pool of calcium and its rate of turnover are reduced, changes that reflect decreased bone formation and resorption. Because levels of parathyroid hormone are often slightly increased, some degree of resistance to its action may be present; levels of 1,25(OH)2D are also increased.

Levels of calcium and phosphorus in serum are usually normal, but calcium may be slightly elevated. The alkaline phosphatase level is usually below normal in infantile and juvenile hypothyroidism. Bone density may be increased. The radiologic appearance of the skeleton in cretinism and juvenile hypothyroidism are discussed subsequently.

Renal Function: Water and Electrolyte Metabolism

Renal blood flow, glomerular filtration rate, and tubular reabsorptive and secretory maxima are reduced. Blood urea nitrogen and serum creatinine levels are normal, but uric acid levels may be increased. Urine flow is reduced, and delay in the excretion of a water load may result in reversal of the normal diurnal pattern of urine excretion. The delay in water excretion appears to be due to decreased volume delivery to the distal diluting segment of the nephron as a result of the diminished renal perfusion; evidence supporting inappropriate secretion of vasopressin (syndrome of inappropriate antidiuretic hormone) is less compelling. These changes are reversed by treatment with thyroid hormone. The ability to concentrate urine may be slightly impaired. Mild proteinuria may occur.

The impaired renal excretion of water in the retention of water by the hydrophilic deposits in the tissues result in an increase in total body water, even though plasma volume is reduced. This increase accounts for the hyponatremia occasionally noted because the level of exchangeable sodium is increased. The amount of exchangeable potassium is usually normal in relation to lean body mass. Serum magnesium concentration may be increased, but exchangeable magnesium levels and urinary magnesium excretion are decreased.
In response to the diminished oxygen requirements and decreased production of erythropoietin, the red blood cell mass is decreased; this is evident in the mild normocytic, normochromic anemia that often occurs. Less commonly, the anemia is macrocytic, sometimes from deficiency of vitamin B₁₂. Reference has already been made to the high incidence of pernicious anemia (and of achlorhydria and vitamin B₁₂ deficiency without overt anemia) in primary hypothyroidism. (see Chapter 37). Conversely, overt and subclinical hypothyroidism is present in 12% and 15% of patients, respectively, with pernicious anemia. Folate deficiency by malabsorption or dietary inadequacy may also cause macrocytic anemia. The frequent menorrhagia and the defective absorption of iron resulting from achlorhydria may contribute to a microcytic, hypochromic anemia.

The total and differential white blood cell counts are usually normal, and platelets are adequate, although platelet adhesiveness may be impaired. If pernicious anemia or significant folate deficiency is present, the characteristic changes in peripheral blood and bone marrow will be found. The intrinsic clotting mechanism may be defective because of decreased concentrations in plasma of factors VIII and IX, and this, together with an increase in capillary fragility and the decrease in platelet adhesiveness, may account for the bleeding tendency that sometimes occurs.

### Pituitary and Adrenocortical Function

In long-standing hypothyroidism of thyroid origin, hyperplasia of the thyrotropes may cause the pituitary gland to be enlarged. This feature can be detected radiologically as an increase in the volume of the pituitary fossa. Rarely, the pituitary enlargement compromises the function of other pituitary cells and causes pituitary insufficiency or visual field defects. Patients with severe hypothyroidism may have increased serum prolactin levels that correlate with the level of serum TSH, and galactorrhea may develop in some patients. Treatment with thyroid hormone corrects serum prolactin and TSH levels and causes disappearance of galactorrhea, if present. The cause of hyperprolactinemia in hypothyroidism is uncertain but may result from enhanced sensitivity of the lactotropes to thyrotropin-releasing hormone (TRH). In severe primary hypothyroidism, the response of growth hormone to provocative stimuli, such as insulin-induced hypoglycemia or growth hormone-releasing hormone may be subnormal.

As a result of the decreased rate of turnover of cortisol due to decreased hepatic 11-hydroxysteroid dehydrogenase, type 1 (11-HSD-1), the 24-hour urinary excretion of cortisol and 17-hydroxycorticosteroids is decreased but the plasma cortisol level is usually normal (see Chapter 14). The responses of urinary 17-OH-corticosteroid to exogenous adrenocorticotropic hormone and metyrapone are usually normal but may be decreased. The response of plasma cortisol to insulin-induced hypoglycemia may be impaired.

In severe, long-standing primary hypothyroidism, pituitary and adrenal function may be secondarily decreased and adrenal insufficiency may be precipitated by stress or by rapid replacement therapy with thyroid hormone. The rate of turnover of aldosterone is decreased, but the plasma level is normal. Plasma renin activity is decreased, and sensitivity to angiotensin II is increased (see Chapter 15).

### Reproductive Function

In both sexes, thyroid hormones influence sexual development and reproductive function. Infantile hypothyroidism, if untreated, leads to sexual immaturity, and juvenile hypothyroidism causes a delay in the onset of puberty followed by anovulatory cycles. Paradoxically, primary hypothyroidism may also cause precocious sexual development and galactorrhea.

In adult women, severe hypothyroidism may be associated with diminished libido and failure of ovulation. Secretion of progesterone is inadequate, and endometrial proliferation persists, resulting in excessive and irregular breakthrough menstrual bleeding. These changes may be due to deficient secretion of luteinizing hormone. Rarely, in primary hypothyroidism, secondary depression of pituitary function may lead to ovarian atrophy and amenorrhea. Fertility is reduced, and spontaneous abortion may result, although many pregnancies are successful. Hypothyroidism in men may cause diminished libido, impotence, and oligospermia.

Values for plasma gonadotropins are usually in the normal range in primary hypothyroidism; in postmenopausal women, levels are usually somewhat lower than in euthyroid women of the same age but are nevertheless within the menopausal range. This provides a valuable means of differentiating primary from secondary hypothyroidism.

The metabolism of both androgens and estrogens is altered in hypothyroidism. Secretion of androgens is decreased, and the metabolism of testosterone is shifted toward estradiol and estrone rather than androsterone. With respect to estradiol and estrone, hypothyroidism favors metabolism of these steroids via 16-hydroxylation over that via 2-oxoxygenation, with the result that formation of estriol is increased and that of 2-hydroxyestrone and its derivative, 2-methoxyestrone, is decreased. The sex hormone-binding globulin in plasma is decreased, with the result that the plasma concentrations of both testosterone and estradiol are decreased, but the unbound fractions are increased. The alterations in steroid metabolism are corrected by restoration of the euthyroid state.

### Catecholamines

The plasma cyclic adenosine monophosphate (cAMP) response to epinephrine is decreased, suggesting a state of decreased adrenergic activity. The fact that the responses of plasma cAMP to glucose and parathyroid hormone are also decreased suggests that thyroid hormones have a general modulating influence on cAMP generation. The mechanism underlying the decreased adrenergic responsiveness is uncertain but probably results from impaired cAMP responses to norepinephrine.

### Energy Metabolism: Protein, Carbohydrate, and Lipid Metabolism

The decrease in energy metabolism and heat production is reflected in the low basal metabolic rate, decreased appetite, cold intolerance, and slightly low basal body temperature. Both the synthesis and the degradation of protein are decreased, the latter especially so, the net effect being one of lipid accumulation, especially of low-density lipoprotein (LDL) and triglycerides. The decrease in the lipid degradation rate may reflect the decrease in post-heparin lipolytic activity, as well as reduced LDL receptors. High-density lipoprotein (HDL) concentrations are reduced. The increase in serum cholesterol in primary (but not central) hypothyroidism is accompanied by increased levels of serum phospholipids, serum triglycerides, and LDL. Plasma free fatty acid levels are decreased, and the mobilization of free fatty acids in response to fasting, catecholamines, and growth hormone is impaired. All of these abnormalities are relieved by treatment.
In a random group of 1149 Dutch women, 11% were found to have subclinical hypothyroidism (TSH > 4.0 µU/L with normal free thyroxine. There was an increased prevalence of aortic atherosclerosis (odds ratio, 1.7; confidence interval, 1.1 to 2.6) and myocardial infarction (odds ratio, 2.3; confidence interval, 1.3 to 4.0). Also, hyperhomocysteinemia, a risk factor for vascular disease, is present in patients with primary hypothyroidism and responds rapidly to levothyroxine replacement.
Current Clinical Picture

In the adult, the onset of hypothyroidism is usually so insidious that the typical manifestations may take months or years to appear and go unnoticed by family and friends. The gradual development of the hypothyroid state is due to slow progression both of thyroid hypofunction and of the clinical manifestations after thyroid failure is complete. This course is in contrast with the more rapid development of the hypothyroid state when replacement therapy is discontinued in a patient with treated primary hypothyroidism or when the thyroid gland of a normal subject is surgically removed. In such patients, manifestations of frank hypothyroidism are present by 6 weeks; by 3 months myxedema appears.

The clinical picture of hypothyroidism in the third millennium is, in general, much milder than that described 50 years ago in the first edition of this textbook. New scales for assessment of clinical symptoms suggesting hypothyroidism have been developed (Fig. 12-5). In general, many of the symptoms are similar but much less prevalent than they were and do not significantly discriminate the hypothyroid from the euthyroid patient (e.g., cold intolerance or pulse rate). The reason for this is largely the ready availability of sensitive and specific laboratory tests for hypothyroidism that allow recognition of the primary form of the disease long before severe symptoms have developed. Nonetheless, early symptoms are variable and nonspecific. There should be a low threshold for screening patients for primary hypothyroidism with a TSH determination because in every series there are a number of patients with significant biochemical abnormalities who do not score high on symptom and sign testing.

With respect to physical signs of hypothyroidism, the presence of coarse skin, peri-orbital puffiness that obscures the curve of the malar bone (see Fig. 12-1), cold skin, and delayed ankle reflex relaxation phase all are signs that should lead to appropriate diagnostic tests.

The unusual syndrome of acute hypothyroidism in the previously hyperthyroid patient that is characterized by painful cramping of large muscle groups is discussed under the topic ‘Treatment of Graves’ Disease’ (see Chapter 11).

Severe hypothyroidism is seldom apparent at birth, hence the requirement for systematic screening. The age at which symptoms appear depends on the degree of impairment of thyroid function (see Fig. 12-3 and Fig. 12-4). Severe hypothyroidism in infancy is termed cretinism. As the age at onset increases, the clinical picture of cretinism merges imperceptibly with that of juvenile hypothyroidism. Retardation of mental development and growth, the hallmark of cretinism, becomes manifest only in later infancy and is largely irreversible. Consequently, early recognition is crucial and can be achieved by population screening by measuring serum T4 or TSH concentrations routinely in neonates. During the first few months of life, symptoms of hypothyroidism include feeding problems, failure to thrive, constipation, a hoarse cry, and somnolence. In succeeding months, especially in severe cases, protuberance of the abdomen, dry skin, poor growth of hair and nails, and delayed eruption of the deciduous teeth become evident. Retardation of mental and physical development is manifested by delay in reaching the normal milestones of development, such as holding up the head, sitting, walking, and talking.

Impairment of linear growth results in dwarfism, with the limbs disproportionately short in relation to the trunk (see Fig. 12-3). Delayed closure of the fontanelles causes the head to be large in relation to the body. The naso-orbital configuration remains infantile. Maldevelopment of the femoral epiphyses results in a waddling gait. The teeth are malformed and susceptible to caries. The characteristic appearance includes a broad, flat nose, widely set eyes; peri-orbital puffiness; large protruding tongue; sparse hair; rough skin; short neck; and protuberant abdomen with an umbilical hemia. Mental deficiency is usually severe.

Radiologic examination of the skeleton is diagnostic. The skull shows a poorly developed base; delayed closure of the fontanelles; widely set orbits; and a short, flat nasal bone. The pituitary fossa may be enlarged. Shedding of deciduous teeth and eruption of permanent teeth are delayed (see Fig. 12-4).

The radiologic picture of epiphyseal dysgenesis is virtually pathognomonic of hypothyroidism in infancy and childhood and may involve any center of endochondral ossification, depending on the age at onset of the hypothyroid state; it is usually best seen in the femoral and humeral heads and the navicular bone of the foot. The centers of ossification appear late, so bone age is retarded in relation to chronologic age, and when they eventually appear, instead of a single center, multiple small centers are scattered throughout the length and shape of the bone (see Fig. 12-4). These small centers of ossification eventually coalesce and form a single center with an irregular outline and a stippled appearance (stippled epiphysis). Epiphyseal dysgenesis is evident only in centers that normally ossify at a time after the onset of the hypothyroidism. After a normal metabolic state is restored by treatment, centers destined to ossify at a later age develop normally.

Hypothyroidism that begins in childhood is termed juvenile hypothyroidism. The clinical manifestations are intermediate, between those of infantile and those of adult hypothyroidism, in that the developmental retardation is not as severe as that of cretinism and the manifestations of full-blown adult myxedema are rarely seen. Growth and sexual development are affected predominantly. Linear growth is severely retarded, and the rate of linear growth is usually less than that of weight gain. Sexual maturation and the onset of puberty are delayed. The result is a child who appears much younger than the chronologic age. Rarely, precocious puberty and galactorrhea occur. Intellectual performance is poor, but the mental deficiency is not as severe as that in cretinism.

The manifestations of adult hypothyroidism are present to a varying, but usually milder, degree. On radiologic examination, epiphyseal dysgenesis may be present and epiphyseal union is always delayed, resulting in a bone age that is retarded in relation to chronologic age.
Laboratory Evaluation

Primary and Central Hypothyroidism

A decrease in secretion of the thyroid hormones is common to all varieties of hypothyroidism, except for consumptive hypothyroidism and resistance to thyroid hormone (see later). In patients with primary thyroid disease, the cause of hypothyroidism in more than 99% of the patients, there is a significant increase in basal serum TSH concentration. A strategy for evaluating the patient suspected of hypothyroidism involves an initial TSH determination. If the suspicion of hypothyroidism is strong, a goiter is present or central hypothyroidism is part of the differential diagnosis, a free T3 or free T4 index (FT4I) should be included (see Chapter 10). If hypothyroidism is thought to be unlikely but must be excluded, only a TSH determination is required because primary hypothyroidism is almost always the cause. If TSH is elevated, an FT3 can be added to the same determination (see Chapter 10). As hypothyroidism becomes more severe, the serum TSH increases and serum FT4 and later serum triiodothyronine (T3) concentrations become subnormal, the former more rapidly than the latter (see Table 12-1). The persistence of a normal serum T3 is, in part, due to preferential synthesis and secretion of T3 by residual functioning thyroid tissue under the influence of the increased plasma TSH. In addition, the efficiency of conversion of T4 to T3 by D2 is increased as the serum T4 level falls. Consequently, the serum T3 concentration may remain within the normal range and is not a useful index of thyroid function in the hypothyroid patient.

The differentiation of hypothyroidism due to intrinsic thyroid failure from hypothyroidism due to diminished TSH secretion

<table>
<thead>
<tr>
<th>TSH, Free T4 Index</th>
<th>TPO Antibodies</th>
<th>Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>TSH &gt; 10 mU/L</td>
<td>+</td>
<td>Primary hypothyroidism due to autoimmune thyroid disease</td>
</tr>
<tr>
<td>Low-normal</td>
<td>+</td>
<td>Primary “subclinical” hypothyroidism (autoimmune)</td>
</tr>
<tr>
<td>Low or low-normal</td>
<td>-</td>
<td>Recovery from systemic illness</td>
</tr>
<tr>
<td></td>
<td></td>
<td>External irradiation, drug-induced, congenital hypothyroidism</td>
</tr>
<tr>
<td>Normal</td>
<td>+/-</td>
<td>Consider TSH or T4 assay artifacts</td>
</tr>
<tr>
<td>Elevated</td>
<td>-</td>
<td>Thyroid hormone resistance</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Blockade of T4-to-T3 conversion (amiodarone) or a congenital 5'-deiodinase deficiency</td>
</tr>
<tr>
<td>TSH 500 mU/L</td>
<td>Low, low-normal</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Low, low-normal</td>
<td>-</td>
</tr>
<tr>
<td>Elevated</td>
<td>+/-</td>
<td>Consider thyroid hormone resistance</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T4-to-T3 conversion blockade (e.g., amiodarone)</td>
</tr>
<tr>
<td>TSH 0.55 mU/L</td>
<td>Low, low-normal</td>
<td>+/-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Salicylate or phenytoin therapy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Desiccated thyroid or T3 replacement</td>
</tr>
<tr>
<td>TSH &lt;0.5 mU/L</td>
<td>Low, low-normal</td>
<td>+/-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T3 or desiccated thyroid excess</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Postexcess levothyroxine withdrawal</td>
</tr>
</tbody>
</table>

Initial tests: Serum TSH, serum free T4 index, antibodies to TPO. A parenthesis indicates that the result is less common but may occur.

Results are grouped with respect to the serum TSH concentration. The different diagnoses possible for low, normal, or elevated free T4 index and TPO antibody results are indicated.

from hypothyroidic or pituitary disease (central or secondary hypothyroidism) is the most critical decision point in this pathway (see Fig. 12-6). A low thyroid hormone level with a normal or low TSH level should lead to an evaluation for the possibility of failure of other endocrine systems that require trophic pituitary hormones for normal function (Table 12-2) (see Chapter 7 and Chapter 8). The only exception to this is posthypothyroidal hypothyroidism, in which TSH levels may remain suppressed for several months even though hypothyroidism has been induced by 131I surgery, or antithyroid drugs (see Table 12-1). In some patients with central hypothyroidism, the basal serum TSH concentration (and the response to TRH) may even be somewhat elevated, but the TSH has reduced biologic potency even though it is immunologically reactive. In patients with an elevated TSH level and a reduced FT4I, the presence or absence of thyroid peroxidase (TPO) antibodies should be ascertained (see Fig. 12-6). The presence of TPO antibodies generally points to autoimmune thyroid disease (Hashimoto’s disease) as the cause of the hypothyroidism. On the other hand, the absence of TPO antibodies requires a search for less common causes of hypothyroidism such as transient hypothyroidism, infiltrative thyroid disorders, and external irradiation, as discussed later (see Table 12-1).

Tests that employ radiiodine to assess the function of the thyroid gland display a variable pattern, depending on the underlying thyroid disorder. When the amount of
thyroid tissue is reduced, the radioactive iodine uptake (RAIU) is subnormal. However, the diagnostic value of this finding in North America is minimized by the low normal range resulting from the high dietary iodine intake. Yet when hypothyroidism results primarily from a biochemical defect in thyroid hormone synthesis rather than thyroid cell destruction (thus leading to compensatory goitrogenesis), RAIU may be normal or increased. Specific functional patterns in relation to the causes of hypothyroidism are discussed later. Nonetheless, measurement of RAIU is almost never required in the diagnostic evaluation of the hypothyroid patient.

Differential Diagnosis

The clinical picture of fully developed hypothyroidism is characteristic enough to leave the diagnosis in little doubt. Despite the availability of inexpensive and specific tests, it is still surprising how often what is retrospectively obvious, severe, primary hypothyroidism is overlooked by experienced clinicians. If the diagnosis is not considered during the first meeting with the patient, it may take another 6 months before it is recognized that multiple, seemingly disparate complaints are due to hypothyroidism. A high index of suspicion is required to avoid this oversight.

For the milder forms of hypothyroidism, it may be necessary to differentiate them from several other states. The fact that these disorders often occur in older patients is partly responsible for the diagnostic uncertainty. In some cases, slowing of mental and physical activity, dry skin, and loss of hair may mimic similar findings in hypothyroidism. Furthermore, older people often become hypothermic with cold exposure. In patients with chronic renal insufficiency, anorexia, torpor, periorbital puffiness, sallow complexion, and anemia (e.g., see Fig. 12-1) may suggest hypothyroidism and may call for specific testing. Distinguishing nephrotic states from hypothyroidism by clinical examination alone may be even more difficult. In this disorder, waxy pallor, edema, hypercholesterolemia, and

---

TABLE 12-2 -- Causes of Hypothyroidism

<table>
<thead>
<tr>
<th>Primary Hypothyroidism with Goiter</th>
<th>Acquired</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hashimoto's thyroiditis (autoimmune thyroiditis type 2A)</td>
<td>Iodine deficiency (endemic goiter)</td>
</tr>
<tr>
<td>Drugs blocking synthesis or release of T₄ (e.g., lithium, ethionamide, sulfonamides, iodide)</td>
<td>Goitrogens in foodstuffs or as endemic substances or pollutants</td>
</tr>
<tr>
<td>Cytokines (interferon , interleukin-2)</td>
<td>Thyroid infiltration (amyloidosis, hemochromatosis, sarcoidosis, Riedel's struma, cystinosis, scleroderma)</td>
</tr>
<tr>
<td>Congenital</td>
<td>Iodide transport or utilization defect (NIS or pendrin mutations)</td>
</tr>
<tr>
<td>Pseudohypoparathyroidism</td>
<td>Iodotyrosine dehalogenase deficiency</td>
</tr>
<tr>
<td>Organification disorders (TPO deficiency or dysfunction)</td>
<td>Defects in thyroglobulin synthesis or processing</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Atrophic Hypothyroidism</th>
<th>Acquired</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hashimoto's disease (autoimmune thyroiditis type 2B)</td>
<td>Postablative due to ¹³¹I, surgery, or therapeutic irradiation for nonthyroidal malignancy</td>
</tr>
<tr>
<td>Congenital</td>
<td>Thyroid agenesis or dysplasia</td>
</tr>
<tr>
<td>TSH receptor defects</td>
<td>Thyroidal Gₛ protein abnormalities (pseudohypoparathyroidism type 1a)</td>
</tr>
<tr>
<td>Idiopathic TSH unresponsiveness</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Transient (Post-thyroiditis) Hypothyroidism</th>
<th>Following subacute, painless, or postpartum thyroiditis</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Consumptive Hypothyroidism</th>
<th>Rapid destruction of thyroid hormone due to D₃ expression in large hemangiomata or hemangioendotheliomata</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Central Hypothyroidism</th>
<th>Acquired</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pituitary origin (secondary)</td>
<td>Hypothalamic disorders (tertiary)</td>
</tr>
<tr>
<td>Bezaleotone (RXR receptor agonist)</td>
<td>Dopamine and/or severe illness</td>
</tr>
<tr>
<td>Congenital</td>
<td>TSH deficiency or structural abnormality</td>
</tr>
</tbody>
</table>

Figure 12-6 -- Strategy for the laboratory evaluation of patients with suspected hypothyroidism. The principal differential diagnosis is between primary and central hypothyroidism. The serum thyrotropin (TSH) concentration is the critical laboratory determination that, in general, allows recognition of the cause of the disease. An exception is the individual with a recent history of thyrotoxicosis (and suppressed TSH) in whom a low free thyroxine (T₄) level may be associated with a reduced TSH level for several months after relief of the thyrotoxicosis. In patients with primary hypothyroidism, the absence of thyroid peroxidase (TPO) antibodies raises a possible diagnosis of transient hypothyroidism following an undiagnosed episode of subacute or postviral thyroiditis. In such patients, a trial of a reduced levothyroxine dosage after 4 months may reveal recovery of thyroid function thus avoiding permanent levothyroxine replacement. MRI, magnetic resonance imaging.
TSH receptor defect

**Resistance to Thyroid Hormone**

Generalized

*"Pituitary" dominant*

TPO, thyroid peroxidase; TSH, thyroid-stimulating hormone.

Hypometabolism may suggest hypothyroidism. In addition, the total serum T₄ concentration may be decreased if significant thyroglobulin is lost in the urine but the FT₄ and TSH would be normal.

In patients with pernicious anemia, psychiatric abnormalities, pallor, and numbness and tingling of the extremities may mimic similar findings in hypothyroidism. Although there is a clinical and immunologic overlap between primary hypothyroidism and pernicious anemia, this association is not invariable (see Chapter 37). The presence of hypothyroidism is often suspected in patients who are severely ill, especially in the elderly. In such patients, the total T₄ concentration may be decreased, often markedly so, but the FT₄ is generally normal unless the patient is severely ill (see Chapter 10). These features, together with the absence of an elevation of serum TSH, usually serve to differentiate the ill euthyroid patient from one with primary hypothyroidism.

Hypothyroidism may develop either because of some extrinsic factor or acquired condition or because of a congenital defect impairing thyroid hormone biosynthesis (see Table 12-2). Inadequate synthesis of hormone leads to hypersecretion of TSH, which in turn produces both goiter and stimulation of all steps in hormone biosynthesis capable of response. If the compensatory response is inadequate, goitrous hypothyroidism results. In some instances, however, the compensatory response overcomes the impairment in hormone biosynthesis, and the patient is euthyroid with a compensatory goiter. The latter condition is discussed in Chapter 13 under the topic "Simple or Nontoxic Goiter." Less commonly, hypothyroidism is associated with an atrophic gland or, in the case of a congenital abnormality, one that never developed properly. From a clinical standpoint, it is useful to classify patients with hypothyroidism into those with and those without goiter (see Table 12-2).
Hashimoto's disease, or autoimmune thyroiditis type 2A, is the most common cause of goitrous hypothyroidism in areas of the world in which dietary iodine is sufficient. Before discussing this entity, it is important to redefine the term autoimmune thyroid disease (Table 12-3). Many use the term autoimmune thyroiditis to cover both primary myxedema (nongoitrous) and classic Hashimoto's disease (goitrous). These are differing clinical manifestations of the same disorder that is also closely related to autoimmune hyperthyroidism/Graves' disease.

Autoantibodies to the TSH receptor that act as TSH antagonists may be the cause of some cases of the thyroid atrophy seen in primary myxedema and are seen less often in goitrous Hashimoto's disease. However, both Graves' and Hashimoto's diseases may occur within the same families and may share human leukocyte antigen (HLA) and other genetic susceptibility haplotypes. Furthermore, thyroid failure occurs in some patients with Graves' disease and hyperthyroidism and even orbitopathy develop in some patients with Hashimoto's disease. Both types of patients may have autoantibodies to thyroglobulin, TPO, and the TSH receptor. Hence, the diseases must be closely related, and autoimmune thyroid disease can be viewed as a spectrum from hyperthyroidism to hypothyroidism.

To bring some clarity to this situation, we should also redefine the term chronic thyroiditis. In pathologic terms, thyroiditis implies the presence of both a mononuclear cell infiltrate and destruction of thyroid follicles. However, these are arbitrary criteria. The term chronic thyroiditis is more appropriately defined simply as evidence of “intrathyroidal lymphocytic infiltration” without the necessity of follicular damage. Because by this definition patients with both Graves' disease and Hashimoto's disease have thyroiditis, replacement of the term autoimmune thyroid disease with the more correct term autoimmune thyroiditis allows a simple classification for autoimmune thyroid disease (Table 12-3).

Until the demonstration of circulating thyroid antigenspecific T cells and thyroid autoantibodies, the diagnosis of Hashimoto's disease could be confirmed only by biopsy of the thyroid. The ease with which we can now demonstrate high levels of circulating antibodies and thyroid antigenspecific T cells in most patients with Hashimoto's disease led to the use of the term autoimmune thyroiditis, which, as explained earlier, we prefer to use for any mononuclear infiltrate.
Hashimoto’s disease is common and may be increasing in frequency. The mean incidence in women is in the order of 3.5 cases per 1000 people per year and in men is 0.8 cases per 1000 people per year. No age group is exempt, although the prevalence increases with age. Hashimoto’s disease is the most common cause of goitrous hypothyroidism in areas of iodine insufficiency.

Pathophysiology.

Impairment of hormone synthesis is due to apoptotic destruction of the thyroid cells. The sick cells exhibit a defect in organic binding of thyroid iodide, as evidenced by a positive perchlorate discharge test (see Chapter 11 for a discussion of pendrin). In addition, release of iodopeptides, mostly thyroglobulin, is enhanced by cell lysis. Approximately 90% of the thyroid gland must be destroyed before hypothyroidism develops. The presence of lymphocytic infiltration of the thyroid (hence the older term lymphocytic thyroiditis), circulating thyroid autoantibodies, and clinical or immunologic overlap with other diseases with autoimmune components indicate that Hashimoto’s disease is an autoimmune thyroid disorder.

The current understanding of autoimmune mechanisms has been discussed earlier in Chapter 11. However, autoimmune thyroiditis is characterized by thyroid cell apoptosis leading to follicular destruction rather than thyroid stimulation and thyroid cell hyperplasia. Although both autoantibodies to thyroid peroxidase (TPOAb) and thyroglobulin (TgAb) may be complement-fixing and cytotoxic, the thyroid gland is infiltrated by both B cells and by T cells; the latter are armed with Fas ligand and capable of destroying thyroid cells expressing Fas via apoptosis (Fig. 12-7) (see also Color Plate). In addition, other cell death pathways may be involved. Fas expression on thyroid cells may be secondary to elaboration of a variety of cytokines from T cells that undergo blast transformation when exposed to thyroid antigens (thyrotropin receptor,

TPO, and thyroglobulin), suggesting that cell-mediated autoimmune mechanisms are pathogenetically involved. Indeed, a T-cell clone specifically cytotoxic for autologous thyroid cells in a patient with Hashimoto’s disease (Fig. 12-8) is reminiscent of animal models of cytotoxic T cells associated with experimental autoimmune thyroiditis. These manifestations of autoimmunity in Hashimoto’s disease and other autoimmune thyroid disorders reflect a hereditary susceptibility to thyroid disease that allows the survival and persistence of B cells and T cells directed against thyroid antigens. The fact that infusion of interleukin-2 and lymphokine-activated killer cells causes progression or development of hypothyroidism in patients with detectable TPOAb is additional evidence of the autoimmune nature of this disease.

Histopathology.

The thyroid gland is pale and firm. Histopathologic changes vary in type and extent but usually consist of diffuse lymphocytic infiltration with germinal center formation, obliteration of thyroid follicles by widespread apoptosis, and fibrosis (Fig. 12-9). In most cases, there is destruction of epithelial cells and degeneration and fragmentation of the follicular basement membrane. The remaining epithelial cells may be larger and show oxyphilic changes in the cytoplasm; these so-called Askanazy cells are virtually pathognomonic.

In some cases, epithelial hyperplasia may be prominent. Colloid is sparse. The interstitial tissue is infiltrated with lymphocytes that may form typical lymphoid follicles with germinal centers. Plasma cells may be prominent. Fibrosis is generally present in the older lesions but not to the extent seen in Riedel’s thyroiditis.

Histologically, two variants can be distinguished:

1. The oxyphilic variant displays more oxyphilic change, less fibrosis, and striking infiltration with lymphocytes forming germinal centers.
2. The fibrous variant is infiltrated mainly with plasma cells and displays more fibrosis.

Clinical differences between the two types are described later, but there is no evidence that the etiologic mechanisms are different. In the past, the diagnosis of Hashimoto’s disease required the presence of Askanazy cells or lymphoid follicles, but now the primary observation should be follicular destruction, often with macroscopic feature of the follicular spaces. The degree of lymphocytic infiltration usually correlates with the levels of circulating thyroid autoantibodies. Unlike the situation of diabetes mellitus, formal linkage of specific histocompatibility antigens with autoimmune thyroid disease has been difficult to demonstrate.

Hashimoto’s disease occurs with increased frequency in Down’s syndrome and (probably) gonadal dysgenesis. The fact that thyroid cells can express HLA-DR antigens, at least as a secondary phenomenon, indicates the potential role of these cells in perpetuating the immune response and may be related to the propensity of autoimmune disease for certain HLA-DR subgroups. Hashimoto’s disease almost certainly is associated with a polygenic susceptibility, HLA being one gene involved. Efforts are under way to identify non-HLA susceptibility genes in families with autoimmune thyroid disease. In this regard, polymorphisms in the CTLA4 gene have been linked to the propensity to secrete thyroid autoantibodies and may, in turn, be an important risk factor for disease itself.

Nongenetic Risk Factors.

Many of the factors that have been identified as increasing the risk for Graves’ disease (pregnancy, drugs, age and sex, infection, and irradiation) apply equally to autoimmune thyroiditis. These are detailed in Chapter 11 and are briefly considered here.

Pregnancy.

The recognition of transient postpartum thyroiditis as an important clinical entity has also provided an example of the immune manipulation of thyroid disease with a predictable onset and recovery (see Chapter 11). Maternal microchimerism may be an important component of this risk analysis. The disease is essentially postpartum Hashimoto’s disease except for its transient nature. Data suggest that 8% to 10% of women experience thyroiditis in the postpartum period with a variety of consequences (see later). Pregnancy is, therefore, an important risk factor, with transient postpartum thyroiditis developing in some patients and thyroid failure
differentiated permanently or in the early years after pregnancy in a significant proportion.

**Iodine and Drugs.**

Iodine and iodine-containing drugs (such as amiodarone) precipitate autoimmune thyroiditis in susceptible populations. This form should be distinguished from direct blockade and destruction of the thyroid gland by iodine. The mechanism of this precipitation is unknown. However, much evidence accumulated in animal models suggests that increased iodination of thyroglobulin enhances its immunoreactivity and that T-cell-reactive peptides may also be more antigenic when iodinated.

---

**Differential Diagnosis.**

Differential diagnosis of Hashimoto's disease from other uncomplicated disorders of the thyroid is facilitated by the demonstration that high levels of thyroid autoantibodies are common and in higher concentrations than thyroglobulin autoantibodies. Sometimes part of a gland with autoimmune thyroiditis may look and feel like a firm thyroid nodule, and ultrasonography should be performed to resolve the issue.

**Laboratory Tests.**

The results of the common tests of thyroid function depend on stage of disease. Rarely, the tests may suggest thyroid hyperfunction with a suppressed TSH but without overproduction of hormone. The RAU may be increased, but serum T₃ and T₄ levels may remain normal. At this stage, the patient may be euthyroid. As the TSH level rises, the glandular response at first compensates for the impairment of hormone biosynthesis. With time, the ability of the thyroid to respond to TSH diminishes, and the RAU and serum T₃ level decline to subnormal values. The serum T₃ concentration, however, may be slightly increased, probably reflecting maximal stimulation of the failing thyroid by the increased serum TSH level. The foregoing sequence, with still normal T₃ and T₄ and increased TSH levels, reflects the development of mild (subclinical) hypothyroidism.

The diagnosis of Hashimoto's disease is confirmed by the presence of thyroid autoantibodies in the serum, usually in high levels. TPO autoantibodies are more common and in higher concentrations than thyroglobulin autoantibodies. Sometimes part of a gland with autoimmune thyroiditis may look and feel like a firm thyroid nodule, and ultrasonography should be performed to resolve the issue.

**Cytokines.**

Treatment of patients with interferon-2 or interferon may precipitate the appearance of autoimmune thyroid disease. Destructive thyrotoxicosis may appear suddenly, but persistent Graves' disease may also develop in such patients. Autoimmune thyroid disease is more common in patients with preexisting TPOAbs.

**Infection.**

There is no direct evidence that infection causes autoimmune thyroiditis in humans. A number of viral infections in animals, however, do precipitate thyroid autoimmunity. In addition, there is evidence from long-term follow-up of patients with subacute thyroiditis (see later), thought to be a reaction to a viral infection, that persistent signs of thyroid autoimmune disease can be found. Infection remains a likely cause of a local or distant insult that is considered to be needed to precipitate autoimmune disease in susceptible individuals (see "Mechanisms in Graves' Disease," Chapter 11).

**Clinical Picture.**

Goiter, the hallmark of classic Hashimoto's disease, usually develops gradually and may be found during routine examination or by ultrasonography. On occasion, the thyroid gland enlarges rapidly and, when accompanied by pain and tenderness, may mimic de Quervain's or subacute thyroiditis. Some patients, particularly those with the fibrous variant, are hypothyroid when first seen. The goiter is generally moderate in size and firm in consistency and moves freely on swallowing. The surface is either smooth or scalloped, but well-defined nodules are unusual. Both lobes are enlarged, but the gland may be asymmetrical. The pyramidal lobe may be enlarged, and adjacent structures, such as the trachea, esophagus, and recurrent laryngeal nerves, may be compressed. Enlargement of regional lymph nodes is unusual.

Although nongloutous Hashimoto's disease (atrophy hypothyroidism) is thought to be the end result of autoimmune destruction of the thyroid, the progression of goitrous Hashimoto's disease to the atrophied state is not commonly seen in the individual patient. Indeed, the histopathologic picture tends to remain rather static except for an increase in fibrous tissue.

Clinically, the untreated goiter remains unchanged or enlarges gradually over many years. The manifestations of hypothyroidism may develop over several years in patients who are initially euthyroid. Some, but not all, studies suggest an increased prevalence of thyroid carcinoma in Hashimoto's disease.

Occasionally, hyperthyroidism develops in patients with Hashimoto's disease. In other patients with early autoimmune thyroiditis, transient thyrotoxicosis (painless or silent thyroiditis with thyrotoxicosis) occurs as the result of thyroid cell destruction. In such cases, evidence of ongoing thyroid hyperfunction is lacking because the thyroid RAU is depressed. As described earlier, a phase of transient hypothyroidism begins 3 to 6 months post partum in 30% of women with autoimmune thyroiditis, as evidenced by the presence of TPOAbs. The history may suggest earlier mild thyrotoxicosis (see syndromes associated with transient hyperthyroidism in Chapter 11).

**Iodine and Drugs.**

Iodine and iodine-containing drugs (such as amiodarone) precipitate autoimmune thyroiditis in susceptible populations. This form should be distinguished from direct blockade and destruction of the thyroid gland by iodine. The mechanism of this precipitation is unknown. However, much evidence accumulated in animal models suggests that increased iodination of thyroglobulin enhances its immunoreactivity and that T-cell-reactive peptides may also be more antigenic when iodinated.
occurs more commonly in Hashimoto's disease than in other thyroid disorders. The frequent coexistence of hypothyroidism and Hashimoto's disease serves to distinguish this disease from nontoxic goiter and thyroid neoplasms.

Differentiation of euthyroid Hashimoto's disease from a diffuse nontoxic goiter is often difficult, although diffuse nontoxic goiter tends to be softer than that of Hashimoto's disease and ultrasonic examination may reveal the heterogeneity of Hashimoto's disease. In adolescents, differentiation of Hashimoto's disease from diffuse nontoxic goiter is even more difficult because in this age group Hashimoto's disease may not be accompanied by such high levels of thyroid autoantibodies. The presence of well-defined nodules usually distinguishes nontoxic multinodular goiter from Hashimoto's disease.

Differentiation between Hashimoto's disease and thyroid carcinoma can sometimes be made on clinical grounds. Thyroid carcinoma is usually nodular and firm or hard and may be fixed to adjacent structures. Compression of the recurrent laryngeal nerve with hoarseness is virtually pathognomonic of thyroid carcinoma but occurs late in the disease progression. A history of a recent enlargement of the goiter is more common in thyroid carcinoma than in Hashimoto's disease. Enlargement of regional lymph nodes also suggests thyroid carcinoma. In thyroid carcinoma, ultrasound examination or radioiodine scanning of the thyroid may reveal only the isolated lesion. In Hashimoto's disease, activity is usually heterogeneous.

Treatment.

In many patients, no treatment is required because the goiter is small and the disease is asymptomatic, with the TSH level remaining in the normal range (autoimmune thyroiditis type 1). In other patients, treatment with thyroid hormone is directed at alleviating goiter, hypothyroidism, or both (autoimmune thyroiditis type 2A).

Levothyroxine treatment is indicated in patients when the goiter presses on adjacent structures or is unsightly, and it is most effective in goiters of recent onset. In long-standing goiter, treatment with thyroid hormone is usually ineffective, possibly because of fibrosis.

Glucocorticoids may cause regression of the goiter and decrease autoantibody levels, but these agents are not recommended in the usual case because of untoward side effects and the return of activity after treatment is withdrawn.

Full-replacement doses of thyroid hormone should be given when hypothyroidism or subclinical hypothyroidism supervenes. Surgery is justified if pressure symptoms or unsightly enlargement persists after a trial of suppressive therapy. Administration of levothyroxine should be continued after surgery because hypothyroidism is inevitable. The importance of maintaining the serum TSH level within the normal range is covered later, under Treatment.

Iodine Deficiency (Endemic Goiter)

The term endemic goiter denotes any goiter occurring in a region where goiter is prevalent. As mentioned, endemic goiter almost always occurs in areas of environmental iodine deficiency. Although this condition is estimated to affect more than 200 million people throughout the world and is of major public health significance, it is most common in mountainous areas, such as the Alps, Himalayas, and Andes, or in the Great Lakes and Mississippi Valley regions of the United States, owing to the depletion of iodine consequent to the persistent glacial run-off in these regions.

The causative role of iodine deficiency in the genesis of endemic goiter is supported by the inverse correlation between the iodine content of soil and water and the incidence of goiter, the kinetics of iodine metabolism in patients with the disorder, and a decrease in incidence after iodine prophylaxis. The latter accounts for its absence in the population residing in the Great Plains region of the United States.

The occurrence of endemic goiter can be spotty, even within an area of known iodine deficiency; the role of dietary minerals or naturally occurring goitrogens and of pollution of water supplies has been suggested in instances of this type. For example, in the Cauca Valley of Colombia, waterborne goitrogens have been implicated, and in many areas of endemic iodine deficiency, consumption of cassava meal, which gives rise to thiocyanate, aggravates the iodine-deficient state by inhibiting thyroid iodide transport.

Most abnormalities in iodine metabolism in patients with endemic goiter are consistent with the expected effects of iodine deficiency (see Chapter 10, “Iodine Metabolism”). Thyroid iodide clearance rates and RAU are increased in proportion to the decrease in the urinary excretion of stable iodine. The absolute iodine uptake is normal or low. In areas of moderate iodine deficiency, the serum T4 concentration is usually in the lower range of normal; in areas of severe deficiency, however, values are decreased. Nevertheless, most patients in these areas do not appear to be in a hypothyroid state because of an increase in the synthesis of the catabolically more efficient T3 at the expense of T4, and because of an increase in the activity of thyroidal D1 and D2 (see Fig. 10-12).

The incidence and severity of endemic goiter and the metabolic state of the goitrous patient depend mainly on the degree of iodine deficiency. In the absence of hypothyroidism, the effects of the goiter are mainly cosmetic. When the goiter becomes nodular, however, hemorrhage into a nodule may cause acute pain and swelling, mimicking subacute thyroiditis or neoplasia. The goiter may also compress adjacent structures, such as the trachea, esophagus, and recurrent laryngeal nerves. The borderline nature of the iodine supply in many countries of western Europe is exemplified by the development in Belgium of compensatory goiter during pregnancy due to the increased requirement for thyroid hormone during gestation. It is possible that some pregnant patients may have inadequate iodine supplies. An annual injection of iodized oil is another effective means of administering iodine, and endemic goiter can be treated by the addition of iodine to communal drinking water.

Administration of iodine has little, if any, effect on a long-standing endemic goiter, but it causes the early endemic hyperplastic goiter of iodine deficiency to regress. Similarly, thyroid hormone usually has no effect on long-standing goiter or on established mental or skeletal changes, but it should be given in full-replacement doses if there is evidence of hypothyroidism. This is of paramount importance in pregnant women. Surgical treatment is indicated if the adjacent structures are compressed or if the goiter is either very large or is enlarging rapidly.

Endemic Cretinism

Endemic cretinism is a developmental disorder that occurs in regions of severe endemic goiter. Both parents of an endemic cretin are usually goitrous, and in addition to the features of sporadic cretinism described earlier, endemic cretins often have deaf-mutism, spasticity, motor dysfunction, and abnormalities in the basal ganglia demonstrable by magnetic resonance imaging.

Three types of cretins can be discerned: (1) hypothyroid cretins, (2) neurologic cretins, and (3) cretins with combined features of the two. The pathogenesis of neurologic cretinism is obscure but may be due to severe thyroid hormone deficiency during a critical early phase of central nervous system development in utero. Some cretins are goitrous, but the thyroid may also be atrophic, possibly as a consequence of exhaustion atrophy from continuous overstimulation or the lack of iodine.

Iodide Excess

Goiter and hypothyroidism, either alone or in combination, are sometimes induced by chronic administration of large doses of iodine in either organic or inorganic form (see Table 10-8). Iodide-induced goiter was formerly seen in patients with chronic respiratory disease, who were given potassium iodide as an
Patients with cystic fibrosis.
Patients with Graves' disease, especially after its treatment with radioiodine.

Among these groups, many, but not all, individuals display a positive iodide-thioracil discharge test, indicating a defect in the thyroid organic binding mechanism (see Chapter 10, "Iodine Metabolism"). However, intrinsic thyroid disease need not be present because a propensity to develop iodide goiter and hypothyroidism has also been demonstrated in patients who have undergone hemithyroidectomy for a solitary thyroid nodule in whom the remaining lobe was histologically normal. In these patients, as in those with Hashimoto's disease or Graves' disease studied prospectively, individuals with the highest basal serum TSH concentrations, even within the normal range, were those who developed iodide goiter. Iodinated contrast material, amiodarone and povidone iodine are common sources.

Goiter and hypothyroidism commonly occur in newborn infants born to women given large quantities of iodine during pregnancy, and death from neonatal asphyxia has been reported (see Fig. 10-13). In such cases, the mother is usually free from goiter. Pregnant women should not receive large doses of iodine (>1 mg/ day) over prolonged periods (>10 days), especially near term. Maternal amiodarone therapy causes thyroidal dysfunction in up to 20% of newborns. It is not known whether iodide goiter in newborns results from an inherent hypersensitivity of the fetal thyroid or from the fact that the placenta concentrates iodide several-fold or both.

As discussed earlier (see Chapter 10, "Regulation of Thyroid Function"), large doses of iodine cause an acute inhibition of organic binding that abates in the normal individual, despite continued iodine administration (acute Wolff-Chaikoff effect and escape). Iodide goiter appears to result from a more pronounced inhibition of organic binding and the failure of the escape phenomenon. As a consequence of decreased hormone synthesis and the consequent increase in TSH, iodide transport is enhanced. Because inhibition of organic binding is a function of the intrathyroidal concentration of iodide, a vicious circle, augmented by this increase in serum TSH, is set in motion.

The disorder usually appears as a goiter with or without hypothyroidism, although in rare instances iodine may produce hypothyroidism unaccompanied by goiter. Usually the thyroid gland is firm and diffusely enlarged, often greatly so. Histopathologic examination reveals intense hyperplasia. The FT4 concentration is low, TSH concentration is increased, and the 24-hour urinary iodine excretion and the serum inorganic iodide concentration are increased. The disorder regress after iodine is withdrawn. Thyroid hormone may also be given to relieve severe symptoms.

### Drugs Blocking Thyroid Hormone Synthesis or Release

Ingestion of compounds that block thyroid hormone synthesis or release may cause goiter with or without hypothyroidism. Apart from the agents used in the treatment of hyperthyroidism, antithyroid agents may be encountered either as drugs for the treatment of disorders unrelated to the thyroid gland or as natural agents in foods. Goiter with or without hypothyroidism can occur in patients given lithium, usually for bipolar manic-depressive psychosis. Like iodide, lithium inhibits thyroid hormone release, and in high concentrations can inhibit organic binding reactions. At least acutely, iodide and lithium act synergistically in the latter respect. The mechanisms underlying the several effects of lithium are uncertain; what differentiates patients who develop goiter during lithium therapy from those who do not is also unclear. Underlying autoimmune thyroiditis may be at least one factor because many patients with this combination have autoimmune thyroid disease.

Other drugs that occasionally produce goitrous hypothyroidism include para-aminosalicylic acid, phenylbutazone, aminoglutethimide, and ethionamide. Like the thionamides, these drugs interfere with both the organic binding of iodine and the later steps in hormone biosynthesis. Although soybean flour is not an antithyroid agent, soybean products in feeding formulas formerly resulted in goiter in infants by enhancing fecal loss of hormone, which, together with the low iodine content of soybean products, produced a state of iodine deficiency. Feeding formulas containing soybean products are now enriched with iodine.

Cigarette smoking increases the risk of hypothyroidism in patients with underlying autoimmune thyroid disease. Although the mechanism is unclear, certain components of cigarette smoke, including thiocyanate, hydroxyperidine, and benzopyrene derivatives, may be responsible.

Both the goiter and the hypothyroidism usually subside after the antithyroid agent is withdrawn. If continued administration of pharmacologic goitrogens is required, however, replacement therapy with thyroid hormone causes the goiter to regress.

### Goitrogens in Foods or as Endemic Substances or Pollutants

Antithyroid agents also occur naturally in foods. These are widely distributed in the family Cruciferae or Brassicaceae, particularly in the genus Brassica, including cabbages, turnips, kale, kohlrabi, rutabaga, mustard, and various plants that are not eaten by humans but that serve as animal fodder. It is likely that some thiocyanate is present in such plants (particularly cabbage). Cassava meal, a dietary staple in many regions of the world, contains linamarin, a cyanogenic glycoside, the metabolism of which leads to the formation of thiocyanate. Ingestion of cassava can accentuate goiter formation in areas of endemic iodine deficiency. Except for thiocyanate, dietary goitrogens influence thyroid iodine metabolism in the same manner as do the thiocyanides, which they resemble chemically; their role in the induction of disease in humans is uncertain. Waterborne, sulfur-containing goitrogens of mineral origin are believed to contribute to the development of endemic goiter in certain areas of Colombia.

A number of synthetic chemical pollutants have been implicated in causing goitrous hypothyroidism, including polychlorinated biphenyls and resorcinol derivatives. Perchlorate has also been noted in high concentrations in geographic regions in which explosives were made. It is not clear whether the concentrations are significant enough to produce hypothyroidism.

### Cytokines

Patients with chronic hepatitis C or various malignancies may be given interferon or interleukin-2. Such patients may experience hypothyroidism, which is usually transient but may persist. These agents activate the immune system and can induce a clinical picture suggesting an exacerbation of underlying autoimmune disease such as occurs during postpartum thyroiditis (see Chapter 11) and Chapter 37. Graves’ disease with hypothyroidism may also develop, and ablative therapy may be required to treat this condition. Patients with preexisting evidence of autoimmune thyroid disease who have positive TPO antibodies are probably at higher risk for this complication and should be monitored carefully during and after a course of treatment with either of these cytokines. Autoimmune hypothyroidism may also develop after successful treatment of Cushing's disease, presumably as a result of the release of the glucocorticoid-induced immunosuppression.

### Congenital Causes

Inherited defects in hormone biosynthesis are rare causes of goitrous hypothyroidism and account for only about 10% to 15% of the 1 in 3500 newborns with congenital hypothyroidism. In most instances, the defect appears to be transmitted as an autosomal recessive trait. Individuals with goitrous hypothyroidism are believed to be homozygous for the abnormal gene, whereas euthyroid relatives with slightly enlarged thyroids are presumably heterozygous. In the latter group, appropriate functional testing may disclose a mild abnormality of the same biosynthetic step that is defective in the homozygous individual. In contrast with nontoxic
goiter, which is more common in females than in males, these defects, as a group, affect females only slightly more commonly than males. Although goiter may be present at birth, it usually does not appear until several years later. Therefore, the absence of goiter in a child with functioning thyroid tissue does not exclude the presence of hypothyroidism. The goiter is initially diffusely hyperplastic, often intensely so, suggesting papillary carcinoma, but eventually becomes nodular. In general, the more severe the biosynthetic defect, the earlier the goiter appears, the larger it is, and the greater the likelihood of early development of hypothyroidism or even cretinism. Five specific defects in the pathways of hormone synthesis have been identified.

**Iodide Transport Defect**

Iodide transport defect is rare, a result of impaired iodide transport by the sodium-iodide symporter (NIS) protein mechanism and is reflected by defective iodide transport in the thyroid, salivary gland, and gastric mucosa. Administration of iodide, by raising the plasma concentration, increases the intrathyroidal concentration of iodide sufficiently to permit the synthesis of normal quantities of hormone, demonstrating that this is the cause of the deficiency.

**Defects in Expression or Function of Thyroid Peroxidase**

TPO is a protein that is required for normal synthesis of iodothyronines. Quantitative or qualitative abnormalities of TPO have been identified in 1 in 66,000 infants in the Netherlands. The most common of the 16 mutations identified in 35 families was a GGCC insertion in exon 8, leading to premature termination of TPO synthesis.

**Pendred’s Syndrome**

The most common presentation in patients with Pendred’s syndrome is a defect in iodine organification accompanied by sensory nerve deafness. The abnormality is in the PDS gene encoding pendrin, which is involved in the apical secretion of iodide into the follicular lumen (see Fig. 10-2 and Chapter 10, "Iodine Metabolism"). Thyroid function is only mildly impaired in this disorder.

**Defects in Thyroglobulin Synthesis**

Defects in the synthesis of thyroglobulin due to genetic causes are rare, having been identified only in a small number of families with congenital hypothyroidism. Some defects lead to premature termination of translation, whereas another defect causes deficiency in endoplasmic reticulum processing of the thyroglobulin molecule. The complex regulation and huge size of this gene makes screening for mutations a difficult task, and considerable work is still required to unravel the extent of the defects in this gene.

**Iodotyrosine Dehalogenase Defect**

The pathogenesis of goiter and hypothyroidism in the iodotyrosine dehalogenase defect is complex. The major abnormality is an impairment of both intrathyroidal and peripheral deiodination of iodotyrosines, presumably because of a lack (or dysfunction) of the iodotyrosine dehalogenase. The gene encoding this enzyme has yet to be identified.

As a consequence of intense thyroid stimulation and lack of intrathyroidal recycling of iodide derived from dehalogenation, iodide is rapidly accumulated by the thyroid gland and is rapidly released; moniodotyrosine (MIT) and diiodotyrosine (DIT) are elevated in plasma and, together with their deaminated derivatives, in the urine. Hypothyroidism is presumed to result from the loss of large quantities of MIT and DIT in the urine and to secondary iodine deficiency. The goiter and hypothyroidism are relieved by administration of large doses of iodide.

**Thyroid Infiltration**

A number of infiltrative or fibrosing conditions may cause hypothyroidism. Some are often associated with goiter, such as Riedel’s struma (see later). Others, such as amyloidosis, [hemochromatosis](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6101377/), or [scleroderma](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6101377/) may not be. Although the other manifestations of these conditions are usually obvious and hypothyroidism is only a complication, the presence of significant hypothyroidism without evidence of autoimmune thyroiditis should lead to a consideration of these rare causes of this condition.

**Atrrophic Hypothyroidism**

In some patients, manifestations of hypothyroidism are apparent but there is no obvious thyroid enlargement (atrophic hypothyroidism). This may be due to either acquired or congenital abnormalities, prominent among the former being Autoimmune Thyroiditis Type 2B (Table 12-3). The pathophysiology and thyroid function tests are similar to those found when goiter is present.

**Acquired Causes**

**Nongoitrous Hypothyroidism**

Hypothyroidism in the absence of a classic Hashimoto’s goiter has often been termed primary hypothyroidism (or myxedema); this condition is more common in women than in men and occurs most often between the ages of 40 and 60 years. Many years ago, the presence of circulating thyroid autoantibodies in almost all patients and the clinical and immunologic overlap with autoimmune diseases indicated that this represented the end stage of an autoimmune thyroiditis in which goiter either did not develop or went unnoticed (Autoimmune Thyroiditis 2B). Although most cases are due to autoimmune-induced apoptosis of the thyroid epithelial cells, some cases of nongoitrous hypothyroidism are also associated with TSH receptor antibodies that block the response of thyroid cells to endogenous TSH (see Chapter 11). In primary thyroid failure, the thyroid gland is not usually palpable but may be normal in size or even somewhat enlarged on sonography and of firm consistency. Circulating TPOAbs or TgAbs are detectable in most patients but may be absent in long-standing disease.

**Postablative Hypothyroidism**

Postablative hypothyroidism is a common cause of thyroid failure in adults. One type follows total thyroidectomy usually performed for thyroid carcinoma. Although functioning remnants may be present, as indicated by foci of radiiodine accumulation, hypothyroidism invariably develops. Another etiologic mechanism is subtotal resection of the diffuse goiter of Graves’ disease or multinodular goiter. Its frequency depends on the amount of tissue remaining, but continued autoimmune destruction of the thyroid remnant in patients with Graves’ disease may be a factor because some studies suggest a correlation between the presence of circulating thyroid autoantibodies in thyroiditis and the development of hypothyroidism after surgery. Hypothyroidism can be manifested during the first year after surgery, but, as with postradioiodine hypothyroidism, the incidence increases with time to approach 100%. In some patients, mild hypothyroidism appears during the early postoperative period and then may occasionally remit, as also occurs after radioiodine treatment.

Hypothyroidism after destruction of thyroid tissue with radiiodine is common and is the one established disadvantage of this form of treatment for hyperthyroidism in adults. Its frequency is determined, in large part, by the dose of radiiodine but is also influenced by variations in individual susceptibility, including autoimmune factors. The incidence of post-radioiodine hypothyroidism increases with time, approaching 100%. Although the FT₃ is low in patients with postablative hypothyroidism, serum TSH levels may be anomalously low for several months after either surgical or ¹³¹I-induced hypothyroidism if TSH synthesis has been suppressed for a long period prior to treatment.

Primary atrrophic thyroid failure may also develop in patients with Hodgkin’s disease after treatment with mantle irradiation or after high-dose neck irradiation for other forms of lymphoma or carcinoma. Surgical, radiiodine, or external beam therapy may also lead to a state of subclinical hypothyroidism, which usually...
(5 to 15 mU/L), low-to-normal FT₄, and a normal serum T₃ concentration (see Table 12-1).

### Congenital Causes

#### Thyroid Agenesis or Dysgenesis

Developmental defects of the thyroid are often responsible for the hypothyroidism that occurs in 1 in 3500 newborns. These defects may take the form of complete absence of thyroid tissue or failure of the thyroid to descend properly during embryologic development. Thyroid tissue may then be found anywhere along its normal route of descent from the foramen caecum at the junction of the anterior two thirds and posterior third of the tongue (lingual thyroid) to the normal site or below. Absence of thyroid tissue or its ectopic location can be ascertained by scintiscanning.

As indicated, a number of proteins are known to be required for normal thyroid gland development. These include the thyroid-specific transcription factor PAX8 as well as thyroid transcription factors 1 and 2 (TTF1 and 2). It might be anticipated that defects in one or more of these proteins may explain abnormalities in thyroidal development. These have been identified in several patients with PAX8 mutations, and a mutation in the human TTF2 gene was associated with thyroid agenesis, cleft palate, and choanal atresia. Despite a specific search, no mutations have been found in the TTF1 gene in infants with congenital hypothyroidism.

#### Thyroid Aplasia Due to Thyrotropin Receptor Unresponsiveness

Several families exist in which thyroid hypoplasia, high TSH concentrations, and a low free T₄ level are associated with loss-of-function mutations in the TSH receptor. The thyroid glands were in the normal location but did not trap perchentenate (TCO₄⁻). Somewhat surprisingly, thyroglobulin levels were still detectable. The molecular details of these patients are still under study.

A second type of abnormality that may cause TSH unresponsiveness is a mutation in the Gs protein that occurs in pseudohypoparathyroidism type 1A. These patients have inactivating mutations in the -subunit of the Gs protein and, consequently, mild hypothyroidism. Other as yet unexplained patients with elevated TSH levels and hypothyroidism in which the molecular nature of the defect has not been defined have been reported.

### Transient Hypothyroidism

Transient hypothyroidism is defined as a period of reduced FT₄, I with suppressed, normal, or elevated TSH levels that are eventually followed by a euthyroid state. This unusual form of hypothyroidism usually occurs in the clinical context of a patient with subacute (postviral), lymphocytic (painless), autoimmune, or postpartum thyroiditis. These conditions are reviewed in detail in Chapter 11.

The patient reports mild to moderate symptoms of hypothyroidism of short duration, and serum TSH concentrations are typically elevated, although not greatly so. The patient often has a preceding episode of symptoms consistent with mild or moderate thyrotoxicosis. If these symptoms cannot be elucidated from the history, it may be difficult to distinguish such patients from those with a permanent form of hypothyroidism. In the early phases of post-thyroiditis hypothyroidism, TSH concentrations may still be suppressed even though the FT₄ is low because of the delayed recovery of pituitary TSH synthesis, such as in patients with Graves’ disease or with toxic nodules who have undergone surgery and who have experienced rapid relief of hypothyroidism. In that situation, the TSH response to hypothyroidism may be suppressed for many months, in post-thyroiditis hypothyroidism, this period is rarely longer than a few weeks.

A significant fraction (33%) of women with autoimmune thyroiditis but normal thyroid function have episodes of hypothyroidism during the postpartum period. In some, the preceding hyperthyroidism is relatively asymptomatic, which can make an accurate clinical diagnosis difficult. Patients who have had an episode of typical subacute postviral thyroiditis with pain, tenderness, and hyperthyroidism are not difficult to recognize.

Diagnostic evaluation should include a determination of TSH, FT₄, I, and TPOAbs. Negative or low antibodies argue strongly for a nonautoimmune cause. This is significant, in that it may be possible for the patient not to be treated only temporarily for hypothyroidism. In such patients, a trial of a lower levothyroxine dosage after 3 to 6 months may reveal that thyroid function has recovered. This may also occur in patients with hypothyroidism that follows acute autoimmune thyroiditis (e.g., in the postpartum period), but it is somewhat less likely to occur because of the underlying progressive nature of the autoimmune thyroiditis.

In patients with hypothyroidism due to postviral thyroiditis, the thyroid gland is usually relatively small and atrophic. In patients with hypothyroidism that follows an episode of acute lymphocytic thyroiditis, the gland is usually slightly enlarged and somewhat firm, reflecting the underlying scarring and infiltration associated with that condition.

### Consumptive Hypothyroidism

Consumptive hypothyroidism is the term given to an unusual cause of hypothyroidism that has been identified in infants with visceral hemangiomas or related tumors. The first patient reported with this syndrome presented with abdominal distention caused by a large hepatic hemangioma with respiratory compromise secondary to upward displacement of the diaphragm. However, clinical signs suggested hypothyroidism, which was confirmed by finding a markedly elevated TSH level and undetectable T₄ and T₃ levels. The infant’s response to an initial IV infusion of levothyroxine was transient, leading to the decision to use parenteral thyroid hormone replacement to relieve the clinical hypothyroidism. The accelerated degradation of thyroid hormone was apparent from the fact that it required 96 µg of thyroid hormone replacement to relieve the clinical hypothyroidism. During this phase, the patient is euthyroid but has a modest increase in the serum TSH level.
The presence of an elevated glycoprotein -subunit in patients with pituitary tumor but not in those with thyroid hormone resistance.

3. Development of analogues of thyroid hormone with TR, as opposed to mixed or TR preferential effects, may eventually prove useful in treatment.

Cardiovascular manifestations of the condition.

Individuals this is not the case, suggesting that there may be mutations in coactivator proteins or one of the RXR receptors, which can also present in a similar fashion.

Laboratory results may be the first clear evidence that a patient otherwise thought to have hyperthyroidism has RTH. These tests show the unusual combination of an

Thyroid hormone action.

10% of such individuals. Other neuropsychological abnormalities have also been described.

Abnormalities in neuropsychological development exist, with an increased prevalence of attention deficit hyperactivity disorder, which is found in approximately 10% of such individuals. Other neuropsychological abnormalities have also been described.

Deafness in patients with RTH reflects the important role of

HESX1

in Chapter 7 and Chapter 8, and those causes with relatively specific thyroid-related deficiencies are mentioned here for completeness. In addition to pituitary tumors, hypothalamic disorders, and the like, an unusual cause of secondary hypothyroidism occurs in individuals given bexarotene (a retinoid X [RXR] receptor agonist) for T-cell lymphoma. This drug suppresses the activity of the human TSH -subunit promoter in vitro. Serum T4 concentrations are reduced about 50%, and patients experience clinical benefit from thyroid hormone replacement. Dopamine, dobutamine, high-dose glucocorticoids, or severe illness may suppress TSH release transiently, leading to a pattern of thyroid hormone abnormalities suggesting central hypothyroidism. As discussed earlier, this severe state of hypothalamic-pituitary-thyroid suppression is a manifestation of stage 3 illness.

Although these agents might be expected to have similar effects when given chronically, they do not; nor does somatostatin have a similar effect when given for acromegaly, although it does block the response of TSH to TRH and it has been administered to patients with thyrotropin-secreting pituitary adenomas.

Congenital defects in either the stimulation or the synthesis of TSH or in its structure have been identified as rare causes of congenital hypothyroidism. These include the consequences of defects in several of the homeobox genes, including POU1F1 (formerly termed Pit-1), PROP1, and HESX1. The latter factor is necessary for the development of the hypothalamus, pituitary, and olfactory portions of the brain, and its targeted deficiency in the mouse produces a condition resembling sepi-toptic dysplasia in humans.

Defects in POU1F1 and PROP1 cause hereditary hypothyroidism, usually accompanied by deficiencies in growth hormone and prolactin. One patient has been identified with a familial defect in the TRH receptor gene. All of these conditions are associated with the typical pattern of reduced FT4, T3, and TSH.

Structural defects in TSH have also been described. These include those with a mutation in the CAGYC peptide sequence of the -subunit, thought to be necessary for its association with the -subunit or defects that produce premature termination of the TSH -subunit gene. As mentioned, some of these abnormalities may be associated with elevations in TSH, suggesting the diagnosis of primary hypothyroidism, but the TSH molecule is immunologically, but not biologically, intact.

Resistance to Thyroid Hormone

Patients with resistance to thyroid hormone (RTH) may have features of hypothyroidism if the resistance is severe and affects all tissues. Alternatively, patients with RTH may have hyperthyroidism if the resistance is more severe in the hypothalamic-pituitary axis than in the remainder of the tissues. In clinical terms, patients in the former group are said to have generalized resistance to thyroid hormone, whereas patients in the latter group are said to have pituitary resistance to thyroid hormone. Patients with both forms almost always have mutations in one allele of the TR-beta (TR) gene that interfere with the capacity of that receptor to respond normally to T3, usually by reducing its binding affinity.

The mutations in the TR gene causing RTH cluster in three areas of the thyroid hormone receptor, which have been recognized to have important contacts with the hydrophobic ligand-binding domain cavity of TR as recognized from its crystal structure. The mutations do not interfere with the function of the DNA-binding domain, its co-repressor binding domain, or its region of heterodimerization with RXR. Some mutations affect the activation domain in the carboxy-terminus of the TR receptor.

RTH is probably produced by the heterodimerization of the mutant TR with RXR or homodimerization with a normal TR or TR. These mutant TR-containing dimers compete with wild-type TR-containing dimers for binding to the thyroid hormone response elements (TREs) of thyroid hormone-dependent genes.

Because these complexes bind co-repressor molecules that cannot be released in the absence of T3 binding, genes containing these TREs are more repressed than they would be normally at the prevailing concentrations of circulating thyroid hormones. Receptors that contain mutations in the activation domain may have a combination of both decreased affinity for T3 as well as impaired activating potential.

Thus, the mutant TR complex can interfere with the function of the three normal TR-expressing genes, producing a pattern termed dominant negative inhibition with an autosomal dominant pattern of inheritance. At least 400 families have been identified with this condition, and there are probably many more unreported cases. The gene frequency estimate is about 1 : 50,000, and the study of the function of the mutant receptors in this disorder has provided valuable insights into the mechanism of thyroid hormone action.

Patients with RTH usually are recognized because of thyroid enlargement, which is present in about two thirds of these individuals. Despite one's expectations, patients usually report a peculiar mixture of symptoms of hyperthyroidism and hypothyroidism. With respect to the heart, palpitations and tachycardia are more common than a reduced heart rate; however, patients may also demonstrate growth retardation and retarded skeletal maturation. This has been attributed to the fact that thyroid hormone effects in the heart appear to be primarily dependent on TR rather than TR, whereas the hypothalamic-pituitary axis is primarily regulated through TR, particularly TR2.

Abnormalities in neuropsychological development exist, with an increased prevalence of attention deficit hyperactivity disorder, which is found in approximately 10% of such individuals. Other neuropsychological abnormalities have also been described.

Because patients may present with symptoms suggesting hyperthyroidism, it is important to keep this diagnosis in mind in a patient with tachycardia, gaiter, and elevated thyroid hormones. RTH is discussed here because a reduced response to thyroid hormone is the biochemical basis for the condition. However, the laboratory results may be the first clear evidence that a patient otherwise thought to have hyperthyroidism has RTH. These tests show the unusual combination of an increased FT4, accompanied by normal or slightly increased TSH levels. Thus, the principal differential diagnosis is between a TSH-secreting pituitary tumor and RTH.

Factors that may assist in the differential diagnosis are as follows:

2. Normal thyroid hormone levels in family members of individuals with TSH-induced hyperthyroidism due to pituitary tumor.
3. Presence of an elevated glycoprotein -subunit in patients with pituitary tumor but not in those with thyroid hormone resistance.

A definitive diagnosis requires sequencing of the TR gene demonstrating the abnormality. Although virtually all patients with RTH have such abnormalities, in a few individuals this is not the case, suggesting that there may be mutations in coactivator proteins or one of the RXR receptors, which can also present in a similar fashion.

Treatment is difficult because thyroid hormone analogues designed to suppress TSH, thereby relieving the hyperthyroxinemia, may lead to worsening of the cardiovascular manifestations of the condition. Therapy with 3,5,3'-triiodothyroacetic acid (TRIAC) has been used in several patients. The development of analogues of thyroid hormone with TR, as opposed to mixed or TR preferential effects, may eventually prove useful in treatment.
Treatment

Hypothyroidism, either primary or central, is gratifying to treat because of the ease and completeness with which it responds to thyroid hormone. Treatment is nearly always with levothyroxine, and the proper use of this medication has been reviewed extensively. A primary advantage of levothyroxine therapy is that the peripheral deiodination mechanisms can continue to produce the amount of T <sub>3</sub> required under physiologic control. If one accepts the principle that replicating the natural state is the goal of hormone replacement, it is logical to provide the "prohormone" and allow the peripheral tissues to activate it by physiologically regulated mechanisms.

Pharmacologic and Physiologic Considerations

Levothyroxine has a 7-day half-life; about 80% of the hormone is absorbed relatively slowly and equilibrates rapidly in its distribution volume, therefore avoiding large postabsorptive perturbations in FT <sub>3</sub> levels. With its long half-life, omission of a single day's tablet has no significant effect and the patient may safely take an omitted tablet the following day. In fact, the levothyroxine dosage can be calculated almost as satisfactorily on a weekly, as on a daily, basis.

According to the U.S. Pharmacopeia, the levothyroxine content of replacement tablets must be between 90% and 110% of the stated amount, although narrower restrictions are being introduced in the United States. The availability in many countries of a multiplicity of tablet strengths with content ranging from 25 to 300 µg allows precise titration of the daily levothyroxine dosage for most patients with a single tablet, improving compliance significantly.

The typical dose of levothyroxine, approximately 1.6 to 1.8 µg/kg ideal body weight per day (0.7 to 0.8 µg/pound), generally results in the prescription of between 75 and 112 µg/day for women and 125 to 200 µg/day for men. Replacement doses need not be adjusted upward in obese patients. This dosage is about 20% greater than the T <sub>4</sub> production rate owing to incomplete absorption of the levothyroxine. In patients with primary hypothyroidism, these amounts usually result in serum TSH concentrations that are within the normal range. Because of the 7-day half-life, approximately 6 weeks is required before there is complete equilibration of the FT <sub>3</sub> and FT <sub>4</sub> levels.

Accordingly, assessments of the adequacy of a given dose or the effects of a change in dosage should not be made until this interval has passed.

By and large, levothyroxine products are clinically equivalent, although problems do occur. However, the variation permitted by the U.S. Food and Drug Administration in tablet content can result in slight variations in serum TSH in patients with primary hypothyroidism even when the same brand is used. Although the serum TSH level is an indirect reflection of the levothyroxine effect in patients with primary hypothyroidism, it is superior to any other readily available method of assessing the adequacy of therapy. Return of the serum TSH level to normal is therefore the goal of levothyroxine therapy in the patient with primary hypothyroidism. Some patients may require slightly higher or lower doses than generally used, owing to individual variations in absorption, and a number of conditions or associated medications may change levothyroxine requirements in patients with established hypothyroidism (see later).

In decades past, desiccated thyroid was successfully employed for the treatment of hypothyroidism and still accounts for a small fraction of the prescriptions written for thyroid replacement in the United States. Although this approach was successful, desiccated thyroid preparations contain thyroid hormone derived from animal thyroid glands that have significantly higher ratios of T <sub>3</sub> to T <sub>4</sub> than the 1:11 value in normal human thyroid gland. Accordingly, such preparations may lead to supraphysiologic levels of T <sub>3</sub> in the immediate postabsorptive period (2 to 4 hours) owing to the rapid release of T <sub>3</sub> from thyroglobulin, its immediate and nearly complete absorption, and the 1-day period required for T <sub>4</sub> to equilibrate with its 40-L volume of distribution (see Table 10-5).

Mixtures of liothyronine and levothyroxine (liothrix) contain in a 1-grain (64-mg) equivalent tablet (Thyrolar in the United States), the amounts of T <sub>3</sub> (12.5 µg) and T <sub>4</sub> (50 µg) present in the most popular desiccated thyroid tablet. The levothyroxine equivalency of a 1-grain desiccated thyroid tablet or its liothrix equivalent can be estimated as follows. The 12.5 µg of liothyronine (T <sub>3</sub>) is completely absorbed from desiccated thyroid or from liothrix tablets. Liothyronine is approximately 80% absorbed, and about 36% of the 40 µg of levothyroxine absorbed is converted to T <sub>3</sub>, with the molecular weight of T <sub>3</sub> (651) being 84% that of T <sub>4</sub> (777). Accordingly, a 1-grain tablet should provide about 25 µg of T <sub>3</sub> (12.5 ± 12.1), which would be approximately equivalent to that obtained from 100 µg of levothyroxine. This equivalency ratio can be used as an initial guide in switching patients from desiccated thyroid or liothrix to levothyroxine.

As indicated earlier, the use of levothyroxine as thyroid hormone replacement is a compromise with the normal pathway of T <sub>3</sub> production, in which about 80% of T <sub>3</sub> is derived from T <sub>4</sub> 5-monodeiodination and approximately 20% (6 µg) is secreted directly from the thyroid gland. Studies in thyroidecstomized rats, for example, show that it is not possible to normalize T <sub>3</sub> simultaneously in all tissues by an IV infusion of T <sub>4</sub>. However, it should be recalled from the earlier discussion of T <sub>4</sub> deiodination that the ratio of T <sub>3</sub>/T <sub>4</sub> in the human thyroid gland is about 0.09 but is 0.17 in the rat thyroid gland. Thus, about 40% of the rat's daily T <sub>3</sub> production is derived from the thyroid versus about 20% in humans. Accordingly, the demonstration that T <sub>3</sub> alone cannot provide normal levels of T <sub>3</sub> in all tissues in the rat is of interest but is not strictly applicable to thyroid hormone replacement in humans. Nonetheless, the ratio of T <sub>3</sub>/T <sub>4</sub> in the serum of a patient receiving levothyroxine as the only source of T <sub>3</sub> must be about 20% lower than that in a normal individual.

Similarly, the quantity of levothyroxine required to normalize TSH in an athyreotic patient results in a slightly higher serum T <sub>3</sub> concentration than is present in normal individuals. Although this may, to some extent, compensate for the lack of T <sub>3</sub> secretion, the fact that T <sub>3</sub> has an independent mechanism for TSH suppression owing to the intracellular generation of T <sub>3</sub> in the hypothalamic-pituitary-thyroid axis results in a portion of the feedback regulation being independent of the plasma T <sub>3</sub> concentration.

Does this slightly lower T <sub>3</sub> concentration in patients receiving levothyroxine make any difference physiologically? Probably not, although the question is difficult to answer definitively because the most readily measurable end point, TSH, cannot be used. In one study, patients who received 12.5 µg of T <sub>4</sub> as a substitution for 50 µg of their levothyroxine preparation scored, on average, somewhat higher on tests of mood than when they were taking levothyroxine alone. The dosage of thyroid hormone used in these studies was excessive, as judged by the fact that 20% of the group had serum TSH values below normal on either regimen and the test period was only a few months, making it difficult to extrapolate to the chronic replacement setting.

On the other hand, another study showed that the FT <sub>3</sub> I correlated as closely with the resting energy expenditure, as did TSH levels in a group of patients in whom small supplements or decrements in their ideal replacement levothyroxine dosage were made. The correlation with serum T <sub>3</sub> was not statistically significant, suggesting that in humans, perhaps as a result of differences in the peripheral metabolism of T <sub>4</sub>, from that in rodents, the FT <sub>3</sub> I may be as accurate as the TSH value as an index of satisfactory thyroid hormone replacement. The practical difficulty with the design of tablets providing combinations of T <sub>3</sub> and T <sub>4</sub> is that the approximate dose of 6 µg of T <sub>3</sub> provided would need to be released in a sustained fashion over 24 hours, which is quite different from the rapid absorption of T <sub>3</sub> with a peak at 2 to 4 hours when given in its conventional form. Thus, for the present, it appears that the current approach to thyroid replacement using levothyroxine, although not a perfect replication of the normal physiology, is satisfactory for most patients.

Institution of Replacement Therapy

The initial dose of levothyroxine prescribed depends on the degree of hypothyroidism and the age and general health of the patient. Patients who are young or middle-aged and otherwise healthy with no associated cardiovascular or other abnormalities and mild to moderate hypothyroidism (TSH concentrations 5 to 50 mIU/L)
can be given a complete replacement dose of about 1.7 µg/kg of ideal body weight. The resulting increase in serum \( T_4 \) concentration to normal requires 5 to 6 weeks, and the biologic effects of \( T_4 \) are sufficiently delayed that these patients do not experience adverse effects. At the other extreme, the older patient with heart disease, particularly angina pectoris, without reversible coronary lesions, should be given small initial doses of levothyroxine (25 or even 12.5 µg/day), and the dosage should be increased in 12.5 µg increments at 2- to 3-month intervals with careful clinical and laboratory evaluation.

The goal in the patient with primary hypothyroidism is to return serum TSH concentrations to normal, reflecting normalization of that patient's thyroid hormone supply. This usually results in a mid to high-normal serum FT\(_4\) level. The serum FT\(_4\) should be evaluated 6 weeks after a theoretically complete replacement dose has been instituted to allow minor adjustments to optimize the individual dose. In patients with central hypothyroidism, serum TSH is not a reliable index of adequate replacement and the serum FT\(_4\) should be restored to a concentration in the upper half of the normal range. Such patients should also be evaluated and treated for glucocorticoid deficiency before institution of thyroid replacement (see Chapter 8).

Although the adverse effects of the rapid institution of therapy are unusual, pseudotumor cerebri has been reported in profoundly hypothyroid juveniles between ages 8 and 12 years who were given even modest initial levothyroxine replacement. This complication appears 1 to 10 months after initiation of treatment and responds to acetazolamide and dexamethasone.

The interval between the initiation of treatment and the first evidence of improvement depends on the strength of dose given and the degree of the deficit. An early clinical response in moderate to severe hypothyroidism is a diuresis of 2 to 4 kg. The serum sodium (Na\(^+\)) level increases even sooner if hyponatremia was present initially. Thereafter, pulse rate and pulse pressure increase, appetite improves, and constipation may disappear. Later, psychomotor activity increases and the delay in the deep tendon reflex disappears. Hoarseness abates slowly, and changes in skin and hair do not disappear for several months. In individuals started on a complete replacement dose, the serum FT\(_4\) level should return to normal after 6 weeks; a somewhat longer period may be necessary for serum TSH levels to return to normal, perhaps up to 3 months.

In addition to myxedema coma (see later), it is sometimes clinically appropriate to alleviate hypothyroidism rapidly. For example, patients with severe hypothyroidism withstand acute infections or other serious illnesses poorly and myxedema coma may develop as a complication. In such circumstances, rapid repletion of the peripheral hormone pool in the average adult can be accomplished by a single IV dose of 500 µg of levothyroxine. Alternatively, by virtue of its rapid onset of action, levothyroxine (25 µg orally every 12 hours) can be administered if the patient can take medication by mouth. With both approaches, an initial effect is achieved within 24 hours. Parenteral therapy with levothyroxine is then continued with a dose that is 80% of the appropriate oral dose but not in excess of 1.4 µg/kg of ideal body weight. Because of the possibility that rapid increases in metabolic rate will overtax the existing pituitary-adrenocortical reserve, supplemental glucocorticoid (IV hydrocortisone 5 mg/hour) should also be given to patients with severe hypothyroidism receiving high initial doses of thyroid hormones. Finally, in view of the tendency of hypothyroid patients to retain free water, IV fluids containing only dextrose should not be given.

When replacement therapy is withdrawn for short periods (4 to 6 weeks) for purposes of evaluating therapy for thyroid cancer, rapid reinitiation of levothyroxine using a loading dose of three times the daily replacement dose for 3 days can usually be given unless there are other complicating medical illnesses. When hypothyroidism results from administration of iodine-containing or antithyroid drugs, withdrawal of the offending agent usually relieves both the hypothyroidism and the accompanying goiter, although it is appropriate to provide interim replacement until the gland recovers its function. This is especially true for amiodarone, which may remain in tissues for up to a year.

**Infants and Children**

In infants with congenital hypothyroidism, the determining factor for eventual intellectual attainment is the age at which adequate treatment with thyroid hormone is begun. The therapy for infants with congenital hypothyroidism should consist initially of raising the serum T\(_4\) level to more than 130 nmol/L (10 µg/dL) as rapidly as possible and maintaining it at that level for the first 3 to 4 years of life. This is usually accomplished by administering an initial levothyroxine dose of 50 µg/day, which is higher than the adult dose on a weight basis and in keeping with the higher metabolic clearance of the hormone in the infant. The serum TSH concentration may not return to normal even with this high dose because of residual reset of the pituitary feedback mechanism. After 2 years of age, however, a TSH level in the normal range is an index of optimal therapy as it is in adults.

**Monitoring Replacement Therapy**

Monitoring the adequacy of, and compliance with, thyroid hormone therapy in patients with primary hypothyroidism is easily done by measurement of serum TSH. This value should be within the normal range for an assay sufficiently sensitive to measure, with confidence, the lower limit of the normal range. The normal serum TSH concentration varies between 0.5 and 4.0 mU/L in most second-generation and third-generation assays, and results within this range are associated with the elimination of all clinical and biochemical manifestations of primary hypothyroidism, except in patients with RTH.

After the first 6 months of therapy, the dose should be reassessed because restoration of euthyroidism increases the metabolic clearance of T\(_4\). A dose that was adequate during the early phases of therapy may not be adequate when the same patient is euthyroid owing to an acceleration in the clearance of thyroid hormone. Under normal circumstances, the finding of a normal serum TSH level on an annual basis is adequate to ensure that the proper dose is prescribed and is being taken by the patient. If the serum TSH level is above the normal range and noncompliance is not the explanation, small adjustments, usually in 12-µg increments, can be made with reassessment of TSH concentrations after full equilibration (6 weeks) with the new dose to confirm that such adjustments are appropriate. In North America, this strategy is simplified by the availability of multiple tablet strengths, many of which differ by only 12 µg. Most patients can receive the same dose until they reach the 7th or 8th decade, at which point a downward adjustment of 20% to 30% is indicated because thyroid hormone clearance decreases in the elderly.

Thyroid hormone requirements may be altered in several situations. (Table 12-4). A reduction in replacement dosage may be required in women who are receiving androgen therapy for adjuvant treatment of breast carcinoma. Most other conditions or medications increase the levothyroxine requirement in patients receiving maintenance therapy. During pregnancy, the levothyroxine requirement is increased by 25% to 50%. Hypothyroid patients who are planning a pregnancy should be advised to increase the dose by up to 50% as soon as the diagnosis is confirmed because the change in requirement appears soon after implantation. The increased requirement is probably due to a combination of factors, including increases in thyroid hormone-binding globulin and the volume of distribution of T\(_4\), an increase in body mass, and an increase in D3 in placenta and perhaps the uterus. The increased requirement

<table>
<thead>
<tr>
<th>TABLE 12-4 — Conditions That Alter Levothyroxine Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Increased Levothyroxine Requirements</strong></td>
</tr>
</tbody>
</table>

**Pregnancy**

**Gastrointestinal Disorders**

- Mucosal diseases of the small bowel (e.g., sprue)

**After jejunocolic bypass and small-bowel resection**

**Diabetic diarrhea**

**Therapy with Certain Pharmacologic Agents**

**Drugs That Interfere with Levothyroxine Absorption**

- Diuretics
- Corticosteroids
- Opiates
- Anticonvulsants
- Oral contraceptives
- Miscellaneous

**Cholestyramine**
patients may have RTH in peripheral but not central tissues, a situation that has been documented only rarely.

In rare cases, hypothyroid symptoms are associated with hypometabolism despite normal levels of serum thyroid hormones and TSH.

To the relationship between symptoms of hypothyroidism and the role of thyroid hormone in relieving these, and other causes should be sought for the

In rare circumstances, symptoms consistent with hypothyroidism persist despite appropriate treatment of the hypothyroid state. Such patients should be educated as

Patients with Hypothyroid Symptoms Despite Restoration of Normal Thyroid Function

In rare circumstances, symptoms consistent with hypothyroidism persist despite appropriate treatment of the hypothyroid state. Such patients should be educated as
to the relationship between symptoms of hypothyroidism and the role of thyroid hormone in relieving these, and other causes should be sought for the
symptomatology. In rare cases, hypothyroid symptoms are associated with hypometabolism despite normal levels of serum thyroid hormones and TSH. Such
patients may have RTH in peripheral but not central tissues, a situation that has been documented only rarely.

<table>
<thead>
<tr>
<th>Drugs That Increase the Cytochrome P450 Enzyme (CYP3A4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rifampin</td>
</tr>
<tr>
<td>Carbamazepine</td>
</tr>
<tr>
<td>Estrogen</td>
</tr>
<tr>
<td>Phenytoin</td>
</tr>
<tr>
<td>Sertraline</td>
</tr>
<tr>
<td>? Statins</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Drugs That Block T4 to T3 Conversion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amiodarone</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conditions That May Block Deiodinase Synthesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selenium deficiency</td>
</tr>
<tr>
<td>Cirrhosis</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Decreased Levothyroxine Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aging (65 years and older)</td>
</tr>
<tr>
<td>Androgen therapy in women</td>
</tr>
</tbody>
</table>

T4, thyroxine; T3, triiodothyronine.

persists throughout pregnancy but returns to normal within a few weeks after delivery. Therefore, the dose should be reduced to the original pre-pregnancy level at
the time of delivery. Maternal T4 is critically important to the athyreotic fetus, and pregnant patients should be monitored carefully. Other conditions in which levothyroxine requirements are increased (Table 12-4) include malabsorption due to either bowel disease or adsorption of levothyroxine to coadministered medications such as sucralfate, aluminum hydroxide and perhaps calcium carbonate, ferrous sulfate, lovastatin, or various resins. Certain medications, notably rifampin, carbamazepine, phenytoin, and sertraline increase the clearance of levothyroxine by inducing CYP3A4 in the liver. Estrogen given to postmenopausal women may act in the same way, although the changes in thyroglobulin and distribution volume make the exact resolution of the cause of the increased levothyroxine requirement uncertain. Amiodarone increases levothyroxine requirements by blocking conversion of T4 to T3, and perhaps by interfering with T3-thyroid hormone receptor binding. Selenium deficiency is rare, but because it is rate-limiting in the synthesis of D1, a deficiency, such as may occur in patients receiving diets restricted in protein, may increase levothyroxine requirements.

Occasionally, in patients who have been treated with radioactive iodine for Graves’ disease or toxic nodular goiter, some degree of thyroid hormone secretion persists and, although inadequate, is autonomous. Such patients have a suppressed TSH on what otherwise would be considered a replacement dose of levothyroxine. The levothyroxine dose in these individuals should be reduced until TSH levels rise to normal, keeping in mind that several months may be required before TSH secretion recovers after its prolonged suppression. Because of

either the delayed effects of radiiodine or the natural history of Graves’ disease, per se, this autonomous T4 secretion may decrease with time, leading to an increase in levothyroxine requirements in subsequent years. Rarely, the opposite occurs: that is, a patient treated with radiiodine develops an increased TSH level, but, after several months of therapy, the requirement for such replacement is either reduced or eliminated. This may reflect transient impairment of thyroid function by a combination of preirradiation antithyroid drug therapy and immediate effects of radiation on the thyroid. In such patients, frequent monitoring of levothyroxine replacement is required to avoid over-replacement.

In North America, clinical experience with the most commonly used levothyroxine preparations suggests that these products are equally effective, and this is supported by small clinical trials. Nonetheless, the possibility of a problem with tablet levothyroxine content should be considered if a new preparation changes the biologic effects of the same dosage.

Adverse Effects of Levothyroxine Therapy

Although the administration of excessive doses of levothyroxine causes osteoporosis in postmenopausal patients, most authorities believe that returning thyroid status to normal does not have adverse effects on bone density. Administration of excessive doses also increases cardiac wall thickness and contractility and, in elderly patients, increases the risk of atrial fibrillation.

In some patients, TSH levels remain elevated despite the prescription of adequate replacement doses. This is most often a consequence of poor compliance. The combination of normal or even elevated serum FT4 values and elevated TSH levels can occur if the patient does not take levothyroxine regularly but ingests several pills the day before testing. The integrated dose of levothyroxine over prior weeks is best reflected in the serum TSH level, and noncompliant patients require careful education as to the rationale for the treatment. Subtle changes in dietary habits, such as increasing the ingestion of bran-containing products, may decrease levothyroxine absorption, and their recognition requires a careful history.

Patients with Hypothyroid Symptoms Despite Restitution of Normal Thyroid Function

In rare circumstances, symptoms consistent with hypothyroidism persist despite appropriate treatment of the hypothyroid state. Such patients should be educated as
to the relationship between symptoms of hypothyroidism and the role of thyroid hormone in relieving these, and other causes should be sought for the
symptomatology. In rare cases, hypothyroid symptoms are associated with hypometabolism despite normal levels of serum thyroid hormones and TSH. Such
patients may have RTH in peripheral but not central tissues, a situation that has been documented only rarely.
Special Aspects of Hypothyroidism

Subclinical Hypothyroidism

The term subclinical hypothyroidism designates a situation in which an asymptomatic patient has a low-normal FT₄ but a slightly elevated serum TSH level. Other terms for this condition are mild hypothyroidism, predichondal hypothyroidism, biochemical hypothyroidism, and decreased thyroid reserve (see Table 12-1). The TSH elevation in such patients is modest, with values typically between 5 and 15 mIU/L. This syndrome is most often seen in patients with early Hashimoto’s disease and is a common phenomenon, occurring in 7% to 10% of older women.

A number of studies on the effects of thyroid hormone treatment in such patients have used physiologic end points (e.g., measurements of various serum enzymes, systolic time intervals, serum lipids, psychometric testing), and results have been variable. In the most carefully controlled studies, one or another of the parameters has returned to normal in about 25% to 50% of patients. In general, FT₄ and TSH levels normalize, but FT₃, usually normal at the outset, does not change. In one study that employed a double-blind, crossover approach, the 4 of 17 women who improved could be differentiated from the remainder only by a somewhat lower serum free T₃ at the start of the study. Modest improvements in cardiac indices have been noted in some but not all reports, and the same is true for lipids. Thus, when one is confronted with this clinical situation, there is no clearly correct approach.

One factor favoring a decision to recommend levothyroxine therapy is the presence of antibodies to TPO or thyroglobulin or the presence of a goiter. There is a risk of progression of thyroid dysfunction in patients with Hashimoto’s disease, and this premonitory sign of thyroid failure is, to many, a justification for initiating therapy. To be weighed against this are the expense and bother of daily medication, not acceptable to some patients, and the possibility that overdosage with levothyroxine may exacerbate osteoporosis or cause cardiac arrhythmias. If a therapeutic trial is performed, the TSH concentration should be monitored carefully and should not be reduced below normal. If no therapy is given, such patients should be monitored at intervals of 6 to 12 months both clinically and by measurements of serum TSH.

Metabolic Insufficiency

Non-specific symptoms of true hypothyroidism include mild lassitude, fatigue, slight anemia, constipation, apathy, cold intolerance, menstrual irregularities, loss of hair, and weight gain (see Fig. 12-5). For this reason, some patients with such complaints but with normal laboratory results have been considered candidates for levothyroxine therapy. The response to thyroid hormone therapy is sometimes gratifying, at least initially, but symptomatic improvement usually disappears after a time unless the dose is increased. Eventually, even larger doses fail to alleviate the symptoms, confirming that they do not arise from a deficiency of thyroid hormone. Thus, thyroid hormone therapy should be avoided in patients with no biochemical documentation of impaired thyroid function. Furthermore, even in patients with subclinical hypothyroidism, symptoms may be out of proportion to FT₄ abnormalities. It is unwise to raise the patient’s expectations that such symptoms will be relieved by correction of mild biochemical abnormalities.

Thyroid Function Testing in Patients Receiving Replacement Therapy

Physicians are frequently confronted with patients receiving levothyroxine in whom the historical diagnosis of hypothyroidism has been made on what appear to be questionable grounds. In this circumstance, it may be impossible to determine, from retrospective clinical or laboratory findings, whether thyroid hormone replacement is indicated. If serum TSH is in the normal range and primary hypothyroidism is suspected, a simple way of assessing the need for levothyroxine therapy is to switch levothyroxine to every-other-day dosage or to reduce the daily dose by 50% and to reevaluate TSH and FT₄ after 4 weeks. If there has been no significant increase in TSH concentration and FT₄ remains constant during that period, levothyroxine is withdrawn and blood tests are repeated 4 and 8 weeks later.

If the initial TSH level is suppressed, indicating overreplacement, the dose should be reduced until TSH becomes detectable before this trial is instituted. If central hypothyroidism is suspected, the FT₄ must be monitored during these procedures.

Emergency Surgery in the Hypothyroid Patient

The perioperative course of patients with untreated hypothyroidism has been evaluated in several studies. In general, such patients were not recognized to be hypothyroid or did not require surgery despite the presence of significant hypothyroidism. Complications were uncommon. Perioperative hypotension, ileus, and central nervous system disturbances were more common in hypothyroid patients, and patients with major infections had fewer episodes of fever than did euthyroid control subjects. Other complications were delayed recovery from anesthesia and abnormal hemostasis, possibly owing to an acquired form of von Willebrand’s disease.

From these studies, one may conclude that emergency surgery should not be postponed in hypothyroid patients but that such patients should be rigorously monitored for evidence of carbon dioxide retention, bleeding, infection, and hyponatremia. These findings are also relevant to the treatment of hypothyroid individuals with symptomatic coronary artery disease. Considering the lack of significant increase in perioperative complications in the hypothyroid patient, the option of surgery for remendable coronary artery lesions is open to hypothyroid individuals without the risk of a myocardial infarction in association with restitution of the euthyroid state (see later).
Coexisting Coronary Artery Disease and Hypothyroidism

In many patients with coronary artery disease and primary hypothyroidism, cardiac function is corrected during institution of levothyroxine therapy because of a decrease in peripheral vascular resistance and improvement in myocardial function. However, patients with preexisting angina pectoris should be evaluated for correctable lesions of the coronary arteries and treated appropriately before levothyroxine is administered. Retrospective studies indicate that this approach is safer than the institution of replacement therapy prior to angiography and angioplasty or even coronary artery bypass grafting.

In a few patients, lesions may not be remediable or small-vessel disease is severe even after bypass grafting, so that complete replacement cannot be instituted. Such patients must receive optimal antianginal therapy combined with β-adrenergic receptor blockers in judicious quantities, and complete restitution of the euthyroid state may not be possible.

Thyroid Hormone for Compromised Cardiovascular Function

In addition to the issues raised in patients with combined hypothyroidism and coronary artery disease, there is interest in the potential therapeutic use of thyroid hormone in the treatment of patients with either cardiomyopathy or status postcoronary artery bypass grafting (CABG) or other cardiac procedures. As expected, T₃ levels are reduced in patients with advanced congestive heart failure, as with any illness. In one report, 23 patients with advanced heart failure (mean ejection fraction, 22%) were given up to 2.7 µg/kg of liothyronine over 6 hours with an increase in cardiac output and decrease in systemic vascular resistance but without increase in heart or metabolic rate. Similar effects were seen with a dose of liothyronine, 110 µg, over 6 hours after CABG. In addition, decreases in the frequency of atrial fibrillation following surgery were found in liothyronine-treated patients, although this was not confirmed in a second study. However, there were no adverse effects associated with the infusion of T₃ in either study.

Liothyronine has also been given postoperatively for congenital heart disease and, again, an improvement in cardiac output and decrease in vascular resistance occurred without adverse side effects. These results suggest that, in certain selected circumstances, liothyronine may be useful as adjunctive therapy in patients with congestive heart failure because of its effect of relaxing vascular smooth muscle. It is conceivable that more selective thyroid hormone analogues may be on the horizon that might produce this effect but not increase myocardial oxygen demands.
Screening for Primary Hypothyroidism

The high incidence of primary hypothyroidism in women, particularly if the 7% to 10% prevalence of subclinical hypothyroidism is included, raises the issue of whether the cost of systematic periodic screening of an asymptomatic population is justified. A number of studies have addressed this complex issue. The conclusions depend, to a great extent, on assumptions regarding the effectiveness and economic value of therapy in asymptomatic patients with TSH elevation alone. One study concluded that the cost of an every-5-year TSH determination for women and men would be approximately $9000 per quality-adjusted life-year in women. Other studies have indicated somewhat lower financial benefits. An assessment of TSH levels at 5-year intervals in women older than age 50 years seems justified, but further analyses of more extensive screening programs are in order.

A second complex issue involves whether women planning pregnancy should be screened for the presence of hypothyroidism as a routine part of a prenatal visit. This question is raised because of the results of a study indicating that mild to modest hypothyroidism is associated with an impairment of mental development in infants of mothers with elevated TSH during the first trimester. The prevalence of hypothyroidism during pregnancy has been found to be approximately 2%, and screening of all patients has been advocated by several professional organizations.

Maternal FT₄I concentrations in the lowest 10% of the normal range, even with normal TSH levels, have also been suggested as a risk factor for impaired neuropsychological development of the fetus. It is not clear why this is a risk factor for impaired fetal neuropsychological development, because such patients are not hypothyroid. For the moment, it appears that any patient with a family history of autoimmune thyroid disease, with symptoms suggesting hypothyroidism, or with thyroid enlargement should be screened for thyroid dysfunction prior to pregnancy or as soon after conception as is feasible.
Myxedema Coma

Myxedema coma is the ultimate stage of severe long-standing hypothyroidism. This state, which almost invariably affects older patients, occurs most commonly during the winter months and is associated with a high mortality rate. It is usually, but not always, accompanied by a subnormal temperature. Values as low as 23°C having been recorded. The external manifestations of severe myxedema, bradycardia, and severe hypotension are invariably present. The characteristic delay in deep tendon reflexes may be lacking if the patient is areflexic. Seizures may accompany the comatose state. Although the pathogenesis of myxedema coma is not clear, factors that predispose to its development include exposure to cold, infection, trauma, and central nervous system depressants or anesthetics. Alveolar hypoventilation, leading to carbon dioxide retention and narcosis, and dilutional hyponatremia resembling that seen with inappropriate secretion of arginine vasopressin (AVP) may also contribute to the clinical state.

From the foregoing, it appears that myxedema coma should be readily recognized from its clinical signs, but this is not the case. After a brain stem infarction, elderly patients with features suggestive of hypothyroidism may be both comatose and hypothermic. In addition, hypothermia of any cause, due for example to exposure to cold, may cause changes suggestive of myxedema, including delayed relaxation of deep tendon reflexes. The importance of the difficulty in diagnosing myxedema coma is that a delay in therapy worsens the prognosis. Consequently, the diagnosis should be made on clinical grounds, and, after sending blood for thyroid function tests, therapy should be initiated without awaiting the results of confirmatory tests because mortality may be 20% or higher.

Treatment consists of administration of thyroid hormone and correction of the associated physiologic disturbances. Because of the sluggish circulation and severe hypometabolism, absorption of therapeutic agents from the gut or from subcutaneous or intramuscular sites is unpredictable, and medications should be administered intravenously if possible. Administration of levothyroxine as a single IV dose of 500 to 800 µg serves to replete the peripheral hormone pool and may cause improvement within hours. Daily doses of IV levothyroxine, 100 µg, are given thereafter. Hydrocortisone (5 to 10 mg/hour) should also be given because of the possibility of relative adrenocortical insufficiency as the metabolic rate increases.

Alternatively, IV liothyronine may be given at a dose of 25 µg every 12 hours. Others have used a combination of 200 to 300 µg T₄ and 25 µg T₃ intravenously as a single dose, followed by 25 µg T₃ and 100 µg T₄ 24 hours later, and then 50 µg T₄ daily until the patient regains consciousness. Hypotonic fluids should not be given because of the danger of water intoxication owing to the reduced free water clearance of the hypothyroid patient. Hypertonic saline and glucose may be required to alleviate severe dilutional hyponatremia and the occasional hypoglycemia.

A critical element in therapy is support of respiratory function by means of assisted ventilation and controlled oxygen administration. Internal warming by gastric perfusion may be useful, but external warming should be avoided because it may lead to vascular collapse due to peripheral vasodilatation. Further heat loss can be prevented with blankets. An increase in temperature may be seen within 24 hours in response to levothyroxine. General measures applicable to the comatose patient should be undertaken, such as frequent turning, prevention of aspiration, and attention to fecal impaction and urinary retention.

Finally, the physician should assess the patient for the presence of coexisting disease, such as infection and cardiac or cerebrovascular disease. In particular, the myxedematous patient may be afebrile despite a significant infection. As soon as the patient is able to take medication by mouth, treatment with oral levothyroxine should be instituted.
THYROIDITIS

Overview

Thyroiditis is a term indicating the presence of thyroid inflammation, and thus comprises a large group of diverse inflammatory conditions. These include the following:

1. Autoimmune or quasi-autoimmune causes.
2. Viral or postviral conditions.
3. Infections, including those of bacterial and fungal origins.
4. A chronic sclerosing form of thyroiditis, termed Riedel's thyroiditis (or struma).
5. Miscellaneous causes of various types, including radiation-induced and granulomatous causes, such as sarcoidosis.

Not only are the causes of thyroiditis extremely varied; their clinical presentations may also be diverse and are difficult to categorize in a simple fashion (Table 12-5). Thus, as already discussed, autoimmune thyroiditis may present with hypothyroidism but often patients remain euthyroid for long periods after the disease is initiated. On the other hand, in a euthyroid patient with Hashimoto's disease who becomes pregnant, the postpartum period is often complicated by an acute form of hyperthyroidism due to the transient exacerbation of thyroiditis, often followed by a period of hypothyroidism.

A similar syndrome has been observed in nonpregnant patients, called silent or painless thyroiditis. It is manifested primarily as thyrotoxicosis of sudden onset without localized pain and often without evidence of autoimmune disease. This condition may be viral in origin in some patients; however, the most classic presentation of viral thyroiditis is as subacute, nonsuppurative thyroiditis, also known as de Quervain's thyroiditis, pseudotuberculous thyroiditis, or migratory or creeping thyroiditis. Unlike typical autoimmune thyroiditis, this condition is characterized by extreme thyroid tenderness, with pain radiating to the oropharynx and ears, and must be differentiated from acute suppurative thyroiditis caused by bacterial or fungal infection. [306]

Thus, inflammatory conditions of the thyroid present a dilemma because one must decide whether to discuss these entities as a group with the common denominator of inflammation or to categorize them according to their principal clinical effects, namely thyrotoxicosis or thyroid hormone deficiency. We have chosen the latter approach and have already discussed autoimmune thyroiditis, the major cause of thyroid gland failure (see Table 12-2). However, patients with acute autoimmune thyroiditis may also develop thyrotoxicosis, such as in postpartum silent or painless thyroiditis (see Chapter 11 on autoimmune thyroiditis). These patients must be differentiated from those with Graves' disease. In addition, some patients with viral thyroiditis have thyrotoxicosis as a major manifestation with varying degrees of neck discomfort ranging from none to full-blown subacute, nonsuppurative (granulomatous) thyroiditis. For that reason, this thyroiditis syndrome is also discussed in Chapter 11, even though the pain associated with the typical form of this condition makes the principal differential diagnosis lie between that and pyogenic thyroiditis. In that context, subacute thyroiditis is also mentioned later.

---

**TABLE 12-5 -- Causes of Thyroiditis**

<table>
<thead>
<tr>
<th>Category</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autoimmune thyroiditis (see Table 12-3)</td>
<td></td>
</tr>
<tr>
<td>Postpartum, silent, or painless thyroiditis</td>
<td>(see Chapter 11)</td>
</tr>
<tr>
<td>Subacute (nonsuppurative) thyroiditis</td>
<td>(see Chapter 11)</td>
</tr>
<tr>
<td>Acute infectious thyroiditis</td>
<td></td>
</tr>
<tr>
<td>Riedel's thyroiditis</td>
<td></td>
</tr>
<tr>
<td>Miscellaneous</td>
<td></td>
</tr>
<tr>
<td>Postirradiation (¹³¹ I or external-beam therapy)</td>
<td></td>
</tr>
<tr>
<td>Sarcoïdosis</td>
<td></td>
</tr>
</tbody>
</table>
Acute Infectious Thyroiditis

Although the thyroid gland is remarkably resistant to infection, congenital abnormalities of the piriform sinus, underlying autoimmune disease, or immunocompromise of the host may lead to the development of an infectious disease of the thyroid gland. The etiology may be any bacterium, including Staphylococcus, Pneumococcus, Salmonella, or Mycobacterium tuberculosis. In addition, infections with certain fungi, including Coccidioides immitis, Candida, or Aspergillus and Histoplasma have been reported.

The most common cause of repeated childhood pyogenic thyroiditis, particularly in the left lobe, is a consequence of an internal fistula extending from the piriform sinus to the thyroid. This sinus is the residual connection following the path of migration of the ultimobranchial body from the fifth pharyngeal pouch to the thyroid gland. The predominance of thyroiditis of the left lobe is explained by the fact that the right ultimobranchial body is often atrophic, whereas this is not the case for the left side. Nonetheless, a patient with a completely normal thyroid gland may develop bacterial thyroiditis. This is an extremely rare disease even as a complication of direct puncture of the thyroid gland, such as in fine-needle aspiration. In individuals with midline infections, persistence of the thyroglossal duct should be considered.

Incidence

Infectious thyroiditis is extremely rare, with no more than a few cases being seen in large tertiary care centers.

### TABLE 12-6 -- Features Useful in Differentiating Acute Suppurative Thyroiditis and Subacute Thyroiditis

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Acute Thyroiditis</th>
<th>Subacute Thyroiditis</th>
</tr>
</thead>
<tbody>
<tr>
<td>History</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preceding upper respiratory infection</td>
<td>88%</td>
<td>17%</td>
</tr>
<tr>
<td>Fever</td>
<td>100%</td>
<td>54%</td>
</tr>
<tr>
<td>Symptoms of thyrotoxicosis</td>
<td>Uncommon</td>
<td>47%</td>
</tr>
<tr>
<td>Sore throat</td>
<td>90%</td>
<td>36%</td>
</tr>
<tr>
<td>Physical examination of the thyroid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Painful thyroid swelling</td>
<td>100%</td>
<td>77%</td>
</tr>
<tr>
<td>Left side affected</td>
<td>85%</td>
<td>Not specific</td>
</tr>
<tr>
<td>Migrating thyroid tenderness</td>
<td>Possible</td>
<td>27%</td>
</tr>
<tr>
<td>Erythema of overlying skin</td>
<td>83%</td>
<td>Not usually</td>
</tr>
<tr>
<td>Laboratory</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elevated white blood cell count</td>
<td>57%</td>
<td>2550%</td>
</tr>
<tr>
<td>Elevated erythrocyte sedimentation rate (&gt;30 mm/hr)</td>
<td>100%</td>
<td>85%</td>
</tr>
<tr>
<td>Abnormal thyroid hormone levels (elevated or depressed)</td>
<td>510%</td>
<td>60%</td>
</tr>
<tr>
<td>Alkaline phosphatase, transaminases increased</td>
<td>Rare</td>
<td>Common</td>
</tr>
<tr>
<td>Needle aspiration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Purulent, bacteria or fungi present</td>
<td>100%</td>
<td>0</td>
</tr>
<tr>
<td>Lymphocytes, macrophages, some polyps, giant cells</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>I uptake low</td>
<td>Uncommon</td>
<td>100%</td>
</tr>
<tr>
<td>Radiologic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abnormal thyroid scan</td>
<td>92%</td>
<td></td>
</tr>
<tr>
<td>Thyroid scan or ultrasound helpful in diagnosis</td>
<td>75%</td>
<td></td>
</tr>
<tr>
<td>Gallium scan positive</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Barium swallow showing fistula</td>
<td>Common</td>
<td>0</td>
</tr>
<tr>
<td>CT scan useful</td>
<td>Rarely</td>
<td>Not indicated</td>
</tr>
<tr>
<td>Clinical course</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical response to glucocorticoid treatment</td>
<td>Transient</td>
<td>100%</td>
</tr>
<tr>
<td>Incision and drainage required</td>
<td>85%</td>
<td>No</td>
</tr>
<tr>
<td>Recurrence following operative drainage</td>
<td>18%</td>
<td>No</td>
</tr>
<tr>
<td>Piriform sinus fistula discovered</td>
<td>96%</td>
<td>No</td>
</tr>
</tbody>
</table>


Clinical Manifestations

The clinical manifestations of infectious thyroiditis are dominated by local pain and tenderness in the affected lobe or entire gland. This is accompanied by painful swelling and difficulty on swallowing. Because of the tendency for referral of pain to the pharynx or ear, the patient may not recognize the tenderness in the anterior neck. Depending on the virulence of the organism and the presence of septicemia, symptoms such as fever and chills may also accompany the condition.

The major differential diagnosis lies between an infectious form of thyroiditis and subacute, nonsuppurative thyroiditis. It is instructive to compare the principal features of these two diseases to arrive at an accurate diagnosis (Table 12-6). By and large, patients with acute thyroiditis caused by a bacterium are much sicker than patients with subacute thyroiditis; they have more severe and localized tenderness and are less likely to have laboratory evidence of hyperthyroidism, which is present in approximately 60% of patients with subacute thyroiditis. Ultrasonographic examination often reveals the abscess in the thyroid gland or evidence of swelling, and needle aspiration may help pinpoint the responsible organism. A gallium scan will be positive as a result of the diffuseness of the inflammation and, particularly in children with thyroiditis of the left lobe, a barium swallow showing a fistula connecting the piriform sinus and left lobe of the thyroid is diagnostic.

Occasionally, pertechnetate scanning is useful in showing normal function of one lobe of the thyroid gland, which is much less common in subacute thyroiditis (which more often affects the entire gland). Needle aspiration should be used to drain the affected lobe, although occasionally surgical drainage may be required. If a piriform sinus fistula can be demonstrated, it must be removed to prevent recurrence of the problem.
gland have been reported.

The prognosis is excellent with preservation of thyroid function in general, although post-thyroiditis thyroid function tests should be monitored to ascertain that thyroid failure has not occurred.
Riedel's Thyroiditis

Riedel's chronic sclerosing thyroiditis is rare and dramatic and occurs chiefly in middle-aged women. The etiologic mechanism is uncertain, although some cases are considered to be an advanced state of Hashimoto's disease. This condition is characterized by fibrosis of the thyroid gland and adjacent structures and may be associated with fibrosis elsewhere, especially in the retroperitoneal area. The presence of eosinophils has been demonstrated histologically, suggesting a unique autoimmune response to fibrous tissue.

Symptoms develop insidiously and are related chiefly to compression of adjacent structures, including the trachea, esophagus, and recurrent laryngeal nerves. Constitutional evidence of inflammation is uncommon. The thyroid gland is moderately enlarged, stony hard, and usually asymmetrical. The consistency of the gland and the invasion of adjacent structures suggest carcinoma, but there is no enlargement of regional lymph nodes. Temperature, pulse, and leukocyte count are normal. Severe hypothyroidism is unusual but does occur, as does loss of parathyroid function. The RAIU may be normal or low. Circulating thyroid autoantibodies are less common and are found in lower titers than in Hashimoto's disease.

Surgery may be required to preserve tracheal and esophageal function. If extensive involvement of perithyroid tissues is present, resection of the isthmus may relieve some symptoms. Treatment with thyroid hormone relieves the hypothyroidism but has no effect on the primary process, which may progress inexorably. Immunosuppressive treatment and even chemotherapy has been tried in individual cases.
Miscellaneous Causes

Only a few causes of generalized inflammation of the thyroid gland have been reported. These include inflammation arising after $^{131}$I treatment of individuals for Graves’ disease, a residual thyroid lobe in a patient with thyroid cancer of the contralateral lobe, and thyroiditis arising from external-beam therapy for conditions such as Hodgkin’s or non-Hodgkin’s lymphoma, breast carcinoma, or other lesions of the oropharynx. In general, only radiiodine-induced thyroiditis is associated with pain and glucocorticoid treatment may be useful in symptomatic therapy.
References


138. Lazarus JH. The effects of lithium therapy on thyroid and thyrotropin-releasing hormone. Thyroid 1998; 8:909-913.
145. Brucker-Davis F. Effects of environmental synthetic chemicals on thyroid function. Thyroid 1998; 8:827-856.
147. Koh LK, Greenspan FS, Yeo PP. Interferon-alphainduced thyroid dysfunction: three clinical presentations and a review of the literature. Thyroid 1997; 7:891-896.
References

41. Lapl P, MacLauric PE, Chinato L, et al. Mutations in the encoding thyroid transcription factor-1 (TTF-1) are not a frequent cause of congenital hypothyroidism (CH) with thyroid dysgenesis. Thyroid 1997; 7:383387.


254. Weiss RE, Stein MA, Refetoff S. Behavioral effects of liothyronine (L-T3) in children with attention deficit hyperactivity disorder in the presence and absence of resistance to thyroid hormone. Thyroid 1997; 7:385393.


264. Wienslund A. Variation in serum levels of T3, T4, FT4, and TSH during thyroxine replacement therapy. Acta Endocrinol (Copenh) 1986; 113:4749.


305. Ross DS. Hyperthyroidism, thyroid hormone therapy, and bone. Thyroid 1994; 4:319326.


Chapter 13 - Nontoxic Goiter and Thyroid Neoplasia

Martin-Jean Schlumberger
Sebastiano Filetti
Ian D. Hay

After thyroid dysfunction and neck pain, the discovery of an apparent structural abnormality of the thyroid gland is the most common reason for a patient to seek the expertise of a clinical thyroidologist. In this chapter, we review the imaging techniques available for evaluating thyroid structural abnormalities; the units of measurement used in evaluation of the radiation dose and radioactivity are defined in Table 13-1.

Goiter resulting in thyrotoxicosis and other thyroid conditions arising from autoimmune thyroid disease are considered in Chapter 11 and Chapter 12.

This chapter describes simple or nontoxic goiter in addition to the increasingly recognized problem of nodular thyroid disease. Moreover, thyroid neoplasia, both benign and malignant, is discussed authoritatively. We consider an appropriate histologic classification and staging of thyroid cancer and present a management program for the most common thyroid cancer types.
EVALUATION OF STRUCTURAL ABNORMALITIES BY IMAGING TECHNIQUES

External Scintiscanning

Localization of functioning or nonfunctioning thyroid tissue in the area of the thyroid gland or elsewhere is made possible by techniques of external scintiscanning. The underlying principle is that isopes that are selectively accumulated by thyroid tissue can be detected and quantified in situ and the data transformed into a visual display. Two types of apparatus are available.

The rectilinear scanner is a device that moves a highly collimated (focused) scintillation detector back and forth across the area of study in a series of parallel tracks. A printing device records a mark whenever a predetermined number of counts has been received to provide a visual representation of the localization of radioactivity.

The stationary scintillation camera has now replaced the rectilinear scanner in most centers. It is equipped with a pinhole collimator that views the entire field of interest and translates the counting rates from specific areas of the field into images. Radioactivity in specific areas can be quantified. These cameras provide better resolution than rectilinear scanners, but anatomic localization may be more difficult.11

Several radioisotopes are employed in thyroid imaging. Technetium 99m (99mTc) pertechnetate is a monovalent anion that is actively concentrated by the thyroid gland but undergoes negligible organic binding and diffuses out of the thyroid gland as its concentration in the blood decreases. The short physical half-life of 99mTc (6 hours), its low fractional uptake, and its transient stay within the thyroid make the radiation delivered to the thyroid gland by a standard dose very low. Consequently, the intravenous administration of large doses (>37 MBq [1 mCi]) permits, about 30 minutes later, adequate imaging of the thyroid.

Two radioactive isotopes of iodine have been used in thyroid imaging. Iodine 131 (131I) was commonly used in the past and is still useful when functioning metastases of thyroid carcinoma are being sought; however, 131I is a beta emitter, its physical half-life is 8.1 days, and the energy of its main gamma ray is high and thus poorly adapted for its detection.12 131I is, in many respects, ideal but is expensive. The energy of its main gamma ray is adapted for its detection by gamma cameras. Its short half-life (0.55 day) and the absence of beta radiation result in a radiation dose to the thyroid that is about 1% of that delivered by a comparable activity of 123I. It is the isotope of choice for thyroid scintigraphy in pediatric practice.

The most important use of scintigraphic imaging of thyroid tissue is to define areas of increased or decreased function ("hot" or "cold" areas, respectively) relative to function of the remainder of the gland, provided that they are 1 cm in diameter or larger. Almost all malignant nodules are hypofunctioning, but more than 80% of benign nodules are also nonfunctioning. Conversely, functioning nodules (hot nodules), particularly if they are either more active than surrounding tissue or the sole functioning tissue, are rarely malignant.

In the past, several nuclear medical tests were used to evaluate thyroid disorders. In patients with a single area of thyroid uptake, scintiscans after administration of exogenous thyrotropin (TSH) may demonstrate the presence of hemiagenesis of the thyroid or document the functional capability of suppressed thyroid tissue. Conversely, scans performed after a period of exogenous thyroid hormone administration (suppression scans) can reveal areas of autonomous function that may not be detectable in baseline studies. These tests should no longer be used because the use of sensitive TSH assays and of scanning with a gamma camera permits the diagnosis of most of these hot nodules. Scintiscanning with radioactive iodine can also be used to demonstrate that intrathoracic masses represent thyroid tissue, to detect ectopic thyroid tissue in the neck, and to detect functioning metastases of thyroid carcinoma.

The choice of the scanning agent depends on many factors. 99mTc pertechnetate delivers a small dose of radiation to the thyroid gland, is readily available, and is inexpensive. Because imaging is performed soon after administration of the scanning agent, the entire procedure requires only a single visit to the laboratory. However, 5% to 10% of thyroid tumors appear to be functioning when examined with 99mTc pertechnetate but not with radiiodine. Because 99mTc pertechnetate imaging is done early, the intravascular activity and the activity in salivary tissue may obscure or confuse the findings. For the same reason, 123I Tc pertechnetate is inappropriate for scanning subternal or intrathoracic goiter or for detecting ectopic tissue in the neck. In these cases, radioactive iodine should be used.

Total-body scanning is performed with 131I in the follow-up of patients with papillary and follicular thyroid carcinoma. As detailed subsequently, radiiodine uptake by neoplastic tissue may be found only after TSH stimulation and is always lower than in normal thyroid tissue. For this reason, sufficiently high doses of 131I should be given, and scanning should be performed 2 to 3 days after the dose (or even later), when background blood activity is low and when the contrast is optimal. Scanning conditions should be optimized, preferably by use of a gamma camera with two opposed heads equipped with thick crystals and high-energy collimators.

Scanning at low speed with spot images on regions of interest is performed. There are two aims: (1) to verify the completeness of ablation and to detect and localize foci of uptake, and (2) to quantify any uptake. This quantification permits a dosimetric evaluation that indicates the usefulness of 131I treatment.

TABLE 13-1 -- Radiation Nomenclature: Traditional and International System (SI) Units

<table>
<thead>
<tr>
<th>Radiation dose</th>
<th>Abbreviation</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Gy = 100 rad = absorption of 1 joule/kg</td>
<td>Gy = gray</td>
<td></td>
</tr>
<tr>
<td>1 rad = 0.01 Gy = 1 cGy</td>
<td>rad = radiation absorbed dose</td>
<td></td>
</tr>
<tr>
<td>1 Sv = 100 rem</td>
<td>rem = roentgen-equivalent-man</td>
<td></td>
</tr>
<tr>
<td>1 Bq = 1 disintegration per second</td>
<td>mCi = millicurie</td>
<td></td>
</tr>
<tr>
<td>1 mCi = 37 MBq</td>
<td>kBq = becquerel</td>
<td></td>
</tr>
<tr>
<td>1 GBq = 10^9 MBq = 10^9 kBq = 10^9 Bq</td>
<td>MBq = megabecquerel</td>
<td></td>
</tr>
<tr>
<td>GBq = gigabecquerel</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The most important use of scintigraphic imaging of thyroid tissue is to define areas of increased or decreased function ("hot" or "cold" areas, respectively) relative to function of the remainder of the gland, provided that they are 1 cm in diameter or larger. Almost all malignant nodules are hypofunctioning, but more than 80% of benign nodules are also nonfunctioning. Conversely, functioning nodules (hot nodules), particularly if they are either more active than surrounding tissue or the sole functioning tissue, are rarely malignant.
Fluorescent Scan

Fluorescent scanning provides information concerning the content of stable iodine within the gland. In this technique, discrete zones of the thyroid gland are subjected to radiation from radioactive americium ($^{241}$Am) or from an x-ray tube. When incident radiation encounters $^{127}$I, a fluorescent x-ray, registered by a suitable detector, is emitted. Nonfunctioning nodules generally have a low iodine content and are therefore cold on fluorescent scan; thyroid iodine is depleted during subacute thyroiditis and increased during chronic iodine overload, such as with amiodarone treatment. The technique has limited clinical utility.
Ultrasonography

Sonography is noninvasive, is less expensive than computed tomography (CT) or magnetic resonance imaging (MRI), and produces no known tissue damage. No special preparation of the patient is necessary, and the technique requires only portable equipment, allowing it to be performed in the physician's examining room. A major limitation of ultrasonography is a high degree of observer dependence.

High-frequency sound waves are emitted by a transducer and reflected as they pass through the body, whereupon the returning echoes are received by the transducer, which also acts as a receiver. The amplitude of the reflections of the sound waves is influenced by differences in the acoustic impedance of the tissues encountered by the sound; for example, fluid-filled structures reflect few echoes and therefore have no or few internal echoes and well-defined margins; solid structures reflect varying amounts of sound and thus have varying degrees of internal echoes and less well-defined margins; and calcified structures reflect virtually all incoming sound and yield pronounced echoes with an acoustic "shadow" posteriorly.

High-frequency sound waves, such as those used in current thyroid sonography, are attenuated rapidly in the body tissues. Therefore, they cannot be used to image structures deeper than about 5 cm from the skin. Fortunately, the thyroid gland is usually well within this limit and can be completely imaged. 1

High-frequency (7 to 13 MHz), small-parts instruments have become widely available since the middle 1980s and provide good spatial resolution and image quality. 2 The theoretical axial resolution of these systems is about 1 mm; no other thyroid imaging method can achieve this degree of resolution. 3 Intrathyroidal nodules as small as 3 mm in diameter and cystic nodules as small as 2 mm can be readily detected. 4 Color flow Doppler ultrasonography allows visualization of very small vessels, so that vascularity of thyroid nodules can be assessed, but its diagnostic performance for malignancy is lower, as compared with fine-needle aspiration biopsy (FNAB).

Thyroid sonography is typically performed with the patient supine. The patient's neck is hyperextended by a pad centered under the scapulae to provide optimal exposure. The examiner usually sits at the head of the examining table and can steady the transducer by resting an elbow or a forearm on the table next to the patient's head. The thyroid gland must be examined thoroughly in transverse and longitudinal planes. Imaging of the lower poles can be enhanced by swallowing, which momentarily raises the thyroid gland in the neck. The examination should cover the entire gland, including the isthmus. Imaging should also include the region of the carotid artery and jugular vein to identify enlarged cervical lymph nodes. 5

The normal thyroid parenchyma has a characteristic homogeneous medium-level echogenicity, with little identifiable internal architecture (Fig. 13-1). The surrounding muscles have the appearance of hypoechoic structures. The air-filled trachea in the midline gives a characteristic curvilinear reflecting surface with an associated reverberation artifact. The esophagus is usually hidden from sonographic visualization by the tracheal air shadow. A portion of the esophagus, however, may swing laterally, usually toward the left, where it may lie adjacent to the posteromedial surface of the thyroid.

Neck ultrasonography may confirm the presence of a thyroid nodule when the findings on physical examination are equivocal. A diagrammatic representation of the neck showing the location or locations of any abnormal finding is a useful supplement to the routine film images recorded during an ultrasound examination. 6 Such a cervical map (Fig. 13-2) can help communicate the anatomic relationships of the pathology more clearly to the referring clinician and serves as a reference for the sonographer on follow-up examinations.

In patients with known thyroid cancer, sonography can be useful in evaluating the extent of disease, both preoperatively and postoperatively. In most instances, sonography is not performed routinely before thyroidectomy but can be useful in patients with large cervical masses for evaluation of nearby structures (e.g., the carotid artery and internal jugular vein) to exclude the possibility of direct invasion or encasement by the tumor.

Alternatively, in patients who present with cervical lymphadenopathy caused by papillary thyroid carcinoma (PTC) but in whom the gland is palpably normal, sonography may be used preoperatively to detect an occult, primary intrathyroidal focus. Some surgeons do regularly obtain a preoperative sonogram in patients with PTC or medulary thyroid carcinoma (MTC) in order to identify prior to surgery the anatomic locations of any sonographically suspicious regional lymph nodes and thereby to permit planning of the extent of nodal dissection. Occasionally, a hand-held ultrasound probe can be used intraoperatively to identify imparable residual cancer that has been identified by preoperative ultrasonography and proved to be cytologically positive by ultrasound-guided FNAB.

After surgery for thyroid cancer, sonography is the preferred method for detecting residual, recurrent, or metastatic disease in the neck. 7 In patients who have undergone less than a near-total thyroidectomy, the sonographic appearance of the remaining thyroid tissue may be an important factor in the decision whether to recommend completion thyroidectomy. Also, it is more sensitive than neck palpation in detecting recurrent disease within the thyroid bed and metastatic disease in cervical lymph nodes. 8,9,10 The location at the lower part of the neck and the sonographic appearance (hypoechoic, without a central echogenic line), the size (>1 cm in diameter), the shape (round), the presence of fine microcalcifications or a cystic component, and the use of color Doppler ultrasonography (hypervascularization) may aid in recognition of lymph node metastases. Sonography may also be useful to guide fine-needle biopsy of thyroid bed masses and lymph nodes, especially when these abnormalities are not palpable. 11,12,13
Computed Tomography

The CT appearance of the anatomic structures depends on the attenuation of the tissue examined. The thyroid gland, because of its high concentration of iodine, has higher attenuation than do the surrounding soft tissues.\[13\]

The diagnostic utility of CT in the evaluation of nodular thyroid disease is limited because thyroid masses, whether benign or malignant, may be hypodense, hyperdense, or isodense compared with adjacent normal thyroid tissue.\[14\] In aggressive pathologic processes, such as anaplastic thyroid carcinoma, CT can define the extension of the tumor to the mediastinum and its relationships to surrounding structures, such as the carotid artery, jugular vein, and trachea, before attempted surgical excision. In patients with known thyroid cancer, CT is less useful in evaluating recurrence in the neck because of the difficulty in detecting small masses in the indistinct tissue planes in the postoperative neck.\[15\] CT imaging can, however, improve the detection of lymph node metastases in the neck, although there is considerable overlap in the appearance of malignant and inflammatory nodes and CT lacks the ability to guide fine-needle biopsy of minimally enlarged nodes.\[7\] In patients with thyroid cancer, CT is used most frequently to search for lymph node metastases in the mediastinum and for distant metastases in the chest and abdomen.

CT scanning can provide useful information regarding the presence and extent of intrathoracic (subternal) goiters. The CT findings of an intrathoracic mass in continuity with the thyroid gland, with high attenuation on noncontrast-enhanced images and marked enhancement after intravenous contrast material injection, all suggest intrathoracic goiter.\[16\] Radionuclide scanning can also be performed in this clinical setting, but false-negative results can occur when little or no functional tissue is present in the intrathoracic goiter. Because of the necessity of infusing iodine-containing contrast agents, CT should be performed at least 4 weeks before any radiiodine therapy.\[16\]
Magnetic Resonance Imaging

Because the hydrogen atoms of different tissues have different relaxation times (termed T1 and T2), a computer-assisted analysis of T1-weighted and T2-weighted signals is used to differentiate the thyroid gland from skeletal muscles, blood vessels, or regional lymph nodes. Normal thyroid tissue tends to be slightly more intense than muscle on a T1-weighted image, and tumors often appear more intense than normal thyroid tissue.

MRI is rapidly evolving, with improvements in spatial resolution, reduction of artifacts, and development of new contrast agents. Currently obtained MR images have superior tissue contrast resolution but poorer spatial resolution than comparable CT images. Like CT, MRI does not distinguish benign from malignant nodules and does not assess functional status; however, it can define the anatomic extent of large goiters with great clarity. Coronal and sagittal images provide a simultaneous view of the cervical and thoracic components of substernal goiters. The relation of the goiter to surrounding vessels in the mediastinum is also well visualized.

Recurrent neoplasms in the thyroid bed or regional lymph nodes can be detected with MRI. MRI is more accurate than palpation and comparable in accuracy to CT. Recurrence is characterized by a mass with low to medium intensity on T1-weighted images and medium to high signal intensity on T2-weighted images. Conversely, scar tissue or fibrous tissue has low signal intensity on both T1-weighted and T2-weighted images. Tumor invasion of adjacent skeletal muscle has high signal intensity on T2-weighted images. Edema or inflammation in the muscle can cause a similar appearance and can be difficult to differentiate from recurrent tumor.
Positron Emission Tomography

Positron emission tomography (PET) is a special nuclear medical imaging technique that is both quantitative and tomographic. The radionuclide used emits a positron that is converted into a pair of photons after a short path of a few millimeters in the tissue. The coincidence detection of the two photons, which travel on a line in opposite directions, permits the localization of the site of the radionuclide decay.

The agent most widely used with PET is $^{18}$F-fluorodeoxyglucose ($^{18}$FDG). This agent is transported and phosphorylated as a glucose substitute but remains metabolically trapped inside tumor cells because of its inability to undergo glycolysis.

PET scanners with a large field of view permit in vivo images related to regional glucose metabolism, with high sensitivity and a spatial resolution less than 5 mm. Superimposition of CT and PET images greatly improves both the sensitivity and specificity of the technique and the anatomic localization of any focus of abnormal uptake. Elevated glucose metabolism is present in most malignant tumor tissues, and PET scanning has been shown to be particularly useful for the detection of lymph node metastases in the neck or mediastinum in patients with papillary and follicular thyroid carcinoma who have no tumoral radioiodine uptake. High uptake has also been observed in several thyroid diseases, such as thyroiditis, but PET cannot be used to differentiate benign from malignant thyroid nodules.
The clustering of goiters within families.

The persistence of goiters in areas where a widespread iodine prophylaxis program has been properly implemented.

Pathogenesis and Pathophysiology

Goiter has been traditionally regarded as the adaptive response of the thyroid follicular cell to any factor that impairs thyroid hormone synthesis. This classic concept no longer appears to encompass the many aspects of goiter. Indeed, goiter is characterized by a variety of clinical, functional, and morphologic presentations, and whether this heterogeneity represents different entities remains to be clarified. Also, iodine deficiency as the sole factor responsible for goiter appears to be an oversimplification. Thus, not all inhabitants in an iodine-deficient region develop goiter; moreover, endemic goiter has been observed in countries with no iodine deficiency, and even with iodine excess, and has not been observed in some regions with severe iodine deficiency.

The role of genetic factors is suggested by several lines of evidence, such as

1. The clustering of goiters within families.
2. The higher concordance rate for goiters in monozygotic than in dizygotic twins.
3. The female/male ratio (1.1 in endemic versus 7.1 to 9.1 in sporadic goiters).
4. The persistence of goiters in areas where a widespread iodine prophylaxis program has been properly implemented.

By studying families affected by goiter, researchers have been able to detect several gene abnormalities involving proteins related to thyroid hormone synthesis, such as mutations in thyroid globulin (Tg), sodium/iodide symporter (NIS), thyroid peroxidase (TPO), pendrin syndrome (PDS), and TSH receptor (TSHR) genes. In addition, two loci for this disorder have been identified. The first locus, identified on chromosome 14q, was designated MNG1 (Online Mendelian Inheritance in Man [OMIM] 138800) for multinodular goiter; the other, MNG2 (OMIM 300273), maps to chromosome Xp22. Although an autosomal dominant inheritance has been demonstrated in several families, multiple genes may be involved in other families. This may explain why predisposing gene alterations remain unidentified in most patients with simple goiter. Such genetic predispositions are believed to cause abnormalities in thyroid hormone synthesis. Thus, in some cases, defects can be detected by abnormalities of perchlorate discharge (see Chapter 10); more often, however, no abnormality can be demonstrated.

Goiter should thus be regarded as a complex trait in which both genetic susceptibility and environmental factors probably contribute to the development of disease. Whereas iodine deficiency represents the main environmental factor in the genesis of endemic goiter, other factors, such as cigarette smoking, infections, drugs, and goitrogenic factors, may play a role in the genesis of goitrous disease together with a genetic background of susceptibility. Interestingly, in a population-based twin study, a critical role of the genetic background in the etiology of goiter was demonstrated in females.

TSH has long been considered the major agent determining thyroid growth in response to any factor that impairs thyroid hormone synthesis. When such factors are operative, hypersecretion of TSH stimulates thyroid growth and increases the aspects of hormone biosynthesis that are capable of response. As a consequence of the increase in thyroid mass and functional activity, a normal rate of hormone secretion is restored and the patient is goitrous but eumetabolic. Indeed, in the rare clinical setting of functioning pituitary tumor, the increased blood TSH levels typically cause an enlargement of the thyroid gland.

It is interesting that goiter is also a typical part of the clinical picture of Graves’ disease, in which a stimulatory growth effect on thyroid tissue is induced by thyroid-stimulating antibody through TSHR activation. Moreover, thyroid enlargement may appear during the course of Graves’ disease when increased TSH levels result from overtreatment with antithyroid drugs. In addition, toxic thyroid hyperplasia is usually present in non-autoimmune autosomal dominant hyperthyroidism, a disorder related to germ lineactivating mutations of the TSHR gene. This clinical condition further emphasizes the role of TSH-TSHR system activation in the genesis of thyroid hyperplasia in diffuse nontoxic or toxic goiter.

This concept of the pathogenesis of nontoxic goiter is inconsistent with the fact that the serum TSH concentration is normal in most patients with nontoxic goiter. Nonetheless, a participatory role of TSH in the maintenance of goiter is indicated by the regression of goiter that sometimes follows administration of suppressive doses of thyroid hormone.

Several possible mechanisms may accommodate these apparently divergent findings. The mechanism with experimental support in rats is that iodine depletion enhances the promotion of thyroid growth by TSH. Hence, any factor that impairs intrathyroidal iodine levels may lead to gradual development of goiter in response to normal concentrations of TSH.

A second possibility is that the increase in serum TSH concentration is significant but too small to be detected by immunoassay methods.

Finally, a goitrogenic stimulus may have been present in the past but may no longer be detectable at the time of study. Thus, the residual normal TSH concentration can maintain but not initiate goiter. However, this primary, if not exclusive, role for TSH in determining thyroid growth and hyperplasia has been challenged.

Indeed, a complex network of both TSH-dependent and -independent pathways directs thyroid follicular cell growth and function and plays a role in the goitrogenic process. In particular, a variety of growth factors, derived either from the blood stream or through autocrine or paracrine secretion, may serve to regulate thyroid cell proliferation and differentiation processes. Among these factors, epidermal growth factor (EGF) and insulin-like growth factor (IGF) have been recognized as thyroid growthpromoting substances in different species.

IGF-I stimulates cell proliferation and differentiation (i.e., thyroglobulin expression) in thyroid tissue both in vitro and in vivo. Indeed, enhanced IGF-I expression may play a role in the goitrogenic process. In this regard, it is worth emphasizing that acromegalic patients with elevated levels of serum growth hormone and IGF-I and normal TSH levels have an increased prevalence of goiter. Similarly, fibroblast growth factor (FGF) stimulates thyroid function, and its expression has been associated with thyroid hyperplasia. Interestingly, the proliferation effect of these growth factors also occurs through stimulation of their respective receptors in normal TSH levels have an increased prevalence of goiter.

In the propylthiouracil-induced goiter in the rat, Wollman and colleagues recognized the importance of the development of new blood vessels in goiter formation and demonstrated that growth of perfusible blood vessels was induced by angiogenic factors produced by follicular cells. Indeed, many molecules involved in promoting or inhibiting thyroid angiogenesis, including vascular endothelial growth factor, angiopoietins 1 and 2, hepatocyte growth factor, endothelin, angiogenin or thrombospondin, angiostatin, and endostatin, have now been identified.

Goitrogenesis, therefore, appears to be a complex process in which TSH, growth factors, and angiogenic substances either play a distinct and separate role or act
synergistically through complex interaction mechanisms.

Another pathogenetic concept is based on autoradiographic and clinical studies of normal thyroid tissue and nontoxic and toxic multinodular goiters. Early in the course of goiter formation, areas of microheterogeneity of structure and function are intermixed and include areas of functional autonomy and small areas of focal hemorrhage. Indeed, as judged from the presence of scattered foci of persistent radioactive uptake in the thyroid glands of patients given suppressive doses of thyroid hormone before surgery, some cells with functional autonomy are present in the normal thyroid gland; this is in accordance with the heterogeneous staining for NIS observed in normal thyroid and goitrous tissues. Thus, in addition to the variability in thyroid microcirculation, heterogeneity may result from clonal differences among cells that give rise to thyroid follicles, some being more and some less responsive to external stimulation factors, including TSH, and others being autonomous from the outset. This concept implies that the anatomic and functional heterogeneity observed within the thyroid at the outset of the disease is exaggerated by prolonged stimulation.

Further insights into the pathogenesis of sporadic multinodular goiter have been gained by assessment of the clonality of individual thyroid nodules. Polyclonality implies a multicellular origin related to the proliferation of a group of cells, whereas a monoclonal tumor is thought to be formed by expansion of a single cell. Studies involving X-chromosome inactivation analysis have produced variable results in multinodular goiters. Some dominant nodules are monoclonal, especially if they showed evidence of recent rapid growth. Other researchers have found a monoclonal pattern in only a minority of large nodules.

Two groups reported that in multinodular glands more than one nodule can be monoclonal, and both monoclonal and polyclonal nodules can coexist within the same gland. Analysis of hyperplastic nodules by rigid criteria also indicated that morphologically indistinguishable hyperplastic thyroid nodules may be either monoclonal or polyclonal. Monoclonal adenomas within hyperplastic thyroid glands may reflect a stage in progression along the hyperplasia-neoplasia spectrum; accumulation of multiple somatic mutations may subsequently confer a selective growth advantage to this single-cell clone.

Cytogenetic and in situ hybridization studies also support the idea of a biologic continuum and karyotypic evolution between hyperplastic nodules and true follicular adenomas.

Eventually, the amount of functionally autonomous tissue in a multinodular goiter may be sufficient to suppress TSH secretion. Ultimately, autonomous hyperfunction may be sufficient to produce subclinical or overt thyrotoxicosis, or thyrotoxicosis may supervene when the patient is exposed to an iodine load. For this reason, patients with nontoxic multinodular goiter should not be given medications that contain iodine and should be observed after radiologic procedures that involve administration of iodinated contrast media. Some investigators administer antithyroid agents to patients with nodular goiter who are to receive agents containing iodine. This is a reasonable suggestion, especially in areas of iodine deficiency where iodbasedow is likely to occur.

Nontoxic goiter has a female preponderance (7:1 to 9:1) and seems to be common during adolescence or pregnancy. There appears to be no physiologic increase in thyroid volume during normal adolescence, and development of a goiter during adolescence is a pathologic rather than a physiologic process. However, as evidenced by sonographic measurement of thyroid volume in women living in an area of moderate iodine intake, normal pregnancy is goitrogenic, especially in women with preexisting thyroid disorders. The increased thyroid volume during pregnancy is associated with biochemical features of thyroid stimulation (i.e., an increased triiodothyronine/thyroxine [T3/T4] ratio) owing to slightly elevated serum TSH levels at delivery or a high human chorionic gonadotropin (hCG) concentration during the first trimester. Repeated pregnancies may play a role in the development of later thyroid disorders, a relation that might explain the high prevalence of thyroid disorders in women.
Pathology

Simple goiter is a noninflammatory, non-neoplastic, diffuse or nodular enlargement of the thyroid gland without hyperthyroidism. The gland is usually large and may have a distorted shape (Fig. 13-3). The cut surface shows areas of nodularity, fibrosis, hemorrhage, and calcification. The nodules vary in size, number, and appearance, the last according to their colloid or cellular content. Single or multiple cystic areas may contain colloid or brown fluid, representing previous hemorrhage.

Histologically, nodules contain irregularly enlarged, involuted follicles distended with colloid or clusters of smaller follicles lined by taller epithelium and containing small colloid droplets. These microfollicles may be surrounded by an edematous or a fibrous stroma. Large nodules tend to compress the surrounding parenchyma and may have a partially developed fibrous capsule. Markedly distended follicles may coalesce to form colloid cysts several millimeters in diameter.

The nodules tend to be incompletely encapsulated and are poorly demarcated from and merge with the internodular tissue, which also has an altered architecture. However, the nodules in some glands appear to be localized, with areas of apparently normal architecture elsewhere. Here, the distinction from a follicular adenoma may be difficult, and some pathologists apply terms such as colloid or adenomatous nodules to such lesions. Studies of clonality may be helpful in distinguishing between focal or nodular hyperplasia and true adenomas. Whereas nodular goiters are polyclonal in origin, solitary thyroid nodules are monoclonal and therefore true benign neoplasms.
Clinical Picture

Diffuse or nodular goiters are usually not associated with abnormal thyroid hormone secretion. Therefore, affected patients do not exhibit clinical symptoms or signs of thyroid dysfunction. The only clinical features of nontoxic goiter are those of thyroid enlargement. Nearly 70% of patients with sporadic nontoxic goiter complain of neck discomfort; the remainder have cosmetic concerns or a fear of possible malignancy.

Large goiter, which may displace or compress the trachea, esophagus, and neck vessels, can be associated with symptoms and signs including inspiratory stridor, dysphagia, and a choking sensation. These obstructive symptoms may be accentuated by the so-called Pemberton maneuver. This maneuver, which consists of "elevating both arms until they touch the sides of the head," is considered positive if, after a minute or so, congestion of the face, some cyanosis, and lastly distress become apparent. Compression of the recurrent laryngeal nerve, with hoarseness, suggests carcinoma rather than nontoxic goiter, but vocal cord paralysis can occasionally result from benign nodular goiters. Hemorrhage into a nodule or cyst produces acute, painful enlargement locally and may enhance or induce obstructive symptoms.

Endogenous subclinical thyrotoxicosis caused by autonomously functioning nodules should be carefully investigated. It is particularly relevant in elderly patients, whose cardiac morphology and function may be affected, thereby increasing their risk of developing cardiac arrhythmias.
Laboratory Tests

Serum TSH, measured in a highly sensitive immunometric assay, combined with a single measurement of free thyroid hormone concentrations may be used as a first-line screening test. Serum free thyroid hormones and TSH are, by definition, within the normal range. However, the T₃/T₄ ratio may be increased, perhaps reflecting defective iodination of Tg. Patients with sporadic nontoxic goiter tend to have high-normal free T₄ and T₃ concentrations and low TSH levels. The prevalence of so-called subclinical hyperthyroidism is higher when patients with nodular goiter have clear-cut autonomous areas on scintigraphy.

An undetectable serum TSH, even associated with normal free thyroid hormone levels, should suggest the possibility of toxic, autonomously functioning nodular areas in the goiter. Such a finding should prompt further cardiac investigation, especially in elderly patients, whose risk of atrial fibrillation may be increased as much as threefold when serum TSH levels are less than 0.1 mU/L. Moreover, it has been demonstrated that this condition, by affecting cardiac morphology and function, has a relevant clinical impact even in young patients and that many patients are, in fact, symptomatic. Therefore, this disorder should be considered a mild form of tissue thyrotoxicosis that may necessitate treatment.

In a cross-sectional study of 102 patients with sporadic nontoxic goiter, the serum TSH level correlated negatively with the thyroid volume, which in turn correlated positively with both the age of the patient and the duration of the goiter. In a prospective study of 242 patients with nodular goiter, no correlation was found between thyroid volume and any thyroid biochemical parameters, but there were significant negative correlations between the number of nodules identified by ultrasonography and the levels of basal TSH and the TSH response to thyrotropin-releasing hormone (TRH) stimulation.
Imaging in Goiter Evaluation

A diagnosis of goiter usually does not warrant the use of imaging procedures. When a nodular goiter is present, however, both scintigraphy and sonography provide useful information for disease management and treatment. Indeed, the former should be used to detect hot or warm nodules in the thyroid tissue. This finding affects the therapeutic approach. Sonography should be used to assess both morphology and size of the goiter. Thus, sonography in patients with a nodular goiter may allow a determination of the number and the individual features of the nodules and serve as guidance for FNAB. Sonography also permits an accurate, objective measure of goiter growth over time or after treatment.

Conventional radiography of the neck and the upper mediastinum should be used to determine the presence of tracheal compression. CT and MRI are indicated in the presence of intrathoracic goiter to define the relationships with surrounding structures.
Differential Diagnosis

The differential diagnosis of nontoxic goiter can be considered in functional and anatomic terms. As indicated, the same factors that lead to goitrous hypothyroidism can, if they are less severe, cause nontoxic goiter. Consequently, some patients with putative nontoxic goiter are slightly hypothyroid. On the other hand, foci of autonomous function may develop in multinodular goiters in which the spectrum of function can range from clinical euthyroidism with intact regulatory control to euthyroidism with some degree of functional autonomy to thyrotoxicosis. [76]

Anatomically, the diffuse stage of nontoxic goiter can resemble the thyroid of either Graves’ or Hashimoto’s disease. If Graves’ disease is not in an actively thyrotoxic phase, and if the ocular manifestations are lacking, there is no way to distinguish between the two except to demonstrate the presence of TSHR antibody in the serum. In one study of 108 patients with diffuse nontoxic goiter observed for more than 5 years, 33% had a family history of autoimmune thyroid disease and five patients developed Graves’ disease during follow-up. [77] Diffuse nontoxic goiter is sometimes also difficult to differentiate from Hashimoto’s disease, although the thyroid of Hashimoto’s disease is usually firmer and more irregular. Demonstration of high titers of antithyroid antibodies should indicate autoimmune disease.

In its multinodular stage, nontoxic goiter may suggest thyroid carcinoma. The approach to distinguishing between the two is discussed in the following section on thyroid neoplasms.
Treatment

Patients with small, asymptomatic goiters can be monitored by clinical examination and evaluated periodically with ultrasound measurements. In fact, goiter growth can be variable, and some patients have stable goiters for many years.

For more than a century, thyroid "feeding" has been employed to reduce the size of nontoxic goiters. The 1953 report of Greer and Astwood, in which two thirds of patients' goiters regressed with thyroid therapy, led to widespread acceptance of suppressive therapy despite some doubts about the value of such therapy. An overview of studies performed from 1960 to 1992 suggested that 60% or more of sporadic nontoxic goiters respond to suppressive therapy. In a prospective placebo-controlled, double-blind randomized clinical trial, 58% of the thyroxine-treated group had a significant response at 9 months, as measured by ultrasonography, in contrast with 5% after placebo. However, ultrasonographic measurement of goiter size demonstrated a return to pretherapy values within 3 months of treatment discontinuation.

Therefore, maintenance of the size reduction may require continuous long-term treatment. Nodular goiters appear to be less responsive than diffuse goiters, and the therapeutic efficacy of thyroxine treatment is increased in younger patients and in those with small or recently diagnosed goiters.

It has been proposed that a basal serum TSH greater than 1 mU/L in a patient with sporadic nontoxic goiter is an indication to administer levothyroxine to lower the serum TSH level to the low-normal range (0.5 to 1.0 mU/L). Others have suggested that TSH levels on treatment should be subnormal but not profoundly suppressed (0.1 to 0.3 mU/L). The validity of this approach remains to be ascertained. If the goiter size decreases or remains stable, treatment should be continued indefinitely, with periodic monitoring of serum TSH levels to detect possible development of functional autonomy.

A major concern in relation to long-term thyroxine suppression therapy is the possibility of detrimental effects on the skeleton and heart. It has been reported that TSH suppression therapy is associated with variable degrees of bone loss, particularly in postmenopausal women. However, other studies did not demonstrate significant change in bone mass after long-term thyroxine therapy. Furthermore, although marginal cardiac changes may occur with levothyroxine therapy, there is no evidence that levothyroxine per se is detrimental to the heart. It is now generally accepted that TSH should be suppressed with the lowest effective dose of levothyroxine, usually between 1.5 and 2.0 μg/kg body weight per day; the risk of deleterious effects may be minimized by monitoring serum TSH and free T<sub>4</sub> concentrations.

Surgery for simple nontoxic goiter is physiologically unsound because it further restricts the ability of the thyroid to meet hormone requirements. Nevertheless, surgery may become necessary because of persistence of obstructive manifestations despite a trial of levothyroxine. Surgery, which should consist of a near-total or total thyroidectomy, rapidly and effectively removes the goiter, but recurrence is seen in about 10% to 20% within 10 years. Surgical complications have been reported in 7% to 10% of cases and are more common with large goiters and with reoperation. Prophylactic treatment with levothyroxine after goiter resection probably does not prevent recurrence of goiter.

Traditionally, the role of <sup>131</sup>I therapy for nontoxic goiter was to reduce the size of a massive goiter in elderly patients who were poor candidates for surgery or to treat goiter that recurs after resection. However, several studies have demonstrated that primary treatment of multinodular goiter with <sup>131</sup>I is followed by a reduction in thyroid volume.

![Figure 13-4 Median changes in thyroid volume alterations after iodine 131 treatment in 39 patients with nontoxic multinodular goiter who remained euthyroid after a single dose. Bars represent quartiles. (From Nygaard B, Hegedus L, Gervil M, et al. Radioiodine treatment of multinodular nontoxic goiter. BMJ 1993; 307:828832.)](image)

A randomized trial comparing levothyroxine therapy with radioactive iodine treatment (120 μCi/g corrected for 24-hour thyroid uptake) showed impressive differences in outcome. After <sup>131</sup>I therapy, 97% of patients responded, with a mean decrease in goiter size of 39% at 1 year and 48% at 2 years; the initial side effects were neck tenderness and slight thyrotoxic symptoms in 12% of patients, and at 2 years 35% of patients were hyperthyroid and 10% had subclinical thyrotoxicosis. In contrast, with levothyroxine therapy 43% of patients responded with a mean decrease of 23% at 1 year and 22% at 2 years; the initial side effect was a mild thyrotoxicosis in 30% of patients and at 2 years a significant decrement in spine bone density.

It was formerly argued that treatment of large goiters or goiters with substernal extension with <sup>131</sup>I should be avoided because of the risks of acute swelling of the gland and consequent tracheal compression. Ultrasonographic studies of thyroid volume after <sup>131</sup>I have failed to demonstrate significant early volume increase. Moreover, decreased tracheal deviation and increased tracheal lumen size were demonstrable by MRI in patients who had compression by nontoxic goiters with substernal extension.

Therefore, it appears that <sup>131</sup>I treatment of nontoxic multinodular goiter is effective and safe, but hypothyroidism may occur in 22% to 40% within 5 years after <sup>131</sup>I therapy. Regular follow-up, preferably by a systematic annual recall scheme, is necessary. Although reassuring data are available on the long-term thyroid and nonthyroidal cancer risk after <sup>131</sup>I treatment in hyperthyroidism, the follow-up of patients with <sup>131</sup>I-treated nontoxic goiters is short-term and involves small numbers of patients. Children and adolescents should not be treated with <sup>131</sup>I. Stimulation with low doses of recombinant human TSH (rTSH) (0.01 to 0.03 mg) increases the thyroid <sup>131</sup>I uptake and therefore may allow the administration of a lower dosage of <sup>131</sup>I. Long-term randomized studies comparing the effects, side effects, and costs and benefits of surgery and <sup>131</sup>I treatment need to be performed.
THYROID NEOPLASIA

In an era when patients are advised on self-examination to detect cancer at an early stage, the finding of a palpable mass in such a superficial location as the thyroid gland can be disconcerting. The affected patient is likely to seek medical evaluation. At the end of an appropriate investigation, the clinician can usually reassure the patient that the nodule is benign. Alternatively, if the evaluation does suggest malignancy, the patient can be advised that the management of typical thyroid cancer is effective and usually consists of surgical resection, followed by medical therapy and regular surveillance. The major challenge in this circumstance is to determine whether the discovered thyroid nodule is malignant.

Some degree of consensus has been achieved with regard to both the initial evaluation of nodular thyroid disease and the management of differentiated thyroid cancer. But important clinical and biologic questions remain unanswered. In the following discussion, we describe a clinical approach to nodular thyroid disease and present a widely used scheme for classifying and staging tumors of the thyroid gland. We also review the features of the principal types of benign and malignant thyroid neoplasms and the controversies in the management of differentiated thyroid carcinoma.

Initial Investigation

Thyroid tumors are the most common endocrine neoplasms. They usually arise as anterior neck nodules that usually can be localized to the thyroid gland by palpation. Most of these nodules are benign hyperplastic (or colloid) nodules or benign follicular adenomas, but about 5% to 10% of nodules coming to medical attention are carcinomas. Differentiating true neoplasms from hyperplastic nodules and distinguishing between benign and malignant tumors are major challenges.

Moreover, with the widespread practice of medical checkups in healthy individuals and the increasing use of imaging technology, this problem is likely to become more common. High-resolution ultrasound studies suggest that the prevalence of nodular thyroid disease in healthy adults is above 60%. However, during 2001 in the United States, only about 19,500 new cases of thyroid cancer were likely to be diagnosed. Therefore, most of these so-called thyroid incidentalomas are obviously benign and do not progress to clinical tumors.

In identifying the nodules that are likely to be malignant, a thorough history and a careful physical examination should be supplemented with laboratory testing, imaging procedures, and, most important, FNAB of the nodule in question. With the use of this approach, it is possible to assess the likelihood of malignancy and to advise appropriate treatment in the majority of patients.

History and Physical Examination

Historical features that favor benign disease include the following:

1. A family history of Hashimoto's thyroiditis, benign thyroid nodule, or goiter.
2. Symptoms of hypothyroidism or hyperthyroidism.
3. A sudden increase in size of the nodule with pain or tenderness, suggesting a cyst or localized subacutethyroiditis.

Historical features that suggest malignancy include the following:

1. Young (<20 years old) or old (>70 years old) age;
2. Male sex;
3. A history of external neck radiation during childhood or adolescence;
4. Recent changes in speaking, breathing, or swallowing;
5. A family history of thyroid cancer or multiple endocrine neoplasia (MEN) type 2.

On physical examination, manifestations of thyroid malignancy should be sought, including firm consistency of the nodule, irregular shape, fixation to underlying or overlying tissues, and suspicious regional lymphadenopathy.

In both prospective and retrospective studies, the sensitivity and specificity rates for detecting thyroid malignancy by history and physical examination were about 60% and 80%, respectively. In these historical series, only about 20% of patients with later confirmed malignancy had, when initially seen, neither suspicious historical features nor evidence of potential malignancy on neck examination. Further testing may include assessment of thyroid function, measurement of tumor markers, genetic screening, thyroid imaging, and the only decisive parameter, FNAB.

Laboratory Evaluation

The serum TSH level is measured to exclude thyroid dysfunction. Patients with thyroid cancer rarely have abnormalities in serum TSH levels. A low (suppressed) serum TSH level may indicate a toxic nodule and should lead to thyroid scintigraphy. Measurement of serum anti-TPO antibody and anti-Tg antibody levels may be helpful in diagnosis of chronic autoimmune thyroiditis, especially if the serum TSH level is elevated. In chronic autoimmune thyroiditis, the thyroid gland's size and consistency may simulate either a solitary nodule or bilateral nodules. Evidence of autoimmune thyroiditis, however, does not preclude the presence of cancer in the gland.

Follicular cell-derived thyroid cancers (FCTCs) may release increased amounts of Tg into the blood stream. Unfortunately, there is overlap of serum Tg levels in FCTCs and in a number of benign conditions, and measurement of serum Tg levels is not useful in the initial work-up of nodular thyroid disease. Similarly, some investigators routinely measure calcitonin (CI) levels in all patients with nodular thyroid disease to identify cases of MTC. In fact, the calcitonin level is increased in virtually all patients with clinical MTC. However, because of the rarity of unsuspected MTC, the high frequency of false-positive results that may prompt a thyroidectomy despite a reassuring cytologic result, and the unknown clinical relevance of medullary microcarcinomas, it is neither cost-effective nor necessary to determine calcitonin levels in patients with nodular thyroid disease in the absence of clinical suspicion of MTC or abnormal cytologic findings.

The molecular abnormality in more than 95% of familial MTC cases is a germline mutation of the RET proto-oncogene that is located on the long arm of chromosome 10. Many investigators advocate RET mutation testing in all patients with MTC, including apparently sporadic cases, because 4% to 6% of such patients have germline mutations of the gene. Such tests are highly accurate, reproducible, and reliable. If a mutation is found, family members at risk are then tested to identify affected individuals. A negative result obviates the need for any further testing, and individuals who harbor such mutations should undergo prophylactic total thyroidectomy to prevent later development of the multicentric MTC that occurs in this disorder.

Thyroid Imaging

The traditional imaging procedure is thyroid scintigraphy using radiiodine, technetium-99m Tc. Most thyroid carcinomas are inefficient in trapping and organifying iodine and appear on scans as areas of diminished isotope uptake, so-called cool or cold nodules. Unfortunately, most benign nodules also do not concentrate iodine and therefore are cold nodules. Furthermore, not all nodules with normal or slightly increased uptake are benign and may appear cold on a thyroid scan with radioactive iodine.

The only situation in which an iodine scan can exclude malignancy with reasonable certainty is in the case of a toxic adenoma, which is characterized by significantly
increased uptake within the nodule and markedly suppressed or absent uptake in the remainder of the gland. These lesions account for fewer than 10% of thyroid nodules and are almost invariably benign.\(^{138}\) When isotopic thyroid scanning is compared with history and physical examination, most authors have found scanning to be of negligible or no value for the diagnosis of malignancy.\(^{139,140}\)

In an attempt to improve the performance of isotopic scanning, a number of radioisotopes other than iodine-related compounds have been tried, such as thallium 201 (\(^{201}\)Tl)\(^{12,13}\) and \(^{99m}\)Tc-labeled methoxysobutyl isonitrile (MIBI). In the hands of dedicated experts, these techniques may be valuable, but they are expensive and their widespread use must await more extensive evaluation.

Ultrasonography is capable of detecting even minute thyroid nodules and increases the sensitivity of carcinoma detection but does little to enhance specificity. In fact, of 1000 normal control subjects, 65% had detectable nodularity on high-resolution scanning.\(^{141}\) Attempts have been made to develop criteria for distinguishing benign and malignant nodules. Echo-free (cystic) and homogeneously hyperechoic lesions are reputed to carry a low risk of malignancy.\(^{12,142}\) Positive predictive criteria of malignancy include solid hypoechoic nodules, presence of calcifications, irregular shape, absence of halo, and absence of cystic elements; however, in one study only 64% of malignant nodules displayed patterns typical of malignancy.\(^{143}\) In addition, nodules that can be clearly identified as benign by sonography are uncommon, limiting the usefulness of ultrasound scanning.

Ultrasonography is useful in identifying hypoechoic nodules that should be submitted to FNAB and also in examining the rest of the thyroid gland and lymph node areas.\(^{144}\) It may also be used in case of nonpalpable nodules to guide FNAB, especially when the diameter of the nodule is 1 cm or more.\(^{12,145}\) Cystic lesions may be recognized by aspiration of the fluid and ethanol injection to avoid recurrence; this is optimally performed under ultrasonographic guidance.\(^{146}\)

CT scanning and MRI in the initial diagnosis of thyroid malignancy do not provide higher-quality images of the thyroid and cervical nodes than those of ultrasonography. CT examination of the lower central neck is preferable when tracheal or mediastinal invasion is suspected.

**TABLE 13-2** — Probability of Malignancy at Histology Based on Fine-Needle Aspiration Biopsy Cytology (Summary of the Literature)

<table>
<thead>
<tr>
<th>Cytology</th>
<th>Percent of Results (%), Mean (Range)</th>
<th>Probability of Malignancy (%), Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inadequate or nondiagnostic</td>
<td>16 (1520)</td>
<td>1020</td>
</tr>
<tr>
<td>Benign</td>
<td>70 (5390)</td>
<td>12</td>
</tr>
<tr>
<td>Suspicious</td>
<td>10 (523)</td>
<td>1020</td>
</tr>
<tr>
<td>Malignant</td>
<td>4 (110)</td>
<td>&gt;95</td>
</tr>
</tbody>
</table>

*The suspicious category includes follicular neoplasms (hyperplastic nodules, follicular adenomas, and follicular carcinomas) and some Hurthle cell tumors.

FNAB of thyroid nodules has eclipsed all other techniques for diagnosing thyroid cancer, with reported overall rates of sensitivity and specificity exceeding 90% in iodine-sufficient areas.\(^{12,146}\) The technique is easy to perform and safe, with only a handful of complications having been reported in the literature.\(^{12,147}\) and causes little discomfort. However, care must be taken to obtain an adequate specimen; most authors recommend between three and six aspirations.\(^{12,148}\) A satisfactory specimen contains at least five or six of the 10 to 15 well-preserved cells. The cells are categorized by their cytologic appearances into benign, indeterminate or suspicious, and malignant (Table 13-2).

The diagnosis of PTC by FNAB on the basis of characteristic nuclear changes is particularly reliable and accurate, with sensitivity and specificity both approaching 100%. For follicular neoplasms, however, the performance of FNAB is inferior. If strict criteria for malignancy are used, sensitivity may be as low as 8%.\(^{12,149}\) If any follicular neoplasm that is not clearly benign on cytologic examination is classified as cancerous, sensitivity rises to about 90% or more. Unfortunately, this increase is associated with a considerable drop in specificity to less than 50% (i.e., a large number of false-positive results). This seriously limits the usefulness of FNAB in iodine-deficient regions, where the incidence of thyroid follicular carcinoma (FTC) approaches that of PTC and where both follicular adenomas and hyperplastic adenomatous nodules are prevalent.

TPO immunochemistry with a monoclonal antibody (MoAb 47) shows promise in improving the accuracy of FNAB for follicular lesions.\(^{12,149}\) For 100% sensitivity, a specificity of almost 70% has been achieved with this technique. Pending independent confirmation of these results, TPO immunocytochemistry may be a valuable adjunct to the standard cytologic techniques. The use of large-needle biopsy in addition to standard FNAB has improved diagnostic accuracy in difficult FNA cases.\(^{12,150}\) but the technique is more exacting than FNAB alone and is associated with increased morbidity and, possibly, increased complication rates.

Particularly for cystic thyroid nodules, sampling from the margin of the nodule, rather than from the cystic fluid and debris in the center, increases accuracy.\(^{12,151}\) Ultrasonographically guided FNA can be used for this purpose (Fig. 13-5). Although such guided biopsies are sometimes helpful, routine use of ultrasound-guided biopsy for clinically palpable nodules is not any better than “freehand” aspiration.\(^{12,152}\) However, some centers are evaluating this approach to allow recognition of FNA of nonpalpable nodules 1 cm or smaller in size and to reduce the number of passes to three.\(^{12,153}\)

In some European centers, both preoperative FNAB and intraoperative frozen section are combined in endemic goitrous regions with high rates of follicular tumors.\(^{12,154}\) In the hands of experienced surgeon-pathologist teams, this approach results in less than 5% misdiagnoses, as evidenced by subsequent review of paraffin-embedded specimens. The approach avoids unnecessarily extensive surgery in patients with benign tumors, achieves resection of nearly all malignant tumors, and rarely necessitates a second operation for completion thyroidectomy.\(^{12,155}\) Such an approach is employed at the Mayo Clinic and at the Institut Gustave Roussay, where intraoperative frozen section is routine.

Apart from its limited utility in the evaluation of follicular neoplasms, the only other limitation of FNAB is nondiagnostic specimens, which may be obtained in up to 20% of cases.\(^{12,156}\) Although repeated aspiration increases both the accuracy and the rate of diagnostic aspirations, even repeated attempts may sometimes fail. Many persistently nondiagnostic FNAB specimens may be neoplastic, possibly 50%.\(^{12,154}\) Hence, either close observation or surgical removal of the nodule is probably the best option. Some authorities recommend a trial of TSH suppression, which can sometimes shrink benign nodules.\(^{12,153}\) However, a significant proportion of benign nodules do not shrink, and some carcinomas do shrink; consequently, the diagnostic value of TSH suppression is doubtful. Whether ultrasound-guided FNAB can help overcome this problem is unclear, but confirmation is required.\(^{12,157}\) Figure 13-6 is an algorithm for the management of nodular thyroid disease in which FNAB is the first diagnostic test and subsequent management is based on cytologic results.
The most expeditious way to diagnose thyroid malignancy is to obtain a thorough history and physical examination, followed by FNAB and evaluation of the sample by an experienced cytologist. In some cases, FNAB should be performed under ultrasound guidance. Imaging procedures, in addition to ultrasonography, and other tests may occasionally be helpful, but diagnostic thyroid scintiscanning, as traditionally practiced, is of little or no value and should be abandoned.

In iodine-sufficient areas with a high relative prevalence of PTC, the combination of history and physical examination and FNAB is usually sufficient to confirm malignancy. Conversely, if history and physical examination, FNAB, and ultrasonography do not suggest malignancy, the chances of missing PTC are probably less than 1%. In areas where the prevalence of follicular tumors is higher, more patients may require neck exploration because FNAB may not be conclusive; in experienced hands, however, intraoperative frozen sections can limit the number of unnecessarily extensive, bilateral procedures. Surgery should also be considered for large tumors (>4 cm), especially in young subjects, in order to avoid repeated evaluations; in addition, because these tumors may be composed of various cell populations, results of FNAB may be less reliable. Finally, micronodules less than 1 cm in diameter, found incidentally during imaging, do not need to be tested any further, unless there are sonographic features suggestive of PTC or MTC. The usual advice is to repeat ultrasonography of such lesions after an interval of 6 to 12 months.
Classification of Thyroid Tumors

Histologic Classification

Two monographs have had a major impact on the histologic classification of thyroid tumors. One is from the World Health Organization (WHO),[147] the other, from the Armed Forces Institute of Pathology (AFIP).[148] The classification described in Table 13-3 is modified from the guidelines described by these organizations.[147][148]

Lesions of follicular cell origin constitute more than 95% of the cases, and the remainder are largely made up of tumors exhibiting C cell differentiation.[149] Mixed medullary and follicular carcinomas, made up of cells with both C-cell and follicular differentiation, are rare and of uncertain histogenesis.[147] Nonepithelial thyroid tumors mainly include malignant lymphomas, which may involve the thyroid gland as the only manifestation of the disease or as part of a systemic disease. True sarcomas and malignant hemangioendotheliomas are exceptional. Blood-borne metastases to the thyroid are not uncommon at autopsy in patients with widespread malignancy but rarely cause clinically detectable thyroid enlargement.

Staging of Thyroid Carcinoma

In addition to the histologic classification of thyroid tumors developed by WHO and AFIP groups, the International Union Against Cancer (UICC) and the American Joint Committee on Cancer have agreed on a staging system in thyroid cancer.[150]

As stated by the AJCC, “the principal purpose served by international agreement on the classification of cancer cases by extent of disease was to provide a method of conveying clinical experience to others without ambiguity.”[150]

The AJCC based its system of classification on the TNM system, which relies on assessing three components: (1) extent of the primary tumor (T), (2) absence or presence of regional lymph node metastases (N), and (3) absence or presence of distant metastases (M).

The TNM system allows a reasonably precise description and recording of the anatomic extent of disease. The classification may be either clinical (cTNM), based on evidence (including biopsy) acquired before treatment, or pathologic (pTNM), by which intraoperative and surgical pathology data are available. Obviously, pTNM classification is preferable because a precise size can be assigned to the primary tumor, the histo-type is identified, and extrathyroid invasion is demonstrated unequivocally.

Typically, the primary thyroid tumor (T) status is defined according to the size of the primary lesion:

T1, greatest diameter 1 cm or smaller

TABLE 13-3 – Classification of Thyroid Neoplasms

<table>
<thead>
<tr>
<th>Primary Epithelial Tumors</th>
<th>Papillary or Follicular</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tumors of Follicular Cells</td>
<td></td>
</tr>
<tr>
<td>Benign: follicular adenoma</td>
<td></td>
</tr>
<tr>
<td>Malignant: carcinoma</td>
<td></td>
</tr>
<tr>
<td>Differentiated</td>
<td></td>
</tr>
<tr>
<td>Papillary</td>
<td></td>
</tr>
<tr>
<td>Follicular</td>
<td></td>
</tr>
<tr>
<td>Poorly differentiated</td>
<td></td>
</tr>
<tr>
<td>Insular</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td></td>
</tr>
<tr>
<td>Undifferentiated (anaplastic)</td>
<td></td>
</tr>
<tr>
<td>Tumors of C Cells</td>
<td></td>
</tr>
<tr>
<td>Medullary carcinoma</td>
<td></td>
</tr>
<tr>
<td>Tumors of Follicular and C Cells</td>
<td></td>
</tr>
<tr>
<td>Mixed medullary-follicular carcinoma</td>
<td></td>
</tr>
<tr>
<td>Primary Nonepithelial Tumors</td>
<td></td>
</tr>
<tr>
<td>Malignant Lymphomas</td>
<td></td>
</tr>
<tr>
<td>Sarcomas</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td></td>
</tr>
<tr>
<td>Secondary Tumors</td>
<td></td>
</tr>
</tbody>
</table>

TABLE 13-4 – American Joint Committee on Cancer Stage Groupings for Thyroid Carcinoma:

<table>
<thead>
<tr>
<th>Stage</th>
<th>Papillary or Follicular</th>
<th>Medullary (Any Age)</th>
<th>Anaplastic (Any Age)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>M0 (Age &lt;45 yr)</td>
<td>T1</td>
<td>T1</td>
</tr>
<tr>
<td>II</td>
<td>M1 (Age 45 yr)</td>
<td>T2T3</td>
<td>T2T4</td>
</tr>
<tr>
<td>III</td>
<td>T4 or N1</td>
<td>N1</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>M1</td>
<td>M1</td>
<td>Any</td>
</tr>
</tbody>
</table>

* T, size of primary thyroid tumor (T1, 1 cm; T2, >1 cm; T3, >4 cm; T4, extrathyroid invasion); N, regional nodal metastases (0, absent; 1, present); M, distant metastases (0, absent; 1, present).
A thyroid tumor with four degrees of T, two degrees of N, and two degrees of M can have 16 different TNM categories.

For purposes of tabulation and analysis, these categories have been condensed into a convenient number of TNM stage-groupings (Table 13-4). Whereas head and neck cancer is usually staged entirely on the basis of anatomic extent, in thyroid cancer staging both the histologic diagnosis and the age of the patient for PTC and FTC are included because of their importance in predicting the behavior and prognosis of thyroid cancer.

According to this staging scheme, all patients younger than age 45 years with PTC or FTC are in stage I, unless they have distant metastases (DM), in which case they would be in stage II. In young patients and especially in children, the risk of recurrence is high and may be underestimated by the TNM staging system. Older patients (aged 45 years or more) with node-negative papillary or follicular microcarcinoma (T1 N0 M0) are in stage I. Tumors between 1.1 and 4.0 cm are classified as stage II, and those with either nodal spread (N1) or extrathyroidal invasion (T4), stage III.

For MTC, the scheme is similar, in that microcarcinoma is stage I and a node-positive tumor is stage III. There is no age distinction for MTC, although age is a significant independent prognostic indicator in most multivariate analyses, and local (extrathyroidal) invasion is grouped within stage II. For patients with MTC and older patients with PTC or FTC, stage IV denotes the presence of DMs. Independent of age or tumor extent, all patients with undifferentiated (anaplastic) cancer are considered to be in stage IV.
Follicular Adenoma

Follicular adenoma is a benign, encapsulated tumor with evidence of follicular cell differentiation. It is the most common thyroid neoplasm and may be found in 4% to 20% of glands examined at autopsy. The tumor has a well-defined fibrous capsule that is grossly and microscopically complete. There is a sharp demarcation and distinct structural difference from the surrounding parenchyma. These adenomas vary in size, but most have a diameter of 1 to 3 cm at the time of excision. Degenerative changes, including necrosis, hemorrhage, edema, fibrosis, or calcification, are common features, particularly in larger tumors.

Follicular adenomas can be classified into subtypes (Table 13-5) according to the size or presence of follicles and degree of cellularity. Each adenoma tends to have a consistent architectural pattern. Microfollicular, normofollicular, and macrofollicular adenomas owe their names to the size of their follicles compared with follicles in the neighboring, non-neoplastic areas of the gland. Trabecular adenomas are cellular and consist of columns of cells arranged in compact cords. They show little follicle formation and rarely contain colloid. A variant, the hyalinizing trabecular adenoma, has unusually elongated cells and prominent hyaline changes in the extracellular space.

The histologic differences between these subtypes are striking but of no clinical importance. The only practical value of the classification is that the more cellular a follicular nodule is, the more one should search for evidence of malignancy in the form of invasion of blood vessels and capsule, either singly or in combination. Atypical adenomas are hypercellular or heterogeneous, or both, with gross and histologic appearances.

### TABLE 13-5 – Subtypes of Follicular Adenoma

<table>
<thead>
<tr>
<th>Conventional</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trabecular/solid (embryonal) adenoma</td>
</tr>
<tr>
<td>Microfollicular (fetal) adenoma</td>
</tr>
<tr>
<td>Normofollicular (simple) adenoma</td>
</tr>
<tr>
<td>Macrofollicular (colloid) adenoma</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyalinizing trabecular adenoma</td>
</tr>
<tr>
<td>Oncocytic (oxyphilic or Hürthle cell) tumor</td>
</tr>
<tr>
<td>Adenomas with papillary hyperplasia</td>
</tr>
<tr>
<td>Hyperfunctioning (&quot;toxic&quot;) adenoma</td>
</tr>
<tr>
<td>Atypical (hypercellular) adenoma</td>
</tr>
</tbody>
</table>

Figure 13-7 Genetic events in thyroid tumorigenesis. Activating point mutations of the RAS genes are found with a similar frequency in follicular adenomas and follicular carcinomas and are considered an early event in follicular tumorigenesis. The PPAR-PAX8 rearrangement was found only in follicular carcinomas. Rearrangements of transmembrane receptors with tyrosine kinase activity (RET, TRK genes) are found only in papillary thyroid carcinomas. Inactivating point mutations of the P53 gene are found only in poorly differentiated and anaplastic thyroid carcinomas. Activation of the cyclic adenosine monophosphate pathway, by point mutation of the thyrotropin receptor (TSH-R) or the subunit of the G protein genes, leads to the appearance of hyperfunctioning thyroid nodules. Gs stimulatory guanylyl nucleotide protein.

that suggest the possibility of malignancy but not invasion. They account for fewer than 3% of all follicular adenomas. Follow-up indicates that this lesion behaves in a benign fashion. The fact that the tumor does not recur or produce metastases after removal does not prove that it is actually benign; removal may have interrupted a natural history that would have culminated in invasion and metastases.

The most important cytologic variant is the oxyphilic or oncocytic (Hürthle cell) adenoma, which is composed predominantly (at least 75%) or entirely of large cells with granular, eosinophilic cytoplasm. Ultrastructurally, the cells are rich in mitochondria and may exhibit nuclear pleomorphism with distinct nucleoli. Although all such neoplasms are thought by some to be potentially malignant, the biologic behavior and clinical course of oncocytic tumors correlate closely with the histology and the size of the initial lesion. The absence of invasion predicts a benign outcome, but larger tumors may rarely be associated with later recurrence or metastases, even in the absence of obvious microscopic evidence of invasion; fortunately, such an occurrence is extremely rare, and generally a diagnosis of benign Hürthle cell adenoma can be reliable. Some normofollicular adenomas may contain pseudopapillary structures that can be confused with the papillae of papillary carcinoma. These structures are probably an expression of localized hyperactivity and are most common in adenomas that show autonomous function.

In the majority of hyperfunctioning follicular adenomas, activating point mutations have been identified in the TSHR or in the subunit of the stimulatory guanyl nucleotide protein (Gαs). Such mutations may impair guanosine triphosphatase (GTP) activity, trapping the G protein in a state of constitutive activation, resulting in enhanced cyclic adenosine monophosphate (cAMP) production and constitutive hyperstimulation of the cells.
Papillary Thyroid Carcinoma

PTC has been defined as "a malignant epithelial tumor showing evidence of follicular cell differentiation, and characterized by the formation of papillae and/or a set of distinctive nuclear changes." The most common thyroid malignancy, PTC constitutes 50% to 90% of differentiated FCTCs worldwide.

Papillary thyroid microcarcinoma (PTM) is defined by the WHO as a PTC 1 cm in diameter or smaller. The incidence rates for clinically diagnosed PTC in the United States are approximately 5 per 100,000 for tumors larger than 1 cm in diameter and 1 per 100,000 for PTM. By contrast, the incidence of PTM in autopsy material from various continents ranges from 4% to 36.

Typically, PTC shows a predominance of papillary structures, consisting of a fibrovascular core lined by a single layer of epithelial cells, but the papillae are usually admixed with neoplastic follicles having characteristic nuclear features.

The nuclei of PTC cells have a distinctive appearance that has a diagnostic significance comparable to that of the papillae. Indeed, the preoperative diagnosis of PTC can often be made on the basis of the characteristic nuclear changes seen in FNA material. Nuclei are larger than in normal follicular cells and overlap, they may be fissured like coffee beans, chromatin is hypodense (ground glass nuclei), limits are irregular, and they frequently contain an inclusion corresponding to a cytoplasmic invagination.

Several subtypes exist:

1. The tumor is designated a follicular variant of PTC when the lining cells of the neoplastic follicles have the same nuclear features as seen in typical PTC and the follicular predominance over the papillae is complete.
2. The diffuse sclerosing variant is characterized by diffuse involvement of one or both thyroid lobes, widespread lymphatic permeation, prominent fibrosis, and lymphoid infiltration.
3. The tall cell variant is characterized by well-formed papillae that are covered by cells twice as tall as they are wide.
4. The columnar cell variant differs from other forms of PTC because of the presence of prominent nuclear stratification.

Molecular Pathogenesis

The thyroid follicular cell may give rise to both benign and malignant tumors, and the malignancy can be of either papillary or follicular histotype. There is no evidence that benign tumors ever undergo malignant transformation into classic PTC. Structural abnormalities of the chromosomes may occur in about 50% of PTCs, frequently involving the long arm of chromosome 10. The RET proto-oncogene is located on chromosome 10q11-2. It encodes a transmembrane receptor with a tyrosine kinase domain. Its ligands are the glial cell linederived neutrophilic factor (GDNF) and the neurturin, both of which induce protein dimerization. RET activation was first demonstrated in transfusion experiments and has been found only in PTC tumors. It was therefore called RET/PTC.

All activated forms of the RET proto-oncogene are the consequence of oncogenic rearrangements fusing the tyrosine kinase domain of the RET gene with the 5' domain of different genes. The foreign gene is constitutively expressed, and its 5' domain acts as a promoter, resulting in permanent expression of the RET gene. Furthermore, these genes have domains that induce RET activation by permanent dimerization; because of this fusion, the chimeric protein is localized in the cytoplasm and not in the plasma cell membrane.

Three major classes of RET/PTC have been identified:

1. RET/PTC, is formed by an intrachromosomal rearrangement fusing the RET tyrosine kinase domain to a gene designated H4 (D10S170), whose function is still unknown.
2. RET/PTC is formed by an interchromosomal rearrangement fusing the RET tyrosine kinase domain to a gene located on chromosome 17 encoding the RI regulatory subunit of protein kinase A.
3. RET/PTC is formed by an intrachromosomal rearrangement fusing the RET tyrosine kinase domain to a gene designated ELE1, whose function is still unknown.

Several variants of RET/PTC have been observed in post-Chernobyl thyroid tumors, including rearrangements formed by fusing the tyrosine kinase domain of the RET gene at other breakpoint sites or with other partners.

The frequency of RET/PTC rearrangements occurring in PTC patients without prior childhood neck irradiation varies between 2.5% and 35%. In these tumors, the frequencies of RET/PTC and RET/PTC were similar and that of RET/PTC was lower. The RET/PTC rearrangements were more frequently found (in 60% to 80% of cases) in PTC cases occurring either after external irradiation in childhood or in children after the Chernobyl accident. RET/PTC was more frequently found in aggressive tumors that occurred early after permanent dimerization; because of this fusion, the chimeric protein is localized in the cytoplasm and not in the plasma cell membrane.

Several additional oncogenes may occasionally be involved in PTC, including NTRK1 (also named TRKA), which codes for a neural growth factor receptor with a tyrosine kinase domain and which is activated by rearrangement in about 10% of PTCs. The receptor for hepatocyte growth factor is a transmembrane tyrosine kinase encoded by the MET oncogene; it is overexpressed in some patients with PTC, and low expression has been associated with the occurrence of DM.

A high incidence of PTC has been reported in patients with adenomatous polyposis coli and Cowden's disease (the multiple hamartoma syndrome), suggesting that the predisposing genes may play a role in the occurrence of papillary carcinoma. About 3% of cases of PTC are familial; their behavior is similar to or slightly more aggressive than that of nonfamilial cases. The gene predisposing to familial thyroid tumors with cellular oxyphilia has been mapped to chromosome 19q13.2, and in a family with PTC and renal carcinoma a separate gene was mapped to chromosome 1q21.1-2.
Although PTCs can occur at any age, most occur in patients between 30 and 50 years of age (mean age, 45 years). Women are affected more frequently (female predominance, 60% to 80%). Most primary tumors are 1 to 4 cm in size; they average about 2 to 3 cm in greatest diameter. Extrathyroidal invasion of adjacent soft tissues is present in about 15% (range 5% to 34%) at primary surgery, and about one third of PTC patients have clinically evident lymphadenopathy at presentation. About 35% to 50% of excised neck nodes have histologic evidence of involvement, and in patients 17 years of age or younger nodal involvement may be present in up to 90%. Only 1% to 7% of PTC patients have DM at diagnosis. Spread to superior mediastinal nodes is usually associated with extensive neck nodal involvement.

The TNM classification is a widely used system for tumor staging. Most PTC patients present with either stage I (60%) or stage II (22%). Patients aged 45 years or older with either nodal metastases or extrathyroidal extension (stage III) account for fewer than 20% of cases. As already noted, few (1% to 7%) of PTC patients present with DM and have stage IV disease (age 45 years or older with any T, any N, M1). Figure 13-8 (upper left) illustrates the distribution of TNM stages in 2284 PTC cases seen at the Mayo Clinic, and Figure 13-9 demonstrates survival by TNM stage in this cohort of PTC patients treated from 1940 to 1997.

### Recurrence and Mortality

Three types of tumor recurrence may occur with PTC:

- **Postoperative nodal metastases (NM)**
- **Local recurrence (LR)**
- **Postoperative distant metastases (DM)**

LR may be defined as “histologically confirmed tumor occurring in the resected thyroid bed, thyroid remnant, or other adjacent tissues of the neck (excluding lymph nodes)” after complete surgical removal of the primary tumor. Nodal or distant spread may be considered postoperative if the metastases are discovered within 180 or 30 days, respectively. Ideally, tumor recurrence should be considered only as it occurs in patients without initial DM who had complete surgical resection of the primary tumors.

Figure 13-10 illustrates rates of PTC recurrence at local, nodal, and distant sites in 2150 patients with PTC treated at one institution from 1940 to 1997. After 20 years of follow-up, postoperative NM had been discovered in 9%, and LR and DM occurred in 5% and 4%, respectively. Both LR and DM are less common in PTC than in FTC.

Cause-specific mortality (CSM) rates for differentiated thyroid cancer are shown in Figure 13-12. CSM rates for PTC were 2% at 5 years, 4% at 10 years, and 5% at 20 years. Among those with lethal PTC, 20% of deaths occurred in the first year after diagnosis, and 80% of the deaths occurred within 10 years. The 25-year cause-specific survival rate of 95% for PTC was significantly higher than the 79%, 71%, and 66% rates seen with MTC, Hürthle cell cancer (HCC), and FTC, respectively.

### Outcome Prediction

Only a fraction (15%) of patients with PTC are likely to experience relapse of disease, and even fewer (5%) have a lethal outcome. Exceptional patients, who have an aggressive course, tend to experience relapse early (Fig. 13-13), and the rare fatalities usually occur within 5 to 10 years of diagnosis. Multivariate analyses have been used to identify variables predictive of CSM. Increasing age of the patient and the presence of extrathyroidal invasion are independent prognostic factors in all studies. The presence of initial DM, although relevant to future nodal recurrence, does not influence CSM.

Several scoring systems based on these significant prognostic indicators have been devised. Each system allows one to assign the majority of PTC patients (80% or more) to a low-risk
Information obtained at surgery, the system probably should not be used to decide the extent of primary surgery.

Tumor surveillance and the appropriateness of adjunctive radioiodine therapy. Because the CIS (completeness of resection, invasion, and size) variables require information obtained at surgery, the system probably should not be used to decide the extent of primary surgery.
Follicular Thyroid Carcinoma

FTC is "a malignant epithelial tumor showing evidence of follicular cell differentiation but lacking the diagnostic features of papillary carcinoma." Such a definition excludes the follicular variant of PTC, and it is also customary to exclude both the poorly differentiated insular carcinoma and the rare mixed medullary and follicular carcinoma. The correct classification of tumors with predominant oncocyctic features (Hürthle cell carcinomas) is controversial. The WHO committee has taken the stance that this tumor is an oxyphilic variant of FTC. The AFIP monograph, by contrast, states that "the tumors made up of this cell type have gross, microscopic, behavioral, cytogenetic (and conceivably etiopathogenic) features that set them apart from all others and justify discussing them in a separate section." Thus categorized, FTC is a relatively rare neoplasm whose identification requires invasion of the capsule, blood vessel, or adjacent thyroid. In epidemiologic surveys, FTC constituted from 5% to 50% of differentiated thyroid cancers and tended to be more common in areas with iodine deficiency. Owing to a combination of changing diagnostic criteria and an increase in the incidence of FTC associated with dietary iodine supplementation, the diagnosis of FTC has decreased in frequency; in one North American experience, minimally invasive nonoxyphilic FTC made up fewer than 2% of thyroid malignancies.

The microscopic appearance of FTC varies from well-formed follicles to a predominantly solid growth pattern. Poorly formed follicles and atypical patterns (e.g., cribiform) may occur, and multiple architectural types may coexist. Mitotic activity is not a useful indicator of malignancy.

FTC is best divided into two categories on the basis of degree of invasiveness:

- Minimally invasive or encapsulated
- Widely invasive

There is little overlap between these two types.

Minimally invasive FTC is an encapsulated tumor whose growth pattern resembles that of a trabecular or solid, microfollicular, or atypical adenoma. The diagnosis of malignancy depends on the demonstration of blood vessel or capsular invasion, or both. The criteria for invasion must therefore be strict. Blood vessel invasion is almost never seen grossly.

Microscopically, the vessels "should be of venous caliber, be located in or immediately outside of the capsule and contain one or more clusters of tumor cells attached to the wall and protruding into the lumen." Interruption of the capsule must involve the full thickness to qualify as capsular invasion. Penetration of only the inner half of the presence of tumor cells embedded in the capsule does not qualify for the diagnosis of FTC. Foci of capsular invasion must be distinguished from the capsular rupture that can result from FNA. The acronym WHAFFT (worrisome histologic alterations following FNA of the thyroid) is applied to such changes.

In contrast, the rare, widely invasive form of FTC can be distinguished easily from benign lesions. Although the tumor may be partially encapsulated, the margins are infiltrative even on gross examination and vascular invasion is often less. The structural features are variable, with solid and trabecular areas, but a follicular element is always present. When follicular differentiation is poor or absent, the tumor may be classified as a poorly differentiated (insular) carcinoma.

Focal or extensive clear-cell changes can occur. A rare clear cell variant of FTC has been described in which glycogen accumulation or dilatation of the granular endoplasmic reticulum is responsible for the clear cells. When more than 75% of cells in an FTC exhibit Hürthle cell (or oncocytic) features, the tumor is classified as a Hürthle cell or an oncocytic carcinoma or an oxyphilic variant FTC.

Molecular Pathogenesis

There is still no accepted paradigm for the pathogenesis of follicular thyroid cancer. A multistep adenoma-to-carcinoma pathogenesis, similar to that for colon cancer and other adenocarcinomas, is not universally accepted because pathologists do not recognize follicular carcinoma in situ and documentation of the evolution of adenoma to carcinoma is rare. Nevertheless, several facts about the pathogenesis of FTC are firmly established.

First, most follicular adenomas and all FTCs are probably of monoclonal origin. Second, oncogene activation, particularly by point mutation of the RAS oncogene, is common both in follicular adenomas and in FTCs (40%), supporting a role in early tumorigenesis. Such RAS mutations are not specific for follicular tumors and also occur in PTC. The RET oncogene does not appear to be significantly involved in follicular tumors. Third, cytogenetic abnormalities and evidence of genetic loss are more common in FTC than in PTC and also occur in follicular adenomas. Losses in FTC are particularly associated with chromosomes 3, 10, 11, and 17.

Of the cytogenetic abnormalities described in FTC, the most common are deletions, partial deletions, and deletion-rearrangements involving the p arm of chromosome 3. Loss of heterozygosity (LOH) on chromosome 3p appears to be limited to FTC because no evidence for 3p LOH has been found in follicular adenomas or PTC. A translocation, t(2;3)(q13;p25), resulting in the fusion of the deoxyribonucleic acid (DNA) binding domains of the thyroid transcription factor PAX-8 to domains of the peroxisome proliferator-activated receptor (PPAR) 1, was detected in five of eight FTCs but not in follicular adenomas, PTCs, or multinodular hyperplasia. The chimeric protein may retard growth inhibition and follicular differentiation normally induced by PPAR 1.

Presenting Features

FTC tends to occur in older people, with the mean age in most studies being more than 50 years, about 10 years older than that for typical PTC. The average median age of patients with oxyphilic FTC (HCC) is about 60 years. As in most thyroid malignancies, women outnumber men by more than 2 to 1. Most patients with FTC present with a painless thyroid nodule, with or without background thyroid nodularity, and they rarely (4% to 6%) have clinically evident lymphadenopathy at presentation. Lymph node metastases to the neck in FTC are so exceptional that "wherever they are observed, the alternative possibilities of follicular variant papillary carcinoma, oncocyctic carcinoma, and poorly differentiated (insular) carcinoma should be considered."

In most series in which tumor sizes were reported, the average tumor in FTC (oxyphilic or nonoxyphilic) was larger than those seen with PTC. Direct extrathyroidal extension, by definition, does not occur with minimally invasive FTC but is not uncommon in the rare patient with invasive FTC. Between 5% and 20% of patients may have DM at presentation. The most common sites for DM in FTC are lung and bone. The bones most often involved are long bones (e.g., femur), flat bones (particularly the pelvis, sternum, and skull), and vertebrae. When a DM is the first manifestation of the disease, definitive proof of its thyroid origin...
should be obtained, usually by a biopsy of a metastasis, before performing any thyroid surgery. It is unusual for patients with FTC to have thyrotoxicosis caused by massive tumor burden.\footnote{315}

Most patients (53\% to 69\%) with FTC or HCC have pTNM stage II disease. Patients aged 45 years or older with nodal metastases or extrathyroidal extension (stage III) account for only 4\% of FTCs and 9\% of HCCs (see Fig. 13-8).\footnote{217} About 5\% of HCCs and 17\% or more of nonoxyphil FTCs have DM at the time of diagnosis (stage IV).

### Recurrence and Mortality

Nodal metastases are rare in typical FTC, and the nodal recurrence rate at 20 postoperative years is the lowest in differentiated thyroid carcinoma, being around 2\% (see Fig. 13-11).\footnote{221} About 6\% of patients with HCC have node involvement at presentation,\footnote{222} but within 25 years after primary surgery about 17\% of HCC patients have nodal recurrence.\footnote{223} When recurrences at either neck or distant sites are taken into consideration, patients with HCC (Fig. 13-17) have the highest numbers of tumor recurrences after 10 to 20 years. As illustrated by Figure 13-11, local recurrences at 20 years have occurred in 20\% of FTCs and 30\% of HCCs. Comparable DM rates are 23\% and 28\%, respectively.

CSM rates vary with the presenting TNM stage in both FTC (Fig. 13-18) and HCC. The death rates tend to parallel the curves for development of DM (see Fig. 13-11).\footnote{224} In more than five decades of experience at the Mayo Clinic, the mortality rate for FTC initially exceeds that of HCC, but by 20 to 30 postoperative years there are no significant differences in cause-specific survival rates between FTC and HCC (Fig. 13-19), both being around 80\% at 20 and 70\% at 30 postoperative years.\footnote{225}

### Outcome Prediction

The risk factors that predict outcome in FTC are largely the same as in PTC.\footnote{219} DM at presentation, increasing age of the patient, large tumor size, and the presence of local (extrathyroidal) invasion. To a lesser degree, increased mortality is associated with male sex and higher grade (less well-differentiated) tumors. In addition, vascular invasiveness, lymphatic involvement at presentation, DNA aneuploidy, and oxyphilic histology are potential prognostic variables unique to FTC.\footnote{220} The importance of vascular invasion is underscored by a study showing that FTC patients with minimal capsular invasion and no evidence of vascular invasion had 0\% CSM at 10-year follow-up.\footnote{224}

Prognostic scoring systems for FTC\footnote{236} allow stratification of patients into high-risk and low-risk categories. A multivariate analysis at the Mayo Clinic found that DM at presentation, patient's age greater than 50 years, and marked vascular invasion predict a poor outcome.\footnote{226} As illustrated by Figure 13-21, if two or more of these factors are present, the 5-year survival rate is only 47\%, and 20-year survival is 8\%. By contrast, if only one of these factors is present, 5-year survival is 99\%, and 20-year survival is 86\%.\footnote{227}

Systems developed to predict outcome in either PTC or FTC have been applied to FTC patients. Specifically, the pTNM as well as the AMES risk group categorization (age, metastasis, extent, size) is useful in FTC.\footnote{228} From a multivariate analysis of 228 patients with FTC treated at the Memorial Sloan Kettering Cancer Center, the independent adverse prognostic factors were identified as age older than 45 years, Hürthle cell histotype, extrathyroidal extension, tumor size exceeding 4 cm, and the presence of DM.\footnote{229} The prognostic importance of FTC of histologic grade was also confirmed,\footnote{230} and this factor was included in assignment of risk groups to low, intermediate, or high categories (Fig. 13-22).\footnote{231}

The AGES scheme, originally developed for PTC, has also been successfully applied to FTC.\footnote{232} It thus appears that scoring systems used in PTC may be cautiously applied in FTC as long as some of the unique features of this tumor, such as vascular invasiveness and the remarkable significance of DNA aneuploidy in HCC, are kept in mind.\footnote{233}

---


\footnote{219} Comparison of cause-specific survival in 1472 papillary thyroid carcinoma (PTC) and 250 follicular thyroid carcinoma (FTC) patients treated at the Mayo Clinic from 1940 to 1999. Numbers in parentheses represent the number of patients in each pTNM stage grouping.

\footnote{220} Deaths related to HCC occur gradually over the first 15 years; however, by 25 years, the average survivor of HCC is 84 years old, and by that time, almost 50\% of the treated cohort would be predicted by the actuarial curve to have died from all causes.

\footnote{221} Curve representing mortality from all causes differ in FTC and HCC. On average, patients with FTC are about 5 years younger, tend to die within the first 10 postoperative years, and have high all-cause mortality for 10 to 30 postoperative years.\footnote{222} (Fig. 13-20). Deaths related to HCC occur gradually over the first 15 years; however, by 25 years, the average survivor of HCC is 84 years old, and by that time, almost 50\% of the treated cohort would be predicted by the actuarial curve to have died from all causes.

\footnote{223} CSM rates vary with the presenting TNM stage in both FTC (Fig. 13-18) and HCC. The death rates tend to parallel the curves for development of DM (see Fig. 13-11).

\footnote{224} In more than five decades of experience at the Mayo Clinic, the mortality rate for FTC initially exceeds that of HCC, but by 20 to 30 postoperative years there are no significant differences in cause-specific survival rates between FTC and HCC (Fig. 13-19), both being around 80\% at 20 and 70\% at 30 postoperative years.

\footnote{225} Curves representing mortality from all causes differ in FTC and HCC. On average, patients with FTC are about 5 years younger, tend to die within the first 10 postoperative years, and have high all-cause mortality for 10 to 30 postoperative years (Fig. 13-20). Deaths related to HCC occur gradually over the first 15 years; however, by 25 years, the average survivor of HCC is 84 years old, and by that time, almost 50\% of the treated cohort would be predicted by the actuarial curve to have died from all causes.

\footnote{226} The risk factors that predict outcome in FTC are largely the same as in PTC. DM at presentation, increasing age of the patient, large tumor size, and the presence of local (extrathyroidal) invasion. To a lesser degree, increased mortality is associated with male sex and higher grade (less well-differentiated) tumors. In addition, vascular invasiveness, lymphatic involvement at presentation, DNA aneuploidy, and oxyphilic histology are potential prognostic variables unique to FTC. The importance of vascular invasion is underscored by a study showing that FTC patients with minimal capsular invasion and no evidence of vascular invasion had 0\% CSM at 10-year follow-up.

\footnote{227} Prognostic scoring systems for FTC allow stratification of patients into high-risk and low-risk categories. A multivariate analysis at the Mayo Clinic found that DM at presentation, patient's age greater than 50 years, and marked vascular invasion predict a poor outcome. As illustrated by Figure 13-21, if two or more of these factors are present, the 5-year survival rate is only 47\%, and 20-year survival is 8\%. By contrast, if only one of these factors is present, 5-year survival is 99\%, and 20-year survival is 86\%.

\footnote{228} Systems developed to predict outcome in either PTC or FTC have been applied to FTC patients. Specifically, the pTNM as well as the AMES risk group categorization (age, metastasis, extent, size) is useful in FTC. From a multivariate analysis of 228 patients with FTC treated at the Memorial Sloan Kettering Cancer Center, the independent adverse prognostic factors were identified as age older than 45 years, Hürthle cell histotype, extrathyroidal extension, tumor size exceeding 4 cm, and the presence of DM. The prognostic importance of FTC of histologic grade was also confirmed, and this factor was included in assignment of risk groups to low, intermediate, or high categories.

\footnote{229} The AGES scheme, originally developed for PTC, has also been successfully applied to FTC. It thus appears that scoring systems used in PTC may be cautiously applied in FTC as long as some of the unique features of this tumor, such as vascular invasiveness and the remarkable significance of DNA aneuploidy in HCC, are kept in mind.
Poorly Differentiated (Insular, Solid, or Trabecular) Carcinoma

Poorly differentiated thyroid carcinoma has been defined as a tumor of follicular cell origin with morphological and biologic attributes intermediate between differentiated and anaplastic carcinomas of the thyroid. The most distinctive histologic feature is the presence of small cells with round nuclei and scant cytoplasm with a diffuse solid pattern or organized in round or oval nests (insulae) or in trabeculae. The predominant pattern of growth is solid, but microfollicles are also seen, some of which contain dense colloid. Extrathyroidal extension and blood vessel invasion are common. Most such tumors exhibit foci of necrosis, are larger than 5 cm in diameter at diagnosis, and have an invasive margin on gross examination.

The mean age at diagnosis is about 55 years, and the female/male ratio is about 2:1. Poorly differentiated carcinoma is aggressive and often lethal. Metastases are common in regional nodes and distant sites (lung, bone, brain). In one series,
Undifferentiated (Anaplastic) Carcinoma

Anaplastic carcinoma constitutes about 5% of all thyroid carcinomas, usually occurs after the age of 60 years, and is slightly more common in women (1.3:1 to 1.5:1). This carcinoma is highly malignant, nonencapsulated, and extends widely. Evidence of invasion of adjacent structures, such as the skin, muscles, nerves, blood vessels, larynx, and esophagus, is common. DM occur early in the course of the disease in lungs, liver, bones, and brain.

On histopathologic examination, the lesion is composed of atypical cells that exhibit numerous mitoses and form a variety of patterns. Spindle-shaped cells, multinucleate giant cells, and squamoid cells usually predominate. Areas of necrosis and polymorphonuclear infiltration are common, and the presence of PTC or FTC suggests that they may be the precursors of anaplastic carcinoma. Mutations of the p53 gene are present in many undifferentiated carcinomas but may not be found in the residual well-differentiated component, suggesting that these mutations occurred after the development of the original tumor and may have played a key role in tumor progression.

The usual clinical complaint is of a rapid, often painful enlargement of a mass that may have been present in the thyroid gland for many years. The tumor invades adjacent structures, causing hoarseness, inspiratory stridor, and difficulty in swallowing. On examination, the overlying skin is often warm and discolored. The mass is tender and is often fixed to adjacent structures. It is stony hard in consistency, but some areas may be soft or fluctuant. The regional lymph nodes are enlarged, and there may be evidence of DM. Anaplastic carcinomas do not accumulate iodine and do not typically produce thyroglobulin.

Treatment should be initiated rapidly to avoid death from locally infiltrative disease and possible suffocation. It consists of surgical resection of the tumor tissue present in the neck, when this is feasible, followed by a combination of external irradiation and chemotherapy. When the extent of disease is limited and when these protocols can be applied, local control may be obtained in about two thirds of the patients and long-term survival in about 20%.
Medullary Thyroid Carcinoma

MTC accounts for less than 10% of thyroid malignancies (see Chapter 36). It arises from the parafollicular or C cells of the thyroid gland, and the tumor cells typically produce an early biochemical signal (hypersecretion of calcitonin). MTC readily invades the intraglandular lymphatics and spreads to other parts of the gland, in addition to the pericapsular and regional lymph nodes. It also regularly spreads through the blood stream to the lungs, bone, and liver.

MTC tumors are firm and usually unencapsulated. On histopathologic examination, the tumor is composed of cells that vary in morphologic features and arrangement. Round, polyhedral, and spindle-shaped cells form a variety of patterns, which may vary from solid, trabecular to endocrine or glandular-like structures. An amyloid stroma is commonly present. Gross or microscopic foci of carcinoma may be present in other parts of the gland, and blood vessels may be invaded. The histopathologic appearance of the metastases resembles that of the primary lesion. In all cases, the diagnosis can be confirmed by positive immunostaining of tumor tissue for calcitonin and carciinoembryonic antigen (CEA).

MTC first appears either as a hard nodule or mass in the thyroid gland or as an enlargement of the regional lymph nodes. Occasionally, a metastatic lesion in a distant site is found first. The neck masses are frequently painful; they are sometimes bilateral and are often localized to the upper two thirds of each lobe of the gland, which reflects the anatomic location of the parafollicular cells.

The tumor occurs in both sporadic and hereditary forms, the latter making up about 20% of the total. The hereditary variety can be transmitted as a single entity, familial MTC, or it can arise as part of MEN syndrome type 2A or 2B. The hereditary form is typically bilateral and is usually preceded by a premalignant C cell hyperplasia. Total thyroidectomy at this premalignant stage can cure the disease in more than 90% of cases. RET proto-oncogene testing should be performed in all MTC patients. The finding of a germline mutation in this gene indicates a hereditary disease; the mutation should then be sought in all first-degree family members.

Early series of MTC mainly described sporadic cases, in which 80% of patients presented with TNM stage II or III. As more patients with familial MTC or MEN 2A have been diagnosed, more patients have curable (stage I) disease, and the survival rate has improved, a trend that should continue with widespread application of RET proto-oncogene testing. Patients with MTC now have outcomes similar to or better than those of patients with nonpapillary FCTC (see Fig. 13-10). The cause-specific survival curves for 218 consecutive MTC cases treated from 1940 to 1997 at the Mayo Clinic, according to TNM stage, are presented in Figure 13-23.

Prognostic factors relevant to outcome in MTC include (1) age at diagnosis, (2) male gender, (3) initial extent of the disease, such as NM and DM, (4) tumor size, (5) extrathyroidal invasion, (6) vascular invasion, (7) calcitonin immunoreactivity and amyloid staining in tumor tissue, (8) postoperative gross residual disease, and (9) postoperative plasma calcitonin levels.

In multivariate analysis, only the age of the patient at initial treatment and the stage of the disease remain significantly independent indicators of survival. This suggests that, in routine practice, clinicians attempting to predict outcome in MTC should take into account not only the presenting disease stage, as assessed by the pTNM system (see Fig. 13-23), but also the age of the patient at diagnosis.

In patients with MTC, Cushing's syndrome may occur because of secretion of corticotropin by the tumor. Prostaglandins, serotonins, kinins, and vasoactive intestinal peptide may also be secreted and are variously responsible for flushing and for the attacks of watery diarrhea that about one third of patients experience, usually at an advanced stage of the disease. In MEN-2A, hyperparathyroidism occurs late and is usually due to parathyroid hyperplasia rather than adenoma. Pheochromocytomas invariably occur later than MTC; they are often bilateral and may be clinically silent, and patients at risk should be screened with measurements of urinary metanephrine excretion. In MEN-2B, MTC and pheochromocytomas are associated with multiple mucosal neuromas (bumpy lip syndrome), a marfanoid habitus, and typical facies, but such patients do not regularly have hyperparathyroidism.

Differentiation of sporadic MTC from other types of thyroid nodule on clinical grounds alone may be difficult. In patients with a family history of thyroid cancer associated with hypertension or hyperparathyroidism, the MEN-2A syndrome should be suspected. FNAB has made it possible to diagnose MTC before surgery. In some patients, however, cytologic findings may be misleading because the type of carcinoma is difficult to determine and HCC may occasionally be confused with MTC.

Positive immunocytochemical staining for calcitonin allows confirmation of the diagnosis. Basal plasma calcitonin levels are elevated in virtually all patients with clinical MTC. Infusions of pentagastrin or calcium elicit secretion of calcitonin, and the response may be exaggerated in patients with either MTC or the antecedent C-cell hyperplasia; its use should be restricted to patients with an undetectable or borderline plasma calcitonin level (see Chapter 36).

When the diagnosis of MTC is made from calcitonin measurements or FNAB, patients should be evaluated for hyperparathyroidism and for pheochromocytoma. If these diagnoses are satisfactorily excluded, a total thyroidectomy with removal of regional nodes can safely be performed. In patients with MEN, surgery should be performed for pheochromocytomas before surgery for MTC is performed. First-degree relatives of patients with MEN or familial MTC should undergo DNA testing for the presence of the mutant RET gene (see Chapter 36). Gene carriers should undergo a prophylactic total thyroidectomy between 5 and 7 years of age.
Primary Malignant Lymphoma

Primary lymphomas of the thyroid are uncommon tumors, constituting fewer than 2% of all thyroid malignancies. The peak incidence is in the seventh decade, and the male/female ratio is 1:3.[237][238][239][240] Thyroid lymphomas are almost invariably seen as a rapidly enlarging, painless neck mass, fixed to surrounding tissues; they cause compressive symptoms and should be differentiated from anaplastic carcinoma. Unilateral or bilateral lymph node enlargement is present in about 50% of affected patients. Clinically evident distant disease is uncommon. The palpated mass is solid and, if studied by imaging, would be hypoechoic on ultrasonography and nonfunctioning on thyroid scintiscan. Most primary thyroid lymphomas arise in patients who have chronic autoimmune thyroiditis. Nonetheless, the disease is a rare complication of Hashimoto's thyroiditis.[241]

Primary thyroid lymphomas should be distinguished from generalized lymphomas with thyroid involvement. FNAB can be useful in distinguishing lymphoid proliferation from epithelial tumors. However, differentiating lymphoma from chronic autoimmune thyroiditis by thyroid cytology may be difficult.[242] Therefore, surgical specimens are needed for diagnosis. Immunohistochemical studies identify lymphoid proliferation if findings are positive for leukocyte common antigen.

Because chronic autoimmune thyroiditis reproduces the exact features of a mucosa-associated lymphoid tissue (MALT), most cases of thyroid lymphoma are considered MALT lymphomas.[243] Those small cell lymphomas are characterized by a low grade of malignancy, slow growth, and a tendency for recurrence in other MALT sites, such as the gastrointestinal or respiratory tract, the thymus, or the salivary glands.

A large proportion of clinical cases are large cell lymphomas and have an aggressive course. With immunohistochemistry, nearly all of them show B-cell markers. Monoclonality for light chain immunoglobulin is considered a strong indication of malignant lymphoma. Usually, immunohistochemistry is positive for BCL2 in small cell and negative in large cell lymphomas.

Although accurate staging is very important for planning treatment, patients are often elderly, in poor condition, or may require urgent therapy to relieve symptoms, thus making a full staging investigation before treatment impractical. Staging includes physical examination; complete blood count; serum lactate dehydrogenase and 2-microglobulin measurements; liver function tests; bone marrow biopsy; CT scanning of the neck, thorax, abdomen, and pelvis; and appropriate biopsies at sites where tumor is suspected. Involvement of Waldeyer's ring and of the gastrointestinal tract has been associated with thyroid lymphomas, and for this reason upper gastrointestinal radiography or endoscopy should be performed.

Disseminated disease necessitates chemotherapy. In patients with disease apparently confined to the neck, therapy is guided by the histologic features of the lymphoma. Chemotherapy with an anthracycline-based regimen and involved-field radiotherapy should be given to all patients with large cell thyroid lymphoma and in some series has provided long-term survival rates of nearly 100%. For small cell MALT lymphomas, radiation alone may be adequate if the disease is determined to be localized after accurate staging.[238][239][240]
SURGICAL TREATMENT OF THYROID CARCINOMA

The extent of surgery appropriate for thyroid malignancy is a matter of controversy. Factors that influence this decision include the histologic diagnosis, the size of the original lesion, the presence of DM, the patient's age, and the risk group category. Obviously, the surgeon must be appropriately skilled in thyroid surgery, and the goal of surgery should be to remove all the malignant neoplastic tissue present in the neck. Therefore, the thyroid gland and affected neck lymph nodes should all be carefully identified and adequately resected.

In the case of PTC and FTC, although some debate still exists regarding the extent of thyroid surgery, many favor a near-total (leaving no more than 2 to 3 g of thyroid tissue) thyroidectomy for all patients. Near-total thyroidectomy reduces the recurrence rate, compared with more limited surgery, because many PTCs are both multifocal and bilateral. Removal of most, if not all, of the thyroid gland facilitates postoperative remnant ablation with 131I.

For extremely low-risk patients (i.e., those with unifocal intra-thyroidal PTM and possibly small [≤2 cm] FTC with only capsular invasion), a lobectomy may be an appropriate primary surgical procedure. In patients who have undergone a previous unilateral lobectomy for a supposedly benign tumor that proves to be an aggressive FTC, a completion thyroidectomy is advisable because it facilitates future follow-up.

Surgery of lymph nodes is routinely performed in patients with PTC. It should include dissection of the central compartment (paratracheal and tracheoesophageal areas) and may also include dissection of the supravclicular area and the lower third of the jugulocarotid chain. A modified neck dissection is performed if palpable lymph node metastases are present in the jugulocarotid chain. Dissection is preferable to lymph node picking. Although this type of lymph node dissection has not been shown to improve the recurrence and survival rates, several arguments support its routine use in patients with papillary carcinomas. These include the fact that histologic evidence of lymph node metastases is present in about two thirds of PTC patients, of whom more than 80% have involvement of the central compartment, and metastases are difficult to detect by palpation in lymph nodes located behind the vessels or in the paratracheal groove. The knowledge of initial lymph node status, acquired by such a routine, helps in the interpretation of any cervical abnormality identified during the subsequent postoperative follow-up. In the case of FTC, lymph node metastases are less frequent, but a lymph node dissection should be performed if FTC has already been diagnosed and palpable lymph nodes are present.

MTC is usually treated by total thyroidectomy, with a dissection of the central compartment of the neck and the lower two thirds of the jugulocarotid chains. A modified neck dissection may be required for MTC affecting the lateral neck nodes.

Ideally, patients with anaplastic carcinoma should be treated with near-total thyroidectomy and lymph node dissection, but lesions are usually too extensive for any procedure but palliative surgery. In these cases, surgery may be performed later in the case of tumor regression after a combination of chemotherapy and external radiotherapy.

In recommending surgery, the endocrinologist should discuss potential operative complications with the patient. Unilateral lobectomy virtually never causes permanent hypocalcemia but can cause vocal cord paralysis in as many as 3% of patients. Near-total thyroidectomy causes temporary hypocalcemia in 7% to 10% of patients and permanent hypocalcemia in 0.5% to 1%; temporary vocal cord paralysis occurs in about 1% to 2%. A total extracapsular thyroidectomy may lead to hypoparathyroidism in as many as 30% of individuals, an unacceptable complication rate for patients with indolent malignancy. In addition, vocal cord paralysis is more common after such a procedure. The experience of the surgeon is important in terms of the finer technical points of thyroidectomy, including preservation of the external branch of the recurrent laryngeal nerve, which is important in the fine regulation of voice pitch.

A history of radiation in childhood increases the risk of both benign and malignant thyroid nodules in later life. The risk increases with a younger age at exposure and larger radiation dose. Several issues are relevant for the thyroidologist. First, given that surgical exploration may be required for patients with a history of thyroid radiation and suspicious thyroid nodules, what should the extent of initial surgery be and how should such patients be treated subsequently?

With respect to the extent of surgery, the protocol described previously should always be applied to patients with a thyroid carcinoma. In cases with benign lesions, individuals with bilateral nodular disease should have a near-total thyroidectomy; when the opposite lobe is macroscopically normal, one must weigh the relative risk of complications associated with a more extensive surgical procedure against the possibility of recurrence of thyroid nodules in the residual thyroid tissue.

In one irradiated population, both benign and malignant nodules recurred after previous subtotal thyroidectomy. The overall risk of recurrence in this study was approximately 20% and was lower in those who had more thyroid tissue removed than in those who had less extensive procedures. In those patients, suppression of TSH by thyroid hormone led to a reduction in recurrence from 35% to approximately 8%, but TSH suppression had no influence on the occurrence of malignant nodules.

Thus, the recommendations for such patients must take into account the estimated risk of developing a thyroid nodule and the experience of the operating surgeon. All irradiated patients who have had thyroid nodules removed should receive TSH-suppressive doses of levothyroxine regardless of the extent of surgery. The appearance of new thyroid nodules is, however, fairly common, and such patients should be monitored indefinitely for this possibility.

It is not clear whether this experience should be extrapolated to prescribe routine TSH suppression therapy for all irradiated patients, even if nodularity is not present, because its beneficial effects have not been quantified and the risks of long-term TSH suppression in women, especially vis-à-vis osteoporosis, have not been clearly defined and may be significant. At present, this approach cannot be recommended for all irradiated patients but can be recommended for patients at high risk of developing a thyroid nodule.
Iodine 131 Therapy

131 I is an effective agent for delivering high radiation doses to the thyroid tissue with low spillover to other portions of the body. The radiation dose to the thyroid tissue is related to the tissue concentration, the ratio between the total tissue uptake and the volume of functional tissue, and the effective half-life of 131 I in the tissue. Thyroid tissue is able to concentrate iodine only after TSH stimulation, but even after optimal TSH stimulation, iodine uptake in neoplastic tissue is always lower than in normal thyroid tissue and may not be detectable in about one third of cases.

131 I therapy is given postoperatively for three reasons. First, it destroys normal thyroid remnants, thereby increasing the sensitivity of subsequent 131 I total-body scanning and the specificity of measurements of serum Tg for the detection of persistent or recurrent disease. Second, it may destroy occult microscopic carcinoma, thereby potentially decreasing the long-term recurrence rate. Finally, it makes it possible to perform a postablative 131 I total-body scan, a sensitive tool for detecting persistent carcinoma.

It cannot be emphasized too strongly that postoperative 131 I therapy should be used selectively and that not all patients with a diagnosis of FCTC benefit from routine postoperative radiiodine ablative therapy. 131 I ablation requires that a dose of at least 300 Gy (30,000 rad) is delivered to thyroid remnants, and a dosimetric study can allow a more precise estimate of the dose to be administered. Obviously, the only patients eligible for such a protocol would be those selected high-risk patients with either PTC or FTC.

Postoperatively, no levothyroxine treatment is given for 4 to 6 weeks but levothyroxine can be substituted for at least 3 to 4 weeks and then discontinued for 2 weeks before radiiodine studies. At that time, the serum TSH level should be greater than 25 to 30 mU/L. Neck uptake may be measured with a tracer dose of 131 I; high uptake (>10%) should lead to completion surgery. 131 I therapy can be administered to the other patients, usually with 24-hour uptakes considerably less than 10%. A total body scan is performed 4 to 7 days after the treatment dose, and levothyroxine suppressive therapy is initiated. Total ablation (defined as no visible uptake) may be verified by an 131 I total-body scan 6 to 12 months later, typically with 2 to 5 mCi (74 to 185 MBq).

Total ablation is achieved after administration of either 100 mCi (3700 MBq) or 30 mCi (1100 MBq) in more than 80% of patients who had at least a near-total thyroidectomy. After less extensive surgery, ablation is achieved in only two thirds of patients with 30 mCi (1100 MBq). Therefore, a near-total thyroidectomy should be performed in all patients who are to be treated with 131 I. Total ablation requires that a dose of at least 300 Gy (30,000 rad) is delivered to thyroid remnants, and a dosimetric study can allow a more precise estimate of the dose to be administered. Obviously, the only patients eligible for such a protocol would be those selected high-risk patients with either PTC or FTC. 131 I ablation therapy does not play a regular role in the management of patients with anaplastic thyroid cancer, MTC, or thyroid lymphoma.

TABLE 13-6 — Indications for 131 I Treatment in Patients with Papillary, Follicular, or Hürthle Cell Thyroid Carcinoma after Initial Definitive Near-Total Thyroidectomy

<table>
<thead>
<tr>
<th>No indication</th>
<th>Indications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients at low risk of cause-specific mortality or of relapse (e.g., PTC patients with MACIS scores &lt; 6 and pTNM stage I FTC or HCC patients)</td>
<td></td>
</tr>
</tbody>
</table>
External Radiotherapy

External radiotherapy to the neck and mediastinum is indicated only for older patients with extensive PTC in whom complete surgical excision is impossible and in whom the tumor tissue does not take up $^{131}$I. Retrospective studies have shown that in these selected patients, external radiotherapy decreases the risk of neck recurrence.\[254\] The target volume encompasses the thyroid bed, bilateral neck lymph node areas, and the upper part of the mediastinum. Typically, 50 Gy (5000 rad) would be delivered in 25 fractions over 5 weeks.

In patients with MTC, this protocol may be applied after incomplete resection of the tumor and also after apparently complete surgery, when plasma calcitonin remains detectable in the absence of DM. In these patients, it may decrease the risk of neck recurrence by a factor of 2 to 4.\[254\]

In patients with anaplastic thyroid carcinoma, when the extent of disease is limited and surgery is feasible, accelerated external radiotherapy in combination with chemotherapy permits local control of the disease in two thirds of the patients and long-term survival in about 20%.\[153\]
Levothyroxine Treatment

The growth of thyroid tumor cells is controlled by TSH, and inhibition of TSH secretion with levothyroxine is thought to improve the recurrence and survival rates. Therefore, levothyroxine should be given to all patients with FCTC, whatever the extent of thyroid surgery and other treatment. The initial effective dose is about 2.5 µg/kg body weight in adults; children require a higher dose. The adequacy of therapy is monitored by measuring serum TSH 3 months after it is begun, the initial goal being a serum TSH concentration of 0.1 mU/L or less. In some centers, the serum free T3 concentration is also documented to be within the normal range. When these guidelines are followed, levothyroxine therapy does not have deleterious effects on the heart or bone.

In patients with anaplastic thyroid carcinoma, MTC, or thyroid lymphoma, a replacement dose of levothyroxine is given with the aim of obtaining a serum TSH level in the normal range.
FOLLOW-UP

In patients with PTC or FTC, the goals of follow-up after initial therapy are to maintain adequate levothyroxine suppressive therapy and to detect persistent or recurrent thyroid carcinoma. Most recurrences occur during the first years of follow-up, but some occur late. Therefore, follow-up is necessary throughout the patient’s life.

Early Detection of Recurrent Disease

Clinical and Ultrasonographic Examinations

Palpation of the thyroid bed and lymph node areas should be routinely performed at all follow-up visits in patients with thyroid cancer. Ultrasonography should be performed in patients at high risk of recurrence and in any patient with clinically suspicious findings. Palpable lymph nodes that are small, thin, or oval; in the posterior neck chains; and especially if they decrease in size after an interval of 3 months are considered benign. By contrast, round shape, hypoechogenicity and absence of a central echogenic line, microcalcifications, a cystic component, and hypervascularization on color Doppler ultrasonography are suspicious findings.

Serum Tg is undetectable in 20% of patients receiving levothyroxine treatment who have isolated lymph node metastases, and undetectable values do not exclude metastatic lymph node disease. If in doubt, ultrasound-guided node biopsy for cytology and Tg measurement in the fluid aspirate may be performed. Sensitive reverse transcriptasepolymerase chain reaction (RT-PCR) to amplify Tg mRNA appears to be even more sensitive but is not yet being used by commercial laboratories.

Radiographs

Bone and chest radiographs are no longer routinely obtained for patients with undetectable serum Tg concentrations. The reason is that virtually all patients with abnormal radiographs have readily detectable serum Tg concentrations.

Serum Thyroglobulin Determinations

Tg is a glycoprotein that is produced only by normal or neoplastic thyroid follicular cells. Methods used for serum Tg determination and serum interferences are detailed in Chapter 10. It should not be detectable in patients who have had total thyroid ablation, and its detection in that setting probably signifies the presence of persistent or recurrent disease (Table 13-7). In patients who are in complete remission after total thyroid ablation, serum Tg antibodies decline gradually to low or undetectable levels. Their persistence or reappearance during follow-up should be considered suspicious for persistent or recurrent disease.

<table>
<thead>
<tr>
<th>TABLE 13-7 -- Percentages of Patients with Detectable (&gt;1 ng/mL) Serum Thyroglobulin Concentrations during Thyroxine Treatment and after Discontinuation of Thyroxine According to the Presence or Absence of Normal Thyroid Tissue.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thyroid Tumor Status</td>
</tr>
<tr>
<td>Thyroxine treatment</td>
</tr>
<tr>
<td>Complete remission</td>
</tr>
<tr>
<td>Lymph node metastases</td>
</tr>
<tr>
<td>Distant metastases with normal radiographs</td>
</tr>
<tr>
<td>Large distant metastases</td>
</tr>
</tbody>
</table>

Most serum thyroglobulin concentrations were above 10 ng/mL.

The production of Tg by both normal and neoplastic thyroid tissue is in part TSH dependent. When serum Tg is detectable during levothyroxine treatment, it increases after TSH stimulation obtained either after the treatment is discontinued or with injections of rhTSH. After rhTSH stimulation, the peak of serum Tg is usually obtained 3 days after the second injection. The serum Tg concentration is an excellent prognostic indicator. Most patients with undetectable serum Tg concentrations who were not receiving levothyroxine therapy remained free of relapse after more than 15 years of follow-up, and only 2% had a neck lymph node recurrence. Conversely, 60% to 80% of patients with serum Tg concentrations above 10 ng/mL after levothyroxine withdrawal and with no other evidence of disease had detectable foci of 131I uptake in the neck or at distant sites after administration of a large dose (3700 MBq, 100 mCi) of 131I.

Several researchers have developed sensitive RT-PCR assays to amplify circulating Tg mRNA. The technique appeared sensitive in that the test was positive in most patients with thyroid tissue, but results were not related to the extent of the disease. This technique may be useful in patients with Tg antibodies.

Isotope 131I Total-Body Scan

The results of a 131I total-body scan depend on the ability of neoplastic thyroid tissue to take up 131I in the presence of high serum TSH concentrations, which are achieved by withdrawing levothyroxine for 4 to 6 weeks. However, the resulting hypothyroidism is poorly tolerated by some patients. This effect can be attenuated by substituting the more rapidly metabolized liothyronine for levothyroxine for 3 to 4 weeks and withdrawing it for 2 weeks or simply by reducing the dose of levothyroxine by 50%.

The serum TSH concentration should be above some arbitrary value (>25 to 30 mU/L) in patients treated in this way; if it is not, 131I administration should be delayed until it is. Intramuscular injections of rhTSH (0.9 mg for 2 consecutive days) are an alternative because levothyroxine treatment need not be discontinued and side effects are minimal. When combining serum Tg measurement and 131I total-body scanning, its efficiency is comparable to that of levothyroxine withdrawal in most patients.

| TABLE 13-8 -- Nonthyroidal Conditions Associated with 131I Accumulation |

Contamination
Consideration given to surgical reexploration if levothyroxine is discontinued or after rhTSH stimulation. If residual neck disease is found on sonography, the diagnosis should be confirmed by guided biopsy and performed annually; neck ultrasonography is frequently performed in case of doubt or in high-risk patients, but any other testing is unnecessary as long as the patient’s serum Tg concentration is undetectable and the patient does not produce an interfering anti-Tg autoantibody.

In low-risk patients considered cured, the dose of levothyroxine is decreased to maintain a low but detectable serum TSH concentration (0.1 to 0.5 mU/L). In high-risk patients, higher doses of levothyroxine are given, the goal being a serum TSH concentration less than 0.1 mU/L.

Follow-up Strategy

If the total-body scan performed after administration of 131I to destroy the thyroid remnants does not show any uptake outside the thyroid bed, physical examination is performed and serum TSH and Tg are measured during levothyroxine treatment 3 months later (Fig. 13-26). In most centers, the serum Tg level is measured and a diagnostic 131I total-body scan is done after thyroid hormone withdrawal or rhTSH stimulation 6 to 12 months later. Visible uptake in the thyroid bed that is too low to be quantified should not be considered evidence of disease in the absence of any other abnormality.

If any significant uptake is detected outside the thyroid bed, a therapeutic dose of 100 mCi (3700 MBq) of 131I is given. Serum Tg determination after TSH stimulation, obtained after either thyroid hormone withdrawal or injections of rhTSH, may help to select for scanning with a large amount of 131I those patients with negative diagnostic 131I total-body scans who have detectable serum Tg levels.

In low-risk patients considered cured, the dose of levothyroxine is decreased to maintain a low but detectable serum TSH concentration (0.1 to 0.5 mU/L). In high-risk patients, higher doses of levothyroxine are given, the goal being a serum TSH concentration less than 0.1 mU/L. Clinical and biochemical evaluations are performed annually; neck ultrasonography is frequently performed in case of doubt or in high-risk patients, but any other testing is unnecessary as long as the patient’s serum Tg concentration is undetectable and the patient does not produce an interfering anti-Tg autoantibody.

In patients receiving levotyroxine in whom serum Tg becomes detectable, neck ultrasonography is performed and serum Tg may be measured again after levotyroxine is discontinued or after rhTSH stimulation. If residual neck disease is found on sonography, the diagnosis should be confirmed by guided biopsy and consideration given to surgical reexploration.
of the neck. If there is no demonstrable neck disease but the stimulated serum Tg concentration increases above 10 ng/mL, even if no uptake is seen on a diagnostic $^{131}$I total-body scan performed with 2 to 5 mCi, consideration should be given to the administration of a therapeutic dose of 100 mCi of $^{131}$I. In the absence of $^{131}$I uptake, spiral CT of the neck and lungs, bone scintigraphy, and FDG PET scanning can be useful in localizing hitherto unrecognized sites of recurrent disease. In patients whose serum Tg levels are initially undetectable during levothyroxine treatment but later become detectable (levels $<$10 ng/mL) after TSH stimulation, another Tg determination should be obtained after TSH stimulation every 2 to 5 years, depending on Tg levels and on prognostic factors.

In low-risk PTC patients who have had a near-total thyroidectomy but who were not given $^{131}$I postoperatively, the intensity of the follow-up strategy depends largely on the serum Tg level. If the Tg is not detectable and a neck ultrasound is negative, $^{131}$I total-body scanning may be avoided. However, if despite adequate TSH suppression, the Tg is readily detectable, a $^{131}$I total-body scan may be performed 6 to 12 months after surgery. An ablative $^{131}$I treatment may rarely be necessary in some of those patients who have either an elevated serum Tg level or abnormal findings on $^{131}$I total-body scanning. The follow-up protocol previously described is then applied on the basis of serum Tg determinations.

In low-risk PTC patients who have initially undergone only a unilateral lobectomy for small ($<$15 mm) tumors, yearly follow-up should consist of a careful neck examination and serum Tg determination during levothyroxine treatment. With time, ultrasonography is likely to show focal nodular abnormalities in the remaining lobe in most patients with detectable Tg concentrations. Usually, biopsies of these lesions can be performed under sonographic guidance, and most prove to be cytologically benign. However, if recurrent PTC is found on biopsy, a completion thyroidectomy should be performed.

For MTC patients, the tumor marker for follow-up is the plasma calcitonin level. In more than 90% of young patients whose disease is treated at a preclinical stage on the basis of a RET oncogene mutation, the postoperative calcitonin level returns to normal and peak levels after stimulation with either pentagastrin or calcium are absent. In patients with a negative pentagastrin stimulation test after two follow-up evaluations are likely to be cured, despite the fact that about 5% of them have subsequent biologic recurrence of the disease.

In adults with sporadic MTC, who most often present with TNM stage III (node-positive) disease, postoperative calcitonin levels are rarely normal, and normal responsiveness to pentagastrin stimulation is unusual. In general, basal and stimulated calcitonin levels correlate with MTC tumor mass, but many MTC patients who have surgery with a curative intent still have postoperative elevations in calcitonin levels without clinical or imaging evidence of persistent disease. In these patients, the localization of neoplastic foci may be difficult and may require a venous sampling catheterization with calcitonin measurements. Reinterventions based on the results of selective venous sampling catheterization allow the removal of neoplastic foci in most patients, but they are not likely to improve the cure rate by more than 5% to 30%. Such a situation may exist for several postoperative years, and slowly rising calcitonin levels may not necessarily imply a prognosis worse than that indicated by the presenting stage of disease.

A second major tumor marker for MTC is CEA. In general, serum CEA levels are higher in more malignant MTC, whereas the plasma calcitonin level is higher in those with better differentiated tumors, leading some authorities to suggest that a rising CEA level postoperatively correlates better with the emergence of a potentially aggressive tumor recurrence.
Papillary and Follicular Thyroid Carcinoma

Locoregional Recurrences

Locoregional recurrences occur in 5% to 20% of patients with PTC and FTC. A recurrence that is palpable or easily visualized with ultrasonography or CT scanning should be excised. Total excision may be facilitated by total-body scanning 4 days after administration of 100 mCi (3700 MBq) of I, because additional tissue that should be excised may be identified. In some selected centers, surgery is performed 1 day later, typically using an intraoperative probe. The completeness of resection is verified 1 to 2 days after surgery by another total-body scan, and in one series this was achieved in 92% of cases. External radiotherapy is indicated only in FTC patients with soft tissue recurrences that cannot be completely excised and that do not take up I. Recently, it has been reported that patients with FTC who were not eligible for further surgery or I therapy have been treated for regional nodal recurrence with ultrasound-guided radiofrequency ablation or percutaneous ethanol injections (PEI). Both techniques appear promising in selected PTC patients with recurrent nodal disease that is not amenable to conventional retreatment with surgery, I, or external irradiation.

Distant Metastases

In a large group of patients with differentiated carcinoma (PTC, FTC, and HCC), only 9% developed DM. Mortality rates at 5 and 10 years after the diagnosis of metastasis were 65% and 75% for all patients with DM, and nearly 80% of the deaths were due to thyroid cancer. Thus, the development of DM in FTC portends an ominous prognosis. Lung metastases are more frequent in young patients with PTC, and the lung is almost the only site of distant spread in children. Bone metastases are more common in older patients and in those with FTC. Other less common sites are the brain, liver, and skin.

Clinical symptoms of lung involvement are uncommon. By contrast, pain, swelling, or fracture occurs in more than 80% of patients with bone metastases. The pattern of lung involvement may vary from macronodular to diffuse infiltrates. The latter, when not detected by chest radiography, are usually diagnosed with I total-body scan and may be confirmed by spiral CT; enlarged mediastinal lymph nodes are often present in patients with PTC, especially children. Bone metastases are osteolytic and are often difficult to visualize on radiographs. Skeletal scintigraphy may show decreased or moderately increased uptake, and bone involvement is better visualized by CT or MRI. Nearly all patients with DM have high serum Tg concentrations unless the lung metastases are not visible on radiographs, and two thirds of such patients have uptake in their sites of metastases.

Palliative surgery is required for bone metastases when there are neurologic or orthopedic complications or a high risk of such complications. Surgery may also be performed with a curative intent in patients with a single or a few bone metastases.

Patients with DM that take I should be treated with 100 to 150 mCi (3700 to 5550 MBq) every 4 to 6 months. Between I treatments, suppressive doses of levothyroxine are given. The radiation dose to the tumor tissue and outcome of I therapy are correlated. A radiation dose higher than 80 Gy (8000 rads) should be delivered to obtain cure; with radiation doses less than 35 Gy (3500 rads), there is little chance for success. For treatment to be effective in this clinical setting, appropriate levels of TSH stimulation and absence of iodine contamination are essential. For this reason, higher doses (200 mCi [7400 MBq] or more) have been advocated in patients with bone metastases, but their effectiveness remains to be demonstrated. Lower doses (1 mCi [37 MBq]/kg body weight) are given to children.

There is no limit to the cumulative dose of I that can be given to patients with DM, although the risk of leukemia rises slightly above a cumulative dose of 500 mCi (18,500 MBq). Furthermore, above this dose, further I therapy may rarely provide benefit. External radiotherapy is given to bone metastases visible on radiographs, even in the presence of iodine uptake. Alternatively, embolization or cement injection may be considered. Chemotherapy is poorly effective and should be given only to patients with progressing and nonfunctioning metastases. Retinoic acid analogues increased iodine uptake by neoplastic tissue and decreased its growth rate in several in vitro models. Further clinical trials are still warranted to assess the role of such therapies.

Complete responses have been obtained in about 45% of patients with DM showing avidity for I, and responses are even more frequent in younger patients and in those with small pulmonary metastases. It was shown by PET scanning that large DM with high FDG uptake almost never respond to I therapy. When response was judged to have been complete after I therapy, subsequent relapse rarely occurred even though serum Tg levels were persistently detectable in some patients.

Overall survival after the discovery of DM is more favorable in young patients with well-differentiated tumors that take up I and have metastases that are small when discovered. When the tumor mass is considered, the location of the DM, be it in the lungs or bone, has no independent prognostic influence. The poor prognosis of patients with bone metastases is linked to the large size of their lesions. The prognostic importance of the small size of the metastases at their discovery has led to the administration of 100-mCi (3700-MBq) doses of I to patients with elevated serum Tg concentrations and no other evidence of disease. Some believe that there is no conclusive evidence that I treatment of these asymptomatic patients meaningfully prolongs life. Others recently have reported a 33% complete remission rate in treated patients who had a positive post-I therapy total-body scan.

Complications of Treatment with Iodine 131

Acute side effects (nausea, sialadenitis) after treatment with I are common but are typically mild and resolve rapidly. Radiation thyroiditis is usually trivial, but if the thyroid remnant is large, the patient may have enough pain to warrant corticosteroid therapy for a few days. Tumor in certain locations, such as the brain, spinal cord, and paratracheal, may swell in response to TSH stimulation or after I therapy, causing compressive symptoms. Radiation fibrosis may develop in patients with diffuse lung metastases and can eventually prove fatal if high doses (>150 mCi [5550 MBq]) are administered at short intervals (<3 months).

Particular attention must be paid to avoid administration of I to pregnant women. After I treatment, spermatogenesis may be transiently depressed, and women may have transient ovarian failure. Genetic damage induced by exposure to I before conception has been a major subject of concern. However, the only anomaly reported to date is an increased frequency of miscarriages in women treated with I during the year preceding the conception. Therefore, it is recommended that conception be postponed for 1 year after treatment with I. There is no evidence that pregnancy affects tumor growth in women receiving adequate levothyroxine therapy. During pregnancy, the serum TSH level should be measured every 2 months, and this frequency leads to an increase in the daily dose of levothyroxine.

Mild pancytopenia may occur after I therapy, especially in patients with bone metastases also treated with external radiotherapy. The overall relative risk of leukemia was found to be increased only in patients treated with a high cumulative dose of I (>500 mCi [18,500 MBq]) or in association with external radiotherapy. In contrast, there is no significant increased risk of solid carcinoma in these patients.
Medullary Thyroid Carcinoma

For patients with locoregional recurrence of MTC, a complete diagnostic work-up should be obtained, principally to exclude DM. Surgery is performed when feasible and is typically followed by external radiotherapy.

DM are usually multifocal in each involved organ and frequently involve multiple organs, including liver, lungs, and bones. They may progress slowly and may be compatible with decades of survival. Systemic chemotherapy is poorly efficient and may be indicated only in cases of rapid tumor progression.
References


80. Rosas DS. Thyroid hormone suppressive therapy of sporadic nontoxic goiter. Thyroid 1992; 2:263-269.


90. Muller CG, Bayal TA, Harrison JE, et al. Possible limited bone loss with suppressive thyroxine therapy is unlikely to have clinical relevance. Thyroid 1995; 5:8187.


References


Guimaraes V, De Groot LJ. Moderate hypothyroidism as preparation for whole body 131I scans and thyroglobulin testing. Thyroid 1996; 6:6973.


Cooper DS, Specker B, Ho M, et al. Thyroidropin suppression and disease progression in patients with differentiated thyroid cancer: results from the National Thyroid Cancer Treatment Cooperative Registry. Thyroid 1998; 8:73744.


Saad MF, Fritsche HA, Samaan NA. Diagnostic and prognostic values of carcinoemobryonic antigen in medullary carcinoma of the thyroid. J Clin Endocrinol Metab 1984; 58:889894.


McDougall IR. Combination of radiodine (I-131) and probe-guided surgery for persistent or recurrent thyroid carcinoma. J Clin Endocrinol Metab 2000; 85:26752680.


HISTORICAL MILESTONES

The anatomy of the adrenal glands was described almost 450 years ago by Bartholomeo Eustacius, and the zonation of the gland and its distinction from the medulla were elucidated shortly thereafter. However, a functional role for the adrenal glands was not accurately defined until the pioneering work of Thomas Addison, who described the clinical and autopsy findings in 11 cases of "Addison's disease" in his classical monograph in 1855. Just a year later, Brown-Séquard demonstrated that the adrenal glands were "organs essential for life" by performing adrenalectomies in dogs, cats, and guinea pigs. In 1896, William Osler first administered adrenal extract to a patient with Addison's disease, a feat that was repeated by others in both animal and human studies over the next 40 years. As a consequence, between 1937 and 1955 the adrenocorticosteroid hormones were isolated, their structures defined, and the hormones synthesized, notable breakthroughs being the discovery of cortisone and the clinical evaluation of its anti-inflammatory effect in patients with rheumatoid arthritis (Reichstein, Hench, Kendall, and Slocumb) and the isolation of aldosterone (Simpson and Tait). The control of adrenocortical function by a pituitary factor was demonstrated in the 1920s, and this led to the isolation of sheep adrenocorticotropic hormone (ACTH) in 1943. Such a concept was supported through clinical studies, notably in 1932 by Harvey Cushing, who associated his original clinical observations of 1912 (a "polyglandular syndrome" caused by pituitary basophilism) with adrenal hyperactivity. The neural control of pituitary ACTH secretion by corticotropin-releasing hormone (CRH) was defined by Harris and other workers in the 1940s, but CRH was not characterized and synthesized until 1981 in the laboratory of Wylie Vale. Jerome Conn described primary aldosteronism in 1956, and the control of adrenal aldosterone secretion by angiotensin II was confirmed shortly afterward. Advances in radioimmunoassays and particularly molecular biology have facilitated an exponential increase in our understanding of adrenal physiology and pathophysiology (Table 14-1).
ANATOMY

The adrenal cortex derives from mesenchymal cells attached to the coelomic cavity lining adjacent to the urogenital ridge. The fetal adrenal is evident from 6 to 8 weeks of gestation and rapidly increases in size so that by midgestation it is larger than its adjacent kidney. In fetal life and up to 12 months post partum two distinct zones are evident: an inner prominent fetal zone and an outer definitive zone that differentiates into the adult adrenal gland. Post partum the fetal zone regresses and the definitive zone containing an inner zona fasciculata and outer glomerulosa proliferates. The innermost zone, the zona reticularis, is evident after 1 year of life. The differentiation of the adrenal cortex into distinct zones has important functional consequences (discussed later) and is thought to be dependent upon the temporal expression of transcription factors including Pref-1/ZOG, inner zone antigen, and steroidogenic factor-1.

The adult gland is a pyramidal structure approximately 4 g in weight, 2 cm wide, 5 cm long, and 1 cm thick lying immediately above the kidney on its posteromedial surface. Beneath the capsule, the zona glomerulosa constitutes approximately 15% of the cortex (depending upon sodium intake). Cells are clustered in spherical nests and are small with smaller nuclei in comparison with other zones. The zona fasciculata makes up 75% of the cortex; cells are large and lipid laden and form radial cords between the fibrovascular radial network. The innermost zona reticularis is sharply demarcated from both the zona fasciculata and adrenal medulla. Cells here are irregular with little lipid content. The maintenance of normal adrenal size appears to involve a progenitor cell population lying between the zona glomerulosa and zona fasciculata; cell migration and differentiation occur within the fasciculata and senescence within the reticularis, but the factors regulating this important aspect of adrenal regeneration are unknown. ACTH administration results in glomerulosa cells adopting a fasciculata phenotype; in turn, the innermost fasciculata cells adopt a reticularis phenotype, which is reversible upon withdrawal of ACTH.

TABLE 14-1 — History of the Adrenal Cortex: Important Milestones

<table>
<thead>
<tr>
<th>Year</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>1563</td>
<td>Eustachius describes the adrenals (published by Lancisi in 1714).</td>
</tr>
<tr>
<td>1849</td>
<td>Thomas Addison, while searching for the cause of pernicious anemia, &quot;stumbles&quot; on a bronzed appearance associated with the adrenal glands &quot;melasma suprarenale.&quot;</td>
</tr>
<tr>
<td>1855</td>
<td>Thomas Addison describes the clinical features and autopsy findings of 11 cases of diseases of the suprarenal capsules, at least 6 of which were tuberculous in origin.</td>
</tr>
<tr>
<td>1856</td>
<td>In adrenalectomy experiments, Brown-Séquard demonstrates that the adrenal glands are essential for life.</td>
</tr>
<tr>
<td>1896</td>
<td>William Osler gives an oral glycerine extract derived from pig adrenals and demonstrates clinical benefit in patients with Addison’s disease.</td>
</tr>
<tr>
<td>1905</td>
<td>Bulloch and Sequiera describe patients with congenital adrenal hyperplasia.</td>
</tr>
<tr>
<td>1929</td>
<td>Liquid extracts of cortical tissue are used to keep adrenalectomized cats alive indefinitely (Swingle and Pfiffner). Subsequently, this extract was used successfully to treat a patient with Addison’s disease (Rowntree and Greene).</td>
</tr>
<tr>
<td>1932</td>
<td>Harvey Cushing associates the &quot;polyglandular syndrome&quot; of pituitary basophilism first described by him in 1912 with hyperactivity of the pituitary-adrenal glands.</td>
</tr>
<tr>
<td>1936</td>
<td>Concept of stress and its effect upon pituitary-adrenal function described by Seyle.</td>
</tr>
<tr>
<td>1937/1952</td>
<td>Isolation and structural characterisation of adrenocortical hormones (Kendall, Reichstein).</td>
</tr>
<tr>
<td>1943</td>
<td>Li and colleagues isolate pure adrenocorticotropic hormone from sheep pituitary.</td>
</tr>
<tr>
<td>1950</td>
<td>Hench, Kendall, and Reichstein share Nobel Prize in medicine for describing the anti-inflammatory effects of cortisone in patients with rheumatoid arthritis.</td>
</tr>
<tr>
<td>1953</td>
<td>Isolation and analysis of the structure of aldosterone (Simpson and Tait).</td>
</tr>
<tr>
<td>1956</td>
<td>Conn describes primary aldosteronism.</td>
</tr>
</tbody>
</table>

The vasculature of the adrenal cortex is complex. Arterial supply is conveyed by up to 12 small arteries from the aorta, inferior phrenic, renal, and intercostal arteries. These branch to form a subcapsular arteriolar plexus from which radial capillaries penetrate deeper into the cortex. In the zona reticularis, a dense sinusoidal plexus is created that empties into a central vein. The right adrenal vein is short, draining directly into the inferior vena cava; the longer left adrenal vein usually drains into the left renal vein.
ADRENAL STEROIDS AND STEROIDOGENESIS

Three main types of hormones are produced by the adrenal cortex: glucocorticoids (cortisol, corticosterone), mineralocorticoids (aldosterone, deoxycorticosterone [DOC]), and sex steroids (mainly androgens). All steroid hormones are derived from the cyclopananoperhydrophenanthrene structure, that is, three cyclohexane rings and a single cyclopentane ring (Fig. 14-2).

Steroid nomenclature is defined in two ways: either by trivial names (e.g., cortisol, aldosterone) or by the chemical structure as defined by the International Union of Pure and Applied Chemistry (IUPAC). The IUPAC classification is inappropriate for clinical use but does provide invaluable insight into steroid structure. The basic structure and trivial and IUPAC names of some common steroids are given in Figure 14-2 and Table 14-2. Estrogens have 18 carbon atoms (C₁₈ steroids), androgens have 19 carbon atoms (C₁₉), and glucocorticoids and progesterogens are C₂₁ steroid derivatives.

Cholesterol is the precursor for all adrenal steroidogenesis. Most of this cholesterol is provided from the circulation in the form of low-density lipoprotein (LDL) cholesterol. Uptake is by specific cell surface LDL receptors present on adrenal tissue; LDL is then internalized by receptor-mediated endocytosis, the resulting vesicles fuse with lysosomes, and free cholesterol is produced after hydrolysis. However, it is clear that this cannot be the sole source of adrenal cholesterol; patients with abetalipoproteinemia who have undetectable circulating LDL and patients with defective LDL receptors in the setting of familial hypercholesterolemia still have normal basal adrenal steroidogenesis. Cholesterol can be generated de novo within the adrenal cortex from acetyl coenzyme A. In addition, there is evidence that the adrenal can utilize HDL cholesterol after uptake through the putative HDL receptor, the class B, type I scavenger receptor (SR-BI).

The biochemical pathways involved in adrenal steroidogenesis are shown in Figure 14-3. The initial hormone-dependent rate-limiting step is the transport of intracellular cholesterol from the outer to the inner mitochondrial membrane for conversion to pregnenolone by cytochrome P450. Human experiments of nature have confirmed the importance of a 30-kd protein, steroidogenic acute regulatory protein (StAR), in mediating this effect (see later). StAR is induced by an increase in intracellular cyclic adenosine monophosphate after binding of ACTH to its cognate receptor, providing the first important rate-limiting step in adrenal steroidogenesis. Other transporters, including the peripheral benzodiazepine-like receptor, may be involved.

Steroidogenesis involves the concerted action of several enzymes, including a series of cytochrome P450 enzymes, all of which have been cloned and characterized (Table 14-3). Cholesterol side-chain cleavage enzyme and the CYPI1B enzymes are localized to the mitochondria and require an electron shuttle system, provided through adrenodoxinreductin reductase, to oxidize and hydroxylate steroids. 17-Hydroxylase and 21-hydroxylase are localized to the microsomalendoplasmic reticulum fraction and are dependent upon a distinct electron shuttle system involving a flavoprotein, cytochrome b₅. Mutations in the genes encoding these enzymes result in human disease, so some understanding of the underlying pathways and steroid precursors is required.

After uptake of cholesterol to the mitochondrion, cholesterol is cleaved by the P450 cholesterol side-chain cleavage enzyme to form pregnenolone. In the cytoplasm, pregnenolone is converted to progesterone by the type II isozyme of 3-hydroxysteroid dehydrogenase (3-HSD) by a reaction involving dehydrogenation of the 3-hydroxy group and isomerization of the double bond at C5. Progesterone is hydroxylated to 17-OHP through the activity of CYPI7-hydroxylase. 17-Hydroxylation is a prerequisite for glucocorticoid synthesis, and the zona glomerulosa does not express 17-hydroxylase. CYPI7 also possesses 17,20-lyase activity, which results in the production of the C₁₈ adrenal androgens, dehydroepiandrosterone (DHEA) and androstenedione. In humans, however, 17-OHP is not an efficient substrate for CYPI7, and there is negligible conversion of 17-OHP to androstenedione. Adrenal androstenedione secretion is dependent on the conversion of DHEA to androstenedione by 3-HSD; this enzyme also converts 17-hydroxy pregnenolone to 17-OHP, but the preferred substrate is pregnenolone.

21-Hydroxylation of either progesterone (zona glomerulosa) or 17-OHP (zona fasciculata) is carried out by the product of the CYPI1B2 gene, 21-hydroxylase, to yield DOC or 11-deoxycortisol, respectively. The final step in cortisol biosynthesis takes place in the mitochondria and involves the conversion of 11-deoxycortisol to cortisol by the enzyme CYPI1B1, 11-hydroxylase. In the zona glomerulosa, 11-hydroxylase may also convert DOC to corticosterone. However, the enzyme CYPI1B2 or aldosterone synthase may also carry out this reaction and, in addition, is required for the conversion of corticosterone to aldosterone through the intermediate 18-OH corticosterone. Thus, CYPI1B2 can carry out 11-hydroxylation, 18-hydroxylation, and 18-methyl oxidation to yield the characteristic C₁₁-18-hemisected structure of aldosterone.

Regulation of Adrenal Steroidogenesis

"Functional Zonation" of the Adrenal Cortex

Glucocorticoids are secreted in relatively high amounts (cortisol 10 to 20 mg/day) from the zona fasciculata under the control of ACTH, and mineralocorticoids are secreted in low amounts (aldosterone 100 to 150 µg/day) from the zona glomerulosa under the principal control of angiotensin II. As a class, adrenal androgens (DHEA, dehydroepiandrosterone sulfate [DHEAS], androstenedione) are the most abundant steroids secreted from the adult adrenal gland (>20 mg/day). In each case, this is facilitated through the expression of steroidogenic enzymes in a specific zonal manner. The zona glomerulosa cannot synthesize cortisol because it does not express 17-hydroxylase. In contrast, aldosterone secretion is confined to the outer zona glomerulosa through the restricted expression of CYPI1B2. Although CYPI1B1 and CYPI1B2 share 95% homology, the 5' promoter sequences differ and permit regulation of the final steps in glucocorticoid and mineralocorticoid biosynthesis by ACTH and angiotensin II, respectively. DHEA is sulfated only in the zona reticularis to form DHEAS.

In the fetal adrenal, steroidogenesis occurs primarily in the inner fetal zone. Because of a relative lack of 3-HSD and high sulfotransferase activity, the principal steroidogenic products are DHEA and DHEAS, which are then aromatized by placental trophoblast to estrogens. Thus, the majority of maternal estrogen across pregnancy is, indirectly, fetally derived. The control and ontogeny of human fetal steroidogenesis are discussed further in Chapter 21.

Classical endocrine feedback loops are in place to control the secretion of both hormones; cortisol inhibits the secretion of both CRF and ACTH from the hypothalamus and pituitary, respectively, and the aldosterone-induced sodium retention inhibits renal renin secretion.
ACTH is the principal hormone stimulating adrenal glucocorticoid biosynthesis and secretion. ACTH has 39 amino acids but is synthesized within the anterior pituitary as part of a much larger 241-amino-acid precursor, pro-opiomelanocortin (POMC). POMC is cleaved in a tissue-specific fashion to yield smaller peptide hormones. In the anterior pituitary this results in the secretion of -lipoprotein and pro-ACTH, the latter being further cleaved to an N-terminal peptide, joining peptide, and ACTH itself (Fig. 14-5). The functions of the N-terminal peptide and -lipoprotein are unknown, although they have weak steroidogenic activity of their own and may augment the effect of ACTH, particularly on stimulating adrenal growth. Indeed, recent studies have cloned a serine protease expressed in the outer adrenal cortex that leaves progamma MSH, releasing shorter fragments that promote adrenal growth.

The first 24 amino acids of ACTH are common to all species, and synthetic ACTH 1 to 24 (Synacthen) is available commercially for clinical testing of the hypothalamic-pituitary-adrenal (HPA) axis and assessment of adrenal glucocorticoid reserve. Melanocyte-stimulating hormones (MSHs, , , and ) are also cleaved products from POMC, but the increased pigmentation characteristic of Addison’s disease is thought to arise directly from increased ACTH concentrations binding to the melanocortin-2 receptor rather than the result of -MSH secretion.

POMC is also transcribed in many extrapituitary tissues, notably brain, liver, kidney, gonad, and placenta. In these normal tissues, POMC messenger ribonucleic acid (mRNA) is usually shorter than the pituitary 1200-bp species.
because of lack of exons 1 and 2 and the 5' region of exon 3. As a result, it is probable that this POMC-like peptide is neither secreted nor active. However, in ectopic ACTH syndrome, additional POMC mRNA species are described that are longer than normal pituitary 1200-bp POMC species (typically 1450 bp) because of the use of alternative promoters in the 5' region of the gene. This may in part explain the resistance of POMC secretion to glucocorticoid feedback in these tumors. Others factors, including interaction with tissue-specific transcription factors and POMC methylation, may explain the ectopic expression of ACTH in some malignant tissues. The cleavage of POMC is also tissue-specific and, at least in some cases of ectopic ACTH syndrome, it is possible that circulating ACTH precursors, notably pro-ACTH, may cross-react in current ACTH radioimmunoassays. The biologic activity of POMC itself upon adrenal function is thought to be negligible.

POMC expression within the hypothalamus and its cleavage appear to be of crucial importance in regulating hair pigmentation and appetite control (see later).

Figure 4C-6 Synthesis and cleavage of pro-opiomelanocortin (POMC) within the human anterior pituitary gland. Prohormone convertase enzymes sequentially cleave POMC to adrenocorticotropic hormone (ACTH). Shaded areas represent melanocyte-stimulating hormone (MSH) structural units. LPH, lipo-protein; LPH, lipo-protein; N-P0C, amino-terminal pro-opiomelanocortin.

Negative Feedback

An important aspect of CRH and ACTH secretion is the negative feedback control exerted by glucocorticoids themselves. Glucocorticoids inhibit POMC gene transcription in the anterior pituitary and CRH and AVP mRNA synthesis and secretion in the hypothalamus. This negative feedback effect is dependent upon the dose, potency, half-life, and duration of administration of the glucocorticoid and has important physiologic and diagnostic consequences. Suppression of the HPA axis by pharmacologic corticosteroids may persist for many months after cessation of therapy, and adrenocortical insufficiency should be anticipated.

The Adrenocorticotropic Hormone Receptor and ACTH Effects on the Adrenal

ACTH binds to a G protein-coupled, melanocortin-2 receptor, of which there are approximately 3500 on each adrenocortical cell. Signal transduction is mediated principally through the stimulation of adenyl cyclase and intracellular cyclic adenosine monophosphate, although both extracellular and intracellular Ca2+ play a role. Other factors synergize with or inhibit the effects of ACTH on the adrenal cortex, including angiotensin II, actinibin, interleukin, and cytokines (tumor necrosis factor and interleukin). Cell-to-cell communication through gap junctions is also important in mediating the effects of ACTH.
Aldosterone production, possibly because of receptor down-regulation or suppression of angiotensin II-stimulated secretion because of a mineralocorticoid effect of aldosterone, is principally by stimulating the early pathways of adrenal steroidogenesis (see earlier), but circulating levels increase by no more than 10% to 20% above baseline.

The effect of ACTH on aldosterone secretion is modest and differs in the acute and chronic situations. An acute bolus of ACTH increases aldosterone secretion, binding of angiotensin II to the surface angiotensin I receptor and activation of phospholipase C. The potassium effect is mediated through membrane depolarization and opening of calcium channels and the angiotensin II effect following calcimodulin kinases. Angiotensin II and potassium stimulate aldosterone secretion principally by increasing the transcription of CYP11B2 through common intracellular signaling pathways.

Prostaglandins also play a role in modulating renin secretion, and indomethacin inhibits renin release. In patients with hypertension, ectopic ACTH syndrome, and renal disease (see Chapter 15), aldosterone is generated through the renin-angiotensin system (Fig. 14-8). Angiotensinogen is an α 2-globulin produced by the liver. Angiotensin I is converted into biologically active angiotensin II by angiotensin-converting enzyme (ACE), mainly in the lung. Angiotensin II increases peripheral vascular resistance, and, together with angiotensin III, stimulates aldosterone (ALDO) secretion, which results in sodium retention and increased plasma volume.

Angiotensin II is generated through the renin-angiotensin system (Figs. 14-8, 14-13). Angiotensinogen is an α 2-globulin synthesized within the liver and is cleaved by renin to form angiotensin I. Angiotensin I is converted to angiotensin II by angiotensin-converting enzyme in lung and many other peripheral tissues; further cleavage to angiotensin III may also occur. Angiotensin I has no apparent biologic activity, but both angiotensin II and angiotensin III are potent in stimulating aldosterone secretion. In addition, angiotensin III is a potent vasconstrictor.

Hypokalemia increases and hyperkalemia decreases renin secretion; in addition, potassium exerts a direct effect on the adrenal cortex to increase aldosterone secretion. The sensitivity of the renin-angiotensin system to changes in circulating potassium is high, with changes in potassium concentrations of only 0.1 to 0.5 mmol/L producing marked changes in aldosterone concentrations. Potassium concentrations also determine the sensitivity of the aldosterone response to a given infusion of angiotensin II, with high potassium intake increasing responsiveness.

Norepinephrine increases renin secretion, and -blockers inhibit renin release. In the clinical assessment of the renin-angiotensin-aldosterone axis, -blockers have a minimal effect on endogenous renin levels and can be given concomitantly if required to control hypertension so that renin and aldosterone concentrations can be measured. Prostaglandins also play a role in modulating renin secretion, and indomethacin inhibits renin release.

Angiotensin II and potassium stimulate aldosterone secretion principally by increasing the transcription of CYP11B2 through common intracellular signaling pathways. Cyclic adenosine monophosphate response elements in the S region of the CYP11B2 gene are activated after an increase in intracellular Ca 2+ and activation of calmodulin kinases. The potassium effect is mediated through membrane depolarization and opening of calcium channels and the angiotensin II effect following binding of angiotensin II to the surface aldosterone I receptor and activation of phospholipase C.

The effect of ACTH on aldosterone secretion is modest and differs in the acute and chronic situations. An acute bolus of ACTH increases aldosterone secretion, principally by stimulating the early pathways of adrenal steroidogenesis (see earlier), but circulating levels increase by no more than 10% to 20% above baseline values. ACTH has no effect on CYP11B2 gene transcription or enzyme activity. Chronic continual ACTH stimulation has either no effect or an inhibitory effect on aldosterone production, possibly because of receptor down-regulation or suppression of angiotensin II-stimulated secretion because of a mineralocorticoid effect of cortisol, DOCC, or angiotensin II. Dopamine and atrial natriuretic peptide inhibit aldosterone secretion, as does heparin.

The separate control of glucocorticoid biosynthesis through the HPA axis and mineralocorticoid synthesis through the renin-angiotensin system has important clinical consequences. Patients with primary adrenal failure invariably have both cortisol and aldosterone deficiency, whereas patients with ACTH deficiency related to
pituitary disease have glucocorticoid deficiency but normal aldosterone concentrations because the renin-angiotensin system is intact.

**Adrenal Androgen Secretion**

Adrenal androgens represent an important component (>50%) of circulating androgens in premenopausal females. In males this contribution is much smaller because of the testicular production of androgens, but adrenal androgen excess even in males may be of clinical significance, notably in patients with CAH. The adult adrenal secretes DHEA at approximately 4 mg/day, DHEAS at 7 to 15 mg/day, androstenedione at 1.5 mg/day, and testosterone at 0.05 mg/day. DHEA is a weak sex steroid but can be converted to androgens and estrogens through the activities of 3-HSD, a superfamily of 17-HSD isozymes, and aromatase, expressed in peripheral target tissues, and this is of clinical importance in many diseases.

ACTH stimulates androgen secretion; DHEA (but not DHEAS because of its increased plasma half-life) and androstenedione demonstrate a similar circadian rhythm to cortisol. However, there are many discrepancies between adrenal androgen and glucocorticoid secretion, which has led to the suggestion of an additional androgen-stimulating hormone (Table 14-4). Many putative androgen-stimulating hormones have been proposed including POMC derivatives such as joining peptide, prolactin, and insulin-like growth factor-I (IGF-I), but conclusive proof is lacking. Adrenal androgen steroidogenesis is dependent upon the relative activities of 3-HSD and 17-hydroxylase and in particular upon the 17.20-lyase activity of 17-hydroxylase. Factors that determine whether 17-hydroxylated substrates, 17-hydroxypregnenolone and 17-OHP, undergo 21-hydroxylation to form glucocorticoid or side-chain cleavage by 17-hydroxylase to form DHEA and androstenedione are unresolved and seem likely to be important in defining the activity of any putative androgen-stimulating hormone.

**TABLE 14-4 -- Dissociation of Adrenal Androgen and Glucocorticoid Secretion: Evidence for an Adrenal-Stimulating Hormone**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Effect on DHEA</th>
<th>Effect on Cortisol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dexamethasone studies: Complete cortisol suppression with chronic high-dose dexamethasone. DHEA falls by only 20%. Greater sensitivity of DHEA to acute low-dose dexamethasone administration.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adrenarche: Rise in circulating DHEA at 68 years of age. Cortisol production unaltered.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aging: Reduction in DHEA production, no change in cortisol.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anorexia nervosa and illness: Fall in DHEA, no change (or increase) in cortisol.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DHEA, dehydroepiandrosterone.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

DHEA, dehydroepiandrosterone.
CORTICOSTEROID HORMONE ACTION

Receptors and Gene Transcription

Both cortisol and aldosterone exert their effects after uptake of free hormone from the circulation and binding to intracellular receptors, termed the glucocorticoid and mineralocorticoid receptors (GR and MR). These are both members of the thyroid/steroid hormone receptor superfamily of transcription factors comprising a C-terminal ligand-binding domain, a central deoxyribonucleic acid (DNA) binding domain interacting with specific DNA sequences on target genes, and an N-terminal hypervariable region. In both cases, although there is only a single gene encoding the GR and MR, splice variants have been described resulting in and variants.

Glucocorticoid hormone action has been studied in more depth than mineralocorticoid action. The binding of steroid to the GR in the cytosol results in activation of the steroid-receptor complex through a process that involves the dissociation of heat shock proteins (HSP 90 and HSP 70). After translocation to the nucleus, gene transcription is stimulated or repressed following binding of dimerized GR-ligand complexes to specific DNA sequences in the promoter regions of target genes. This glucocorticoid response element is invariably a palindromic CGTACAnnnTGTACT sequence that binds with high affinity to two loops of DNA within the DNA binding domain of the GR (zinc fingers). This stabilizes the RNA polymerase II complex, facilitating gene transcription. The GR variant may act as a dominant negative regulator of GR transactivation.

Naturally occurring mutations in the GR (as seen in patients with glucocorticoid resistance, discussed later) and GR variants generated in vitro have highlighted critical regions of the receptor responsible for binding and transactivation. But numerous others factors are also required (coactivators, corepressors) that may confer tissue specificity of response. This is a rapidly evolving field and is reviewed in Chapter 4. However, the interactions between GR and two particular transcription factors are important in mediating the anti-inflammatory effects of glucocorticoids and explain the effect of glucocorticoids on genes that do not contain obvious glucocorticoid response elements in their promoter regions. Activator protein-1 (AP-1) comprises Fos and Jun subunits and is a proinflammatory transcription factor induced by a series of cytokines and phorbol ester. The GR-ligand complex can bind to Jun and prevent interaction with the AP-1 site to repress AP-1 and GR trans-activation functions. Similarly, functional antagonism exists between the GR and nuclear factor B (NF-B). NF-B is a ubiquitously expressed transcription factor that activates a series of genes involved in lymphocyte development, inflammatory response, host defense, and apoptosis. In keeping with the diverse array of actions of cortisol, many hundred glucocorticoid-responsive genes have been identified. Some glucocorticoid-induced genes and repressed genes are given in Table 4-1.

In contrast to the diverse actions of glucocorticoids, mineralocorticoids have a more restricted role, principally to stimulate epithelial sodium transport in the distal nephron, distal colon, and salivary glands. This stimulation is mediated through the induction of the apical sodium channel (comprising three subunits, and ) and the , and subunits of the basolateral Na+. K+-adenosine triphosphatase through transcriptional regulation of an aldosterone-induced gene, serum and glucocorticoid-induced kinase (SGK). Aldosterone binds to the MR, principally in the cytosol (although there is evidence for expression of the unliganded MR in the nucleus), followed by translocation of the hormone-receptor complex to the nucleus through the "prereceptor" metabolism of cortisol through the enzyme 11-HSD, which inactivates cortisol and corticosterone to inactive 11-keto metabolites, enabling aldosterone to bind to the MR.

For both glucocorticoids and mineralocorticoids there is accumulating evidence for so-called nongenomic effects involving hormone responses obviating the genomic GR or MR. A series of responses have been described within seconds or minutes of exposure to corticosteroids and are thought to be mediated by as yet uncharacterized membrane-coupled receptors.
Cortisol-Binding Globulin and Corticosteroid Hormone Metabolism

Over 90% of circulating cortisol is bound, predominantly to the \( \alpha \)-globulin cortisol-binding globulin (CBG). This 383-amino-acid protein is synthesized in the liver and binds cortisol with high affinity. Affinity for synthetic corticosteroids (except prednisolone, which has an affinity for CBG about 50% of that of cortisol) is negligible. Circulating CBG concentrations are approximately 26 mg/dL (700 nmol/L); levels are increased by estrogens and in some patients with chronic active hepatitis but reduced by glucocorticoids and in patients with cirrhosis, nephrosis, and hyperthyroidism. The estrogen effect can be marked, with levels increasing twofold to threefold during pregnancy, and this should also be taken into account when measuring plasma "total" cortisol in pregnancy and in women taking estrogens.

<table>
<thead>
<tr>
<th>Site of Action</th>
<th>Induced Genes</th>
<th>Repressed Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immune system</td>
<td>IB (NFκB inhibitor)</td>
<td>Interleukins</td>
</tr>
<tr>
<td></td>
<td>Haptoglobin</td>
<td>TNF-</td>
</tr>
<tr>
<td></td>
<td>TCR</td>
<td>IFN-</td>
</tr>
<tr>
<td></td>
<td>( p21, p27, ) and ( p57 )</td>
<td>E-selectin</td>
</tr>
<tr>
<td></td>
<td>Lipocortin</td>
<td>ICAM-1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cyclooxygenase 2</td>
</tr>
<tr>
<td>Metabolic</td>
<td>PPAR-</td>
<td>Tryptophan hydroxylase</td>
</tr>
<tr>
<td></td>
<td>Tyrosine aminotransferase</td>
<td>Metalloprotease</td>
</tr>
<tr>
<td></td>
<td>Glutamine synthase</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Glycogen synthase</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Glucose-6-phosphatase</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PEPCK</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Leptin</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \alpha )-Fibrinogen</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cholesterol 7-hydroxylase</td>
<td></td>
</tr>
<tr>
<td>Bone</td>
<td>C/EBP/</td>
<td>Osteocalcin</td>
</tr>
<tr>
<td></td>
<td>Androgen receptor</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Calcitonin receptor</td>
<td>Collagenase</td>
</tr>
<tr>
<td></td>
<td>Alkaline phosphatase</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \text{GF-BP-6} )</td>
<td></td>
</tr>
<tr>
<td>Channels and transporters</td>
<td>Epithelial sodium channel (ENaC) ( \ldots )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Serine and glucocorticoidinduced kinase (SGK)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aquaporin 1</td>
<td></td>
</tr>
<tr>
<td>Endocrine</td>
<td>bFGF</td>
<td>GR</td>
</tr>
<tr>
<td></td>
<td>VIP</td>
<td>PRL</td>
</tr>
<tr>
<td></td>
<td>Endothelin</td>
<td>POMC/CRH</td>
</tr>
<tr>
<td></td>
<td>RXR</td>
<td>PTHrP</td>
</tr>
<tr>
<td></td>
<td>GHRH receptor</td>
<td>Vasopressin</td>
</tr>
<tr>
<td></td>
<td>( \text{Na+} ) -dependent peptide receptors</td>
<td></td>
</tr>
<tr>
<td>Growth and development</td>
<td>Surfactant protein A, B, C</td>
<td>Fibronectin</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \text{Fetoprotein} )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NGF</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Erythropoietin</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G1 cyclins</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cyclin-dependent kinases</td>
</tr>
</tbody>
</table>


bFGF, basic fibroblast growth factor; CRH, corticotropin-releasing hormone; C/EBP/\( \alpha \), CAAT-enhancer binding protein-beta; GR, glucocorticoid receptor; GHRH, growth hormonereleasing hormone; ICAM, intercellular adhesion molecule; IFN, interferon; IGF-BP, insulin-like growth factorbinding protein; IB, inhibitory kappa B; iNOS, inducible nitric oxide synthase; NFB, nuclear factor B; NGF, nerve growth factor; PEPCK, phosphoenolpyruvate carboxykinase; POMC, proopiomelanocortin; PPAR, peroxisome proliferatoractivated receptor; PTHrP, parathyroid hormone-related protein; RXR, retinoid X receptor; SGK, serum and glucocorticoid-induced kinase; TCR, T-cell receptor; TNF-, tumor necrosis factor-alpha; VIP, vasoactive intestinal peptide.

Inherited abnormalities in CBG synthesis are much rarer than those described for thyroid-binding globulin but include elevated CBG, partial and complete deficiency of CBG, or CBG variants with reduced affinity for cortisol. In each case, alterations in CBG concentrations change total circulating cortisol concentrations accordingly but free cortisol concentrations are normal. Only this free circulating fraction is available for transport into tissues for biologic activity. The free cortisol excreted through the kidneys is termed urinary free cortisol and represents only 1% of the total cortisol secretion rate.

The circulating half-life of cortisol varies between 70 and 120 minutes. The major steps for cortisol metabolism are depicted in **Figure 14-12**. These
The interconversion of the 11-hydroxyl (cortisol, Kendall's compound F) to the 11-oxo group (cortisone, compound E) through the activity of 11-HSD (EC 1.1.1.146). The metabolism of cortisol and cortisone then follow similar pathways.

2. Reduction of the C4-C5 double bond to form dihydrocortisol or dihydrocortisone followed by hydroxylation of the 3-oxo group to form tetrahydrocortisol (THF) and tetrahydrocortisone (THE). The reduction of the C4-C5 double bond can be carried out by either 5-reductase or 5-reductase to yield, respectively, 5-THF (THF) or 5-THF (allo-THF). In normal subjects the 5 metabolites predominate (5:5-THF 2:1). THF, allo-THF, and THE are rapidly conjugated with glucuronic acid and excreted in the urine.

Further reduction of the 20-oxo group by either 20HSD or 20-HSD to yield - and -cortols and cortolones from cortisol and cortisone, respectively. Reduction of the C20 position may also occur without A ring reduction, giving rise to 20-hydroxycortisol and 20-hydroxycortisone.

4. Hydroxylation at C6 to form 6-hydroxycortisol.

Cleavage of THF and THE to the C21 hydroxyl group followed by either 20HSD or 20-HSD to yield - and -cortols and cortolones from cortisol and cortisone, respectively.

Further reduction of the 20-oxo group by either 20HSD or 20-HSD to yield - and -cortols and cortolones from cortisol and cortisone, respectively. Reduction of the C20 position may also occur without A ring reduction, giving rise to 20-hydroxycortisol and 20-hydroxycortisol.

6. Oxidation of the C21 position of cortols and cortolones to form the extremely polar metabolites cortolic and cortolonic acids.

Approximately 50% of secreted cortisol appears in the urine as THF, allo-THF, and THE; 25% as cortols and cortolones; 10% as C11 steroids; and 10% as corticotropic and corticosterone acids. The remaining metabolites are free, unconjugated steroids (cortisol, cortisone, 6- and 20-20-metabolites of THF and THE).

The principal site of cortisol metabolism has been considered to be the liver, but many of the preceding enzymes have been described in the mammalian kidney, notably in the inactivation of cortisol to cortisone by 11-HSD. Quantitatively, the interconversion of cortisol to cortisone by 11-HSD is also the most important pathway. Furthermore, the bioactivity of glucocorticoids is in part related to the hydroxyl group at C11; cortisone with a C11 oxo group is an inactive steroid so that 11-HSD expressed in peripheral tissues plays a crucial role in regulating corticosteroid hormone action. Two distinct 11-HSD isozymes have been reported: a type 1 oxoreductase dependent on reduced nicotinamide adenine dinucleotide phosphate and expressed principally in the liver, which confers bioactivity upon orally administered cortisone by converting it to cortisol, and a type 2, nicotinamide adenine dinucleotide-dependent dehydrogenase. It is 11-HSD2, coexpressed with the MR in the kidney, colon, and salivary gland, that inactivates cortisol to cortisone and permits aldosterone to bind to the MR in vivo. If this enzyme-protective mechanism is impaired, cortisol is able to act as a mineralocorticoid; this explains some forms of endocrine hypertension (apparent mineralocorticoid excess, licorice ingestion; see Chapter 15) and the mineralocorticoid excess state that characterizes the ectopic ACTH syndrome (see later).

Hyperthyroidism results in increased cortisol metabolism and clearance and hypothyroidism the converse, principally because of an effect of thyroid hormone on hepatic 11-HSD and 5/5-reductases. IGF-I increases cortisol clearance by inhibiting hepatic 11-HSD (conversion of cortisone to cortisol), and 6-Hydroxylation is normally a minor pathway, but cortisol itself induces 6-hydroxylation so that 6-hydroxycortisol excretion is markedly increased in patients with Cushing's syndrome. Furthermore, some drugs, notably rifampicin and phenytoin, increase cortisol clearance through this pathway. Patients with renal disease have impaired cortisol clearance because of reduced renal conversion of cortisol to cortisone. These observations have clinical implications for patients with thyroid disease, acromegaly, and renal disease and for patients taking cortisol replacement therapy. Adrenal crisis has been reported in steroid-replaced Addisonian patients given rifampicin, and hydrocortisone replacement therapy may need to be increased in treated patients in whom hyperthyroidism develops.

Aldosterone is also metabolized in the liver and kidneys. In the liver it undergoes tetrahydro reduction and is excreted in the urine as a 3-glucuronide tetrahydroaldosterone derivative; however, glucuronide conjugation at the 18 position occurs directly in the kidney, as does 3 and 5/5 metabolism of the free steroid. Because of the aldehyde group at the C18 position, aldosterone is not metabolized by 11-HSD. Hepatic aldosterone clearance is reduced in patients with cirrhosis, ascites, and severe congestive heart failure.
Effects of Glucocorticoids (Fig. 14-13)

Carbohydrate, Protein, and Lipid Metabolism

Glucocorticoids increase blood glucose concentrations through their action on glycogen, protein, and lipid metabolism. In the liver, cortisol stimulates glycogen deposition by increasing glycogen synthase and inhibiting the glycogen-mobilizing enzyme glycogen phosphorylase. Hepatic glucose output increases through the activation of key enzymes involved in gluconeogenesis, principally glucose-6-phosphatase and phosphoenolpyruvate carboxykinase.

In peripheral tissues (muscle, fat), cortisol inhibits glucose uptake and utilization. In adipose tissue lipolysis is activated, resulting in the release of free fatty acids into the circulation. An increase in total circulating cholesterol and triglycerides is observed, but HDL cholesterol levels fall. Glucocorticoids also have a permissive effect on other hormones including catecholamines and glucagon. The resultant effect is to cause insulin resistance and an increase in blood glucose concentrations at the expense of protein and lipid catabolism.

Glucocorticoids stimulate adipocyte differentiation, promoting adipogenesis through the transcriptional activation of key differentiation genes including lipoprotein lipase, glycerol-3-phosphate dehydrogenase, and leptin. Chronically, the effects of glucocorticoid excess on adipose tissue are more complex, at least in humans, in whom the deposition of visceral or central adipose tissue is stimulated, providing a useful discriminatory sign for the diagnosis of Cushing's syndrome. The explanation for the predilection for visceral obesity may be related to the increased expression of both the GR and type 1 isozyme of 11-HSD (generating cortisol from cortisone) in omental compared with subcutaneous adipose tissue.

Skin, Muscle, and Connective Tissue

In addition to inducing insulin resistance in muscle tissue, glucocorticoids cause catabolic changes in muscle, skin, and connective tissue. In the skin and connective tissue, glucocorticoids inhibit epidermal cell division and DNA synthesis and reduce collagen synthesis and production. In muscle, glucocorticoids cause atrophy (but not necrosis), and this seems to be specific for type II or "phasic" muscle fibers. Muscle protein synthesis is reduced.

Salt and Water Homeostasis and Blood Pressure Control

Glucocorticoids increase blood pressure by a variety of mechanisms involving actions on the kidney and vasculature. In vascular smooth muscle they increase sensitivity to pressor agents such as catecholamines and angiotensin II while reducing nitric oxide-mediated endothelial dilation. In the kidney, depending on the activity of the type 2 isozyme of 11-HSD, cortisol can act on the distal nephron to cause sodium retention and potassium loss (mediated by the MR). Elsewhere across the nephron, glucocorticoids increase glomerular filtration rate, proximal tubular epithelial sodium transport, and free water clearance. The last effect involves antagonism of the action of vasopressin and explains the dilutional hyponatremia seen in patients with glucocorticoid deficiency.

Anti-Inflammatory Actions and the Immune System

Glucocorticoids suppress immunologic responses, and this has been the stimulus to develop a series of highly potent pharmacologic glucocorticoids to treat a variety of autoimmune and inflammatory conditions. The inhibitory effects are mediated at many levels. In the peripheral blood, glucocorticoids reduce lymphocyte counts acutely (T lymphocytes > B lymphocytes) by redistributing lymphocytes from the intravascular compartment to spleen, lymph nodes, and bone marrow. Conversely, neutrophil counts increase after glucocorticoid administration. Eosinophil counts fall rapidly, an effect that was used historically as a bioassay for glucocorticoids.

The immunologic actions of glucocorticoids involve direct actions on both T and B lymphocytes that include inhibition of immunoglobulin synthesis and stimulation of lymphocyte apoptosis. Inhibition of cytokine production from lymphocytes is mediated through inhibition of the action of NF-B. NF-B plays a crucial and generalized role in inducing cytokine gene transcription; glucocorticoids can bind directly to NF-B to prevent nuclear translocation and can induce NF-B inhibitor, which sequesters NF-B in the cytoplasm, thereby inactivating its effect.

Additional anti-inflammatory effects involve inhibition of monocyte differentiation into macrophages and macrophage phagocytosis and cytotoxic activity.
Glucocorticoids reduce the local inflammatory response by preventing the action of histamine and plasminogen activators. Prostaglandin synthesis is impaired through the induction of lipocortins, which inhibit phospholipase A₂ activity. Prostaglandin synthesis is impaired through the induction of lipocortins, which inhibit phospholipase A₂ activity.

Central Nervous System and Mood

Clinical observations of patients with glucocorticoid excess and deficiency reveal that the brain is an important target tissue for glucocorticoids, with depression, euphoria, psychosis, apathy, and lethargy being important manifestations (see the following). Both glucocorticoid and mineralocorticoid receptors are expressed in discrete regions of the rodent brain including hippocampus, hypothalamus, cerebellum, and cortex. Glucocorticoids cause neuronal death notably in the hippocampus, and this may underlie the interest in glucocorticoids and cognitive function, memory, and neurodegenerative diseases such as Alzheimer’s. In the eye, glucocorticoids act to raise intraocular pressure through an increase in aqueous humor production and deposition of matrix within the trabecular meshwork, which inhibits aqueous drainage. Steroid-induced glaucoma appears to have a genetic predisposition, but the underlying mechanisms are unknown.

Gut

Chronic but not acute administration of glucocorticoids increases the risk of developing peptic ulcer disease. Pancreatitits with fat necrosis is reported in patients with glucocorticoid excess. The GR is expressed throughout the gastrointestinal tract and the MR in the distal colon, and these mediate the corticosteroid control of epithelial ion transport.

Endocrine: Replacement therapy (Addison’s disease, pituitary disease, congenital adrenal hyperplasia), Graves’ ophthalmopathy

Skin: Dermatitis, pemphigus

Hematology: Leukemia, lymphoma, hemolytic anemia, idiopathic thrombocytopenic purpura

Gastrointestinal: Inflammatory bowel disease (Ulcerative colitis, Crohn’s disease)

Liver: Chronic active hepatitis, transplantation, organ rejection

Renal: Nephrotic syndrome, vasculitides, transplantation, rejection

Central nervous system: Cerebral edema, raised intracranial pressure

Respiratory: Angioedema, anaphylaxis, asthma, sarcoidosis, tuberculosis, obstructive airway disease

Rheumatology: Systemic lupus erythematosus, polyarteritis, temporal arteritis, rheumatoid arthritis

Muscle: polymyalgia rheumatica, myasthenia gravis

excess. The GR is expressed throughout the gastrointestinal tract and the MR in the distal colon, and these mediate the corticosteroid control of epithelial ion transport.

Growth and Development

Although glucocorticoids stimulate growth hormone (GH) gene transcription in vitro, glucocorticoids in excess inhibit linear skeletal growth, probably as a result of catabolic effects on connective tissue, muscle, and bone and inhibition of the effects of IGF-I. Experiments on mice lacking the GR gene emphasize the role of glucocorticoids in normal fetal development. In particular, glucocorticoids stimulate lung maturation through the synthesis of surfactant proteins (SP-A, SP-B, SP-C) and mice lacking the GR die shortly after birth as a result of hypoxia from lung atelectasis. Glucocorticoids also stimulate the enzyme phenylethanolamine N-methyltransferase, which converts norepinephrine to epinephrine in adrenal medulla and chromaffin tissue. Mice lacking the GR do not develop an adrenal medulla.

Endocrine Effects

Glucocorticoids suppress the thyroid axis, probably through a direct action on thyroid-stimulating hormone (TSH) secretion. In addition, they inhibit 5’ deiodinase activity, mediating the conversion of thyroxine to active triiodothyronine. Glucocorticoids also act centrally to inhibit gonadotropin-releasing hormone pulsatility and release of luteinizing hormone (LH) and follicle-stimulating hormone (FSH) (see later).
THERAPEUTIC CORTICOSTEROIDS

Since the dramatic anti-inflammatory effect of cortisol was first demonstrated in the 1950s, a series of synthetic corticosteroids have been developed for therapeutic purposes. These are used to treat a diverse variety of human diseases and rely principally on their anti-inflammatory and immunologic actions (Table 14-6). The main corticosteroids used in clinical practice, together with their relative glucocorticoid and mineralocorticoid potencies, are listed in Table 14-7.

The structures of common synthetic steroids are depicted in Figure 14-14. Biologic activity of a corticosteroid is dependent upon a 4-3-keto, 11-hydroxy, 17,21-trihydroxyl configuration. Conversion of the C11 hydroxyl group to a C11 keto group (cortisol to cortisone) inactivates the steroid. The addition of a 1,2 unsaturated bond to cortisol results in prednisolone, which is four times more potent than cortisol in classical glucocorticoid bioassays such as hepatic glycogen deposition, suppression of eosinophils, and anti-inflammatory actions. Prednisone is the "cortisone equivalent" of prednisolone and relies upon conversion by 11-HSD type 1 in the liver for bioactivity. Potency is further increased by the addition of a 6-methyl group to prednisolone (methylprednisolone). Fludrocortisone is a synthetic mineralocorticoid that has 125-fold greater potency than cortisol in stimulating sodium reabsorption. This is achieved through the addition of a 9-fluoro group to cortisol. Interestingly, fludrocortisone also has glucocorticoid potency (12-fold greater than cortisol), and the addition of a 16-methyl group and 1,2 saturated bond to fludrocortisone results in dexamethasone, a highly potent glucocorticoid (25-fold greater than cortisol) but with negligible mineralocorticoid activity. Betamethasone has the same structure but with a 16-methyl group and is widely used in respiratory and nasal aerosol sprays.

<table>
<thead>
<tr>
<th>Steroid</th>
<th>Anti-inflammatory Action</th>
<th>Hypothalamic-Pituitary-Adrenal Suppression</th>
<th>Salt Retention</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cortisol</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Prednisolone</td>
<td>3</td>
<td>4</td>
<td>0.75</td>
</tr>
<tr>
<td>Methylprednisolone</td>
<td>6.2</td>
<td>4</td>
<td>0.5</td>
</tr>
<tr>
<td>Fludrocortisone</td>
<td>12</td>
<td>12</td>
<td>125</td>
</tr>
<tr>
<td>^{1}Fludrocortisone</td>
<td>14</td>
<td>4</td>
<td>225</td>
</tr>
<tr>
<td>Triamcinolone</td>
<td>5</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Dexamethasone</td>
<td>26</td>
<td>17</td>
<td>0</td>
</tr>
</tbody>
</table>

Corticosteroids are given orally, parenterally, and by numerous topical routes (e.g., eyes, skin, nose, inhalation, rectal suppositories). Unlike hydrocortisone, which has a high affinity for CBG, most synthetic steroids have low affinity for this binding protein and circulate as free steroid (30%) or bound to albumin (70%). Circulating half-lives vary depending upon individual variability and underlying disease, particularly renal and hepatic impairment. Cortisone acetate should not be used parenterally as it requires metabolism by the liver to active cortisol.

It is beyond the consideration of this chapter to describe which steroid should be given and by which route for the nonendocrine conditions listed in Table 14-6. The acute and chronic administration of corticosteroid therapy in patients with hypoadrenalism and CAH is discussed in these sections. In addition to the undoubted benefit that corticosteroids provide, there is increasingly a misuse of corticosteroid therapy, particularly in patients with respiratory or rheumatologic disease, to such an extent that up to 0.5% of the population is now prescribed chronic corticosteroid therapy. Because of their established euphoric effect, corticosteroids often make patients feel better but without any objective measures of improvements in underlying disease parameters. In view of the long-term sequelae of chronic glucocorticoid excess, decisions regarding treatment should be evidence-based and subject to constant review for efficacy and side effects. The endocrinologic consequences, notably suppression of the HPA axis, are an important aspect of modern clinical practice. Endocrinologists need to be aware of the effects of chronic therapy and advise on steroid withdrawal.

**Chronic Corticosteroid Therapy, Hypothalamic-Pituitary-Adrenal Axis Suppression, and Steroid Withdrawal**

The negative feedback control of the HPA axis by endogenous cortisol has already been detailed. Synthetic corticosteroids similarly suppress the function of the HPA axis through a process that is dependent on both dose and duration of treatment. As a result, sudden cessation of corticosteroid therapy may result in adrenal failure. This may also occur after treatment with high doses of the synthetic progestagen medroxyprogesterone acetate, which possesses glucocorticoid agonist activity in patients taking any steroid dose for less than 3 weeks, clinically significant suppression of the HPA axis is rarely a problem and patients can withdraw from steroids suddenly with no ill effect. The possible exception to this is the patient who receives frequent short courses of corticosteroid therapy; for example, patients with recurrent episodes of severe asthma. Conversely, suppression of the HPA axis is invariably in patients taking the equivalent of 15 mg or more of prednisolone.

<table>
<thead>
<tr>
<th>Dose (mg pred/day)</th>
<th>3 wk</th>
<th>&gt;3 wk</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.5 mg</td>
<td>Can stop</td>
<td>Reduce rapidly e.g., 2.5 mg every 34 days</td>
</tr>
</tbody>
</table>

**TABLE 14-8 -- Duration of Glucocorticoid Treatment**
per day chronically. In patients taking lower doses of corticosteroid chronically (prednisolone 5 to 15 mg/day or equivalent), suppression of the HPA axis is variable. Defects in response of the HPA axis to insulin-induced hypoglycemia or exogenous ACTH have been reported in patients taking prednisolone doses as low as 5 mg/day, but clinically significant suppression at these doses is debatable. Alternate-day therapy is associated with less suppression of the HPA axis.

All patients treated chronically with corticosteroids should be treated in a similar fashion to patients with chronic ACTH deficiency; they should carry steroid cards and be offered Steroid Alert bracelets or necklaces. In the event of an intercurrent stress (infection, surgery), supplemental steroid cover should be given, equivalent to hydrocortisone at 100 to 150 mg/day. If the patient is unable to take drugs orally, parenteral therapy is required. Recovery from suppression may take 6 to 9 months. CRH secretion returns to normal, and within a few weeks ACTH levels begin to increase and indeed rise above normal values until adrenal steroidogenesis recovers. In the interim, and without replacement therapy, patients may experience symptoms of glucocorticoid deficiency including anorexia, nausea, weight loss, arthralgia, lethargy, skin desquamation and postural dizziness (see Glucocorticoid Insufficiency). To avoid symptoms of glucocorticoid deficiency, steroids should be withdrawn cautiously over a period of months. Assuming the underlying disease permits steroid reduction, doses should be reduced from pharmacologic levels to physiologic levels (equivalent to prednisolone at 7.5 mg/day) over a few weeks. Thereafter doses should be reduced by 1 mg/day prednisolone every 2 to 4 weeks depending on the patient's well-being. An alternative approach is to change to hydrocortisone at 20 mg/day and reduce the daily dose by 2.5 mg/day every week to 10 mg/day. Doses at nighttime should be avoided as they result in greater suppression of early morning ACTH secretion. After 2 to 3 months of these reduced doses of corticosteroids, endogenous function of the HPA axis can be assessed through a corticotropin (ACTH, Synacthen) stimulation test or an insulin-induced hypoglycemia test (see later). A pass response to these tests indicates adequacy of function of the HPA axis, and corticosteroid therapy can be safely withdrawn. In patients taking physiologic doses of prednisolone (less than 5 to 7.5 mg/day) or equivalent, a corticotropin stimulation test 12 to 24 hours after having omitted steroid therapy will provide an immediate answer on whether sudden or gradual withdrawal of steroid therapy is indicated (Table 14-8).

Iatrogenically induced Cushing's syndrome occurs in patients taking suppressive doses of corticosteroids for more than 3 weeks. The rapidity of onset of clinical features is dependent upon the administered dose but can occur within 1 month of therapy.

**TABLE 14-9 -- Adrenocortical Diseases**

<table>
<thead>
<tr>
<th>Glucocorticoid Excess</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cushing's syndrome</td>
</tr>
<tr>
<td>Pseudo-Cushing's syndromes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Glucocorticoid Resistance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucocorticoid Deficiency</td>
</tr>
<tr>
<td>Primary hypoaldrenism</td>
</tr>
<tr>
<td>Secondary hypoaldrenism</td>
</tr>
<tr>
<td>Post-chronic corticosteroid replacement therapy</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Congenital Adrenal Hyperplasia</th>
</tr>
</thead>
<tbody>
<tr>
<td>21-Hydroxylase, 3-hydroxysteroid dehydrogenase, 17-hydroxylase, 11-hydroxylase, and STAIR deficiencies</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mineralocorticoid Excess</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mineralocorticoid Deficiency</td>
</tr>
<tr>
<td>Defects in aldosterone synthesis</td>
</tr>
<tr>
<td>Defects in aldosterone action</td>
</tr>
<tr>
<td>Hyporeninemic hypoaldosteronism</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Adrenal Incidentalomas, Adenomas, and Carcinomas</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAIR, steroidogenic acute regulatory (protein).</td>
</tr>
</tbody>
</table>

Pred, prednisolone; SST, short Synacthen test; ITT, insulin tolerance test.

* Beware frequent steroid courses, e.g., in asthma.
ADRENOCORTICAL DISEASES

Adrenocortical diseases are relatively rare, but their importance lies in their morbidity and mortality if untreated coupled with the relative ease of diagnosis and the availability of effective therapy. The diseases are most readily classified on the basis of whether there is hormone excess or deficiency (Table 14-9).

Glucocorticoid Excess

In 1912 Harvey Cushing first described a 23-year-old woman with obesity, hirsutism, and amenorrhea and 20 years later postulated that this “polyglandular syndrome” was due to a primary pituitary abnormality causing adrenal hyperplasia. Adrenal tumors were shown to cause the syndrome in some cases but ectopic ACTH production was not characterized until much later in 1962. The term Cushing’s syndrome is used to describe all causes, and Cushing’s disease is reserved for cases of pituitary-dependent Cushing’s syndrome.

Cushing’s syndrome comprises the symptoms and signs associated with prolonged exposure to inappropriately elevated levels of free plasma glucocorticoids (Fig. 14-15). The term glucocorticoid in the definition covers both endogenous (cortisol) and exogenous (e.g., prednisolone, dexamethasone) excess (Table 14-10). Ectopic Cushing’s syndrome is most common occurring to some degree in the majority of patients taking chronic corticosteroid therapy. Endogenous causes of Cushing’s syndrome result in loss of the normal feedback mechanism of the HPA axis and the normal circadian rhythm of cortisol secretion and are rare.

The incidence of pituitary-dependent Cushing’s syndrome is estimated to be 5 to 10 cases per million population per year. The incidence of ectopic ACTH syndrome parallels that of bronchogenic carcinoma, and although 0.5% of lung cancer patients have ectopic ACTH syndrome, the rapid progression of the underlying disease often precludes an early diagnosis. Cushing’s disease and adrenal adenomas are four times commoner in women, and ectopic ACTH syndrome is commoner in men.

Clinical Features of Cushing’s Syndrome

The classical features of Cushing’s syndrome with centripetal obesity, moon face, hirsutism, and plethora are well known.

Table 14-10 – Classification of Causes of Cushing’s Syndrome

<table>
<thead>
<tr>
<th>ACTH-Dependent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cushing’s disease (pituitary-dependent)</td>
</tr>
<tr>
<td>Ectopic ACTH syndrome</td>
</tr>
<tr>
<td>Ectopic CRH syndrome</td>
</tr>
<tr>
<td>Macronodular adrenal hyperplasia</td>
</tr>
<tr>
<td>Iatrogenic (treatment with ACTH 124)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ACTH-Independent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adrenal adenoma and carcinoma</td>
</tr>
<tr>
<td>Primary pigmented nodular adrenal hyperplasia and Carney’s syndrome.</td>
</tr>
<tr>
<td>McCune-Albright syndrome</td>
</tr>
<tr>
<td>Aberrant receptor expression (gastric inhibitory polypeptide, interleukin-1).</td>
</tr>
<tr>
<td>Iatrogenic (e.g., pharmacologic doses of prednisolone, dexamethasone)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pseudo-Cushing’s Syndromes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcoholism</td>
</tr>
<tr>
<td>Depression</td>
</tr>
<tr>
<td>Obesity</td>
</tr>
<tr>
<td>ACTH, adrenocorticotropic hormone; CRH, corticotropin-releasing hormone.</td>
</tr>
</tbody>
</table>

following Cushing’s initial descriptions in 1912 and 1932 (Fig. 14-15 and Fig. 14-16). However, this gross clinical picture is not always present and a high index of suspicion is required in many cases. When the normal physiologic effects of glucocorticoids are appreciated (see Fig. 14-13), the clinical features of glucocorticoid excess are easier to define. These are summarized in Table 14-11 together with the most discriminatory features that assist in distinguishing Cushing’s syndrome from simple obesity.

Obesity

Weight gain and obesity are the commonest sign, and at least in adults this is invariably centripetal in nature. Indeed, generalized obesity is commoner in the general population than in patients with Cushing’s syndrome. One exception to this is childhood, in which glucocorticoid excess may result in generalized obesity. In addition to centripetal obesity, patients develop fat depots over the thoracocervical spine (“buffalo hump”), in the supraclavicular region, and over the cheeks and temporal regions, giving rise to the rounded “moon-like” facies. The epicardial space is another site of abnormal fat deposition, and this may lead to neurologic deficits.

Reproductive Dysfunction

Gonadal dysfunction is common, with menstrual irregularity in females and loss of libido in both sexes. Hirsutism is frequently found in female patients, as is acne. The commonest form of hirsutism is vellous hypertrichosis on the face and should be distinguished from darker terminal differentiated hirsutism, which may...
Muscle weakness

Backache

Fractures

Loss of scalp hair

Obesity

Truncal

Generalized

Plethora

Moon face

Hypertension

Bruising

Red-purple striae

Muscle weakness

Psychiatric Abnormalities

Psychiatric abnormalities occur in approximately 50% of patients with Cushing's syndrome regardless of cause. Agitated depression and lethargy are among the commonest problems, but paranoia and overt psychosis are also well recognized. Memory and cognitive function may also be affected, and increased irritability may be an early feature. Insomnia is common, and both rapid eye movement and delta wave sleep patterns are reduced. Lowering of plasma cortisol by medical or surgical therapy usually results in a rapid improvement in the psychiatric state.

Bone

In childhood the commonest presentation is with poor linear growth and weight gain; as discussed earlier, glucocorticoids have profound effects on growth and development. Many patients with long-standing Cushing's syndrome have lost height because of osteoporotic vertebral collapse. This can be assessed by measuring the patient's height and comparing it with the patient's span; in normal subjects these measurements should be equal. Pathologic fractures, either spontaneous or after minor trauma, are not uncommon. Rib fractures, in contrast to those of the vertebral, are often painless. The radiographic appearances are typical, with exuberant callus formation at the site of the healing fracture. In addition, aseptic necrosis of the femoral and humeral heads, a recognized feature of high-dose exogenous corticosteroid therapy, can occur in endogenous Cushing's syndrome (Fig. 14-17) (Figure Not Available). Hypercalciuria may lead to renal calculi, but hypercalcemia is not a feature.

Skin

Hypercortisolism results in skin thinning, separation, and exposure of the subcutaneous vascular tissue. On examination, wrinkling of the skin on the dorsum of the hand may be seen resulting in a "cigarette paper" appearance (Liddle's sign). Minimal trauma may result in bruising, which frequently resembles the appearance of "senile purpura." The plethoric appearance of the patient with Cushing's syndrome is secondary to the thinning of the skin combined with loss of facial subcutaneous fat and is not due to true polycythemia. Acne and papular lesions may occur over the face, chest, and back.

The typical, almost pathognomonic red-purple livid striae greater than 1 cm in diameter are most frequently found on the abdomen but may also be present on the upper thighs, breasts, and arms. They are common in younger patients and less so in those older than 50 years. They must be differentiated from the paler, less pigmented striae that occur post partum (striae gravidarum) or in association with rapid weight loss.

Increased skin pigmentation is rare in Cushing's disease but common in the ectopic ACTH syndrome and arises because of overstimulation of melanocyte receptors by ACTH.

Myopathy and bruising are two of the most discriminatory features of the syndrome. The myopathy of Cushing's syndrome

| TABLE 14-11 | Prevalence of Symptoms and Signs in Cushing's Syndrome and Discriminant Index Compared with Prevalence of Features in Patients with Simple Obesity |
|-------------|-------------------------------------------------|----------------|----------------|
| **Findings** | **%**               | **Discriminant Index** | **%**               | **Discriminant Index** |
| Symptoms    |                     |                     | Signs            |                     |
| Weight gain | 91                  |                     | Obesity          | 97                  |
| Menstrual irregularity | 84 | 1.6 | Truncal | 46 | 1.6 |
| Hirsutism   | 81                  | 2.8                 | Generalized      | 55                  | 0.8 |
| Psychiatric dysfunction | 62 |     | Plethora      | 94                  | 3.0 |
| Backache    | 43                  |                     | Moon face        | 88                  |
| Muscle weakness | 29 | 8.0 | Hypertension  | 74                  | 4.4 |
| Fractures   | 19                  |                     | Bruising         | 82                  | 10.3 |
| Loss of scalp hair | 13 |     | Red-purple striae | 56                  | 2.5 |
| Muscle weakness | 50 |     | Muscle weakness | 50                  |     |
Lack of "cure" after pituitary stalk section

Pituitary Theory

[Note: The full context and discussion around the pituitary stalk section and its implications are not provided in the given text.]

Etiology.

When iatrogenic causes are excluded, the commonest cause of Cushing's syndrome is Cushing's disease, accounting for approximately 70% of cases. The adrenal glands in these patients show bilateral adrenocortical hyperplasia with widening of the zona fasciculata and zona reticularis.

Cushing's Disease

The condition is most readily classified into ACTH-dependent and ACTH-independent causes. This can be diagnosed by documenting an increase in the ratio of urinary cortisol to cortisone metabolites. In addition, hepatic 5-reductase activity is inhibited, resulting in greater excretion of 5-cortisol metabolites.

Cardiovascular

Hypertension is another prominent feature, occurring in up to 75% of cases; even though epidemiologic data show a strong association between blood pressure and obesity, hypertension is much more common in patients with Cushing's syndrome than in those with simple obesity. This, together with the established metabolic consequences of the disease (diabetes, hyperlipidemia; see the following), is thought to explain the increased cardiovascular mortality in untreated cases. In addition, thromboembolic events may be commoner in Cushing's patients.

Infections

Infections are more common in patients with Cushing's syndrome. In many instances these are asymptomatic and occur because the normal inflammatory response is suppressed. Reactivation of tuberculosis has been reported and has even been the presenting feature in some cases. Fungal infections of the skin (notably tinea versicolor) and nails may occur, as may opportunistic fungal infections. Bowel perforation is commoner in patients with extreme hypercortisolism, and the hypercortisolism may mask the usual symptoms and signs of the condition. Wound infections are commoner and contribute to poor wound healing.

Eye

Ocular effects include raised intraocular pressure and exophtalmos (in up to one third of patients in Cushing's original series), the latter occurring because of increased retro-orbital fat deposition. Cataracts, a well-recognized complication of corticosteroid therapy, seem to be uncommon except as a complication of diabetes. In the author's experience chemosis is a sensitive and underreported feature of Cushing's syndrome.

Classification and Pathophysiology of Cushing's Syndrome

The condition is most readily classified into ACTH-dependent and ACTH-independent causes (see Table 14-10).

Adrenocorticotropic Hormone (ACTH) Dependent Causes

Cushing's Disease

When iatrogenic causes are excluded, the commonest cause of Cushing's syndrome is Cushing's disease, accounting for approximately 70% of cases. The adrenal glands in these patients show bilateral adrenocortical hyperplasia with widening of the zona fasciculata and zona reticularis.

Etiology.

Cushing himself raised the question of whether this disease was a primary pituitary condition or secondary to an abnormality in the hypothalamus, and there has been an ongoing debate on this issue ever since. The hypothalamic theory states that ACTH-secreting adenomas arise because of dysfunctional regulation of corticotrophs through chronic stimulation by CRH (or AVP), but other studies provide data to support a primary pituitary defect as the cause of the condition (Table 14-12).

<table>
<thead>
<tr>
<th>TABLE 14-12</th>
<th>Hypothalamic versus Pituitary Theory Underpinning the Etiology of Cushing's Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hypothalamic Theory</strong></td>
<td><strong>Pituitary Theory</strong></td>
</tr>
<tr>
<td>Neuroendocrine abnormalities</td>
<td>Lack of &quot;cure&quot; after pituitary stalk section</td>
</tr>
<tr>
<td>Loss of circadian rhythm, sleep disturbance, other &quot;hypothalamic defects&quot; (TSH, LH-FSH secretion)</td>
<td>Circulating and CSF CRH levels are suppressed</td>
</tr>
<tr>
<td>Efficacy of centrally acting drugs</td>
<td>Reversal of &quot;hypothalamic defects&quot; upon correction of hypercortisolism</td>
</tr>
</tbody>
</table>

Infections are more common in patients with Cushing's syndrome. In many instances these are asymptomatic and occur because the normal inflammatory response is suppressed. Reactivation of tuberculosis has been reported and has even been the presenting feature in some cases. Fungal infections of the skin (notably tinea versicolor) and nails may occur, as may opportunistic fungal infections. Bowel perforation is commoner in patients with extreme hypercortisolism, and the hypercortisolism may mask the usual symptoms and signs of the condition. Wound infections are commoner and contribute to poor wound healing.

Eye

Ocular effects include raised intraocular pressure and exophtalmos (in up to one third of patients in Cushing's original series), the latter occurring because of increased retro-orbital fat deposition. Cataracts, a well-recognized complication of corticosteroid therapy, seem to be uncommon except as a complication of diabetes. In the author's experience chemosis is a sensitive and underreported feature of Cushing's syndrome.

Classification and Pathophysiology of Cushing's Syndrome

The condition is most readily classified into ACTH-dependent and ACTH-independent causes (see Table 14-10).
The hypothalamic may have an initiating role, and overwhelming evidence is that, at presentation, the condition is pituitary-dependent. In 85% to 90% of cases the disease is due to a pituitary adenoma of mononuclear origin; basophil hyperplasia alone is found in 9% to 33% of pathologic series. On clinical grounds, this condition can be divided into two entities, cases occurring in the setting of highly malignant tumors such as small cell carcinoma of bronchus (Table 14-13) and more indolent cases occurring in patients with underlying neuroendocrine tumors such as bronchial carcinoids. In the former case, the clinical presentation more commonly resembles Addison's disease than Cushing's syndrome. Circulating ACTH concentrations and cortisol secretion rates can be extremely high. As a result, the duration of symptoms from onset to presentation is short (<3 months); patients are commonly pigmented, and the metabolic manifestations of glucocorticoid excess are often rapid and progressive. Weight loss, myopathy, and glucose intolerance are prominent symptoms and signs. The association of these features with hypokalemic alkalosis and peripheral edema should alert the clinician to the diagnosis.

The distinction of pituitary-dependent Cushing's from these indolent causes of ectopic ACTH syndrome is challenging. A key biochemical hallmark of the disease is a relative resistance of ACTH secretion to normal glucocorticoid feedback inhibition. ACTH-secreting pituitary adenomas function at a higher than normal set-point for cortisol feedback. In Cushing's disease the predominant finding is an increase in ACTH pulse amplitude with loss of normal circadian rhythm, but ACTH pulse frequency is also increased in some cases (see Fig. 14-7).

### TABLE 14-13 -- Tumors Associated with the Ectopic Adrenocorticotropic Hormone Syndrome

<table>
<thead>
<tr>
<th>Tumor Type</th>
<th>Approximate Incidence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small cell lung carcinoma</td>
<td>50</td>
</tr>
<tr>
<td>Nonsmall cell lung carcinoma</td>
<td>5</td>
</tr>
<tr>
<td>Pancreatic tumors (including carcinoids)</td>
<td>10</td>
</tr>
<tr>
<td>Thymic tumors (including carcinoids)</td>
<td>5</td>
</tr>
<tr>
<td>Lung carcinoids</td>
<td>10</td>
</tr>
<tr>
<td>Other carcinoids</td>
<td>2</td>
</tr>
<tr>
<td>Medullary carcinoma of thyroid</td>
<td>5</td>
</tr>
<tr>
<td>Pheochromocytoma and related tumors</td>
<td>3</td>
</tr>
<tr>
<td>Rare carcinomas of prostate, breast, ovary, gallbladder, colon</td>
<td>10</td>
</tr>
</tbody>
</table>

The majority of tumors (usually bronchial carcinoids, but larger macroadenomas occur in up to 10% of cases and usually signify a more invasive tumor. Selective surgical removal of a macroadenoma results in cure with a low recurrence rate. However, it is possible, particularly in cases with no identifiable pituitary adenoma, that Cushing's disease may be heterogeneous with different subtypes.

Ectopic Adrenocorticotropic Hormone Syndrome

In 15% of cases, Cushing's syndrome may be associated with nonpituitary tumors secreting ACTH and the ectopic ACTH syndrome. On clinical grounds, this condition can be divided into two entities, cases occurring in the setting of highly malignant tumors such as small cell carcinoma of bronchus (Table 14-13) and more indolent cases occurring in patients with underlying neuroendocrine tumors such as bronchial carcinoids. In the former case, the clinical presentation more commonly resembles Addison's disease than Cushing's syndrome. Circulating ACTH concentrations and cortisol secretion rates can be extremely high. As a result, the duration of symptoms from onset to presentation is short (<3 months); patients are commonly pigmented, and the metabolic manifestations of glucocorticoid excess are often rapid and progressive. Weight loss, myopathy, and glucose intolerance are prominent symptoms and signs. The association of these features with hypokalemic alkalosis and peripheral edema should alert the clinician to the diagnosis.

Depending on local referral practice, approximately 20% of cases of ectopic ACTH syndrome are explained by indolent tumors, such as benign bronchial carcinoids that produce ACTH. In these cases, symptoms and signs are commonly present for 18 months from onset to clinical presentation. Such patients present with the typical features of Cushing's syndrome and may be biochemically similar to patients with Cushing's disease. This, once a diagnosis of Cushing's syndrome is established, the principal diagnostic dilemma is in establishing the distinction of pituitary-dependent Cushing's from these indolent causes of ectopic ACTH syndrome.

### Etiology

POMC is expressed in normal peripheral extrapancreatic tissues and many tumors (lung, testis) irrespective of the presence of Cushing's syndrome, raising the appropriateness of the term "ectopic" ACTH syndrome. Tumors most commonly associated with ectopic ACTH syndrome arise from neuroendocrine tissues, the cells of which have the ability to take up and decarboxylate amine precursors (APUD cells). However, in the case of small cell lung cancer, only 0.5% to 1% of tumors are associated with ectopic ACTH syndrome and the explanation for the development of ACTH secretion remains unclear. POMC mRNA transcripts are usually shorter in tumors not associated with ectopic ACTH syndrome, whereas those with the syndrome express larger POMC mRNA species in addition to the "pituitary" size transcript. In addition to aberrant transcriptional regulation of the POMC gene, interaction with tissue-specific transcription factors or methylation status of the POMC gene may be involved (see earlier). Once secreted, POMC is cleaved in the pituitary by specific serine endopeptidases to produce ACTH precursors; in ectopic ACTH syndrome, aberrant peripheral processing of POMC may lead to increased circulating ACTH precursor concentrations (pro-ACTH, amino-terminal POMC [N-POMC]). In contrast to ACTH-secreting pituitary adenomas, ectopic POMC or ACTH production is not responsive to normal glucocorticoid feedback because of a defective GR or GR signaling mechanism. However, this sensitivity to glucocorticoid feedback is far from clear-cut, which is one reason why the differential diagnosis of ACTH-dependent Cushing's syndrome can be challenging.

### Ectopic Corticotropin-Releasing Hormone Production

This is a rare cause of pituitary-dependent Cushing's syndrome. A number of cases have been described in which a tumor (usually bronchial carcinoid, medullary thyroid, or prostate carcinoma) has been shown to contain CRH but not ACTH. However, CRH is also produced by many normal tissues. Where available, pituitary histology reveals corticotroph hyperplasia but not adenoma formation. Biochemically, these patients usually are similar to patients with ectopic ACTH syndrome with loss of the normal negative glucocorticoid feedback mechanism; 50% are resistant to high-dose dexamethasone therapy. It has been suggested that ectopic CRH production may explain the suppression of cortisol secretion after high-dose dexamethasone found in some patients with the ectopic ACTH syndrome.

### Macronodular Adrenal Hyperplasia

In 10% to 40% of patients with Cushing's disease, there is bilateral adrenocortical hyperplasia associated with one or more nodules that may be up to several centimeters in diameter. Patients tend to be older and have had symptoms for a longer time but otherwise present with the classical clinical features of Cushing's syndrome. Pathologically, the nodules are lobulated and can be markedly enlarged, but interstitial adrenal hyperplasia is invariably found. Macronodular adrenal hyperplasia is thought to result from long-standing adrenal ACTH stimulation that leads to autonomous adenoma formation. Thus, the adrenals in a patient with Cushing's disease become more hyperplastic, they secrete more cortisol for a given ACTH level and the ACTH may ultimately lead to autorepression. Individual clinical cases support this hypothesis, and macronodular adrenal hyperplasia should be regarded as an ACTH-dependent form of Cushing's syndrome, even though...
ACTH levels may be relatively low and dexamethasone suppressibility

less marked than in other cases of Cushing's disease. The adrenals can be a trap for the unwary as they may be mistaken for primary adrenal tumors.

Adrenocorticotropin Hormone-independent Causes

Cortisol-secreting Adrenal Adenoma and Carcinoma

With the exclusion of iatrogenic Cushing's syndrome, adrenal adenomas are responsible for about 10% to 15% of cases and carcinomas for less than 5%. By contrast, in children 65% of cases of Cushing's syndrome have an adrenal etiology (15% adenomas, 50% carcinoma). Onset of clinical features is gradual in patients with adenomas but often rapid in adrenal carcinoma. In addition to the features of hypercortisolism, patients may complain of loin or abdominal pain and a tumor may be palpable. The tumor may secrete other steroids, such as androgens or mineralocorticoids. Thus, in females, there may be features of virilization, with hirsutism, clitoromegaly, breast atrophy, deepening of the voice, temporal recession, and severe acne. With "pure" cortisol-secreting adenomas, hirsutism is uncommon. Subclinical Cushing's syndrome has been reported in patients with adrenal "incidentalomas" (see later).

Primary Pigmented Nodular Adrenal Hyperplasia and Carney Complex

About 100 cases of ACTH-independent Cushing's syndrome have been reported in association with bilateral, small pigmented adrenal nodules. Pathologically, these nodules are usually 2 to 4 mm in diameter (but can be larger) and black or brown on cut section. Adjacent adrenal tissue is strophic, distinguishing this condition from macronodular adrenal hyperplasia. Presentation is with typical features of Cushing's syndrome but is always before 30 years of age and before 15 years of age in 50% of cases. Cases of primary pigmented nodular adrenocortical disease (PPNAD) have been reported without Cushing's syndrome. Bilateral adrenalec-tomy is curative.

In 20% of cases, there is a family history, and it is known that PPNAD forms part of the familial autosomal dominant condition called Carney complex (Table 14-14). This comprises mesenchymal tumors (especially atrial myxomas), spotty skin pigmentation, peripheral nerve tumors, and various tumors including breast lesions and testicular and GH-secreting pituitary tumors.

Genetic linkage studies in affected kindreds

Table 14-14—Clinical Features of the Carney Complex

<table>
<thead>
<tr>
<th>Feature</th>
<th>Prevalence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skin lesions</td>
<td>80</td>
</tr>
<tr>
<td>Pigmented lesions</td>
<td>72</td>
</tr>
<tr>
<td>Blue nevi</td>
<td>45</td>
</tr>
<tr>
<td>Cutaneous myxomas</td>
<td>45 (females only)</td>
</tr>
<tr>
<td>Cardiac myxomas</td>
<td>56 (males only)</td>
</tr>
<tr>
<td>Pigmented nodular adrenal hyperplasia</td>
<td>&lt;5</td>
</tr>
<tr>
<td>Bilateral fibroadenomas</td>
<td>Rare</td>
</tr>
<tr>
<td>Testicular tumors</td>
<td>Rare</td>
</tr>
<tr>
<td>Pituitary lesions, usually growth hormone-secreting</td>
<td>Rare</td>
</tr>
<tr>
<td>Neurogenic tumors (gastric schwannomas)</td>
<td>Rare</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>Rare</td>
</tr>
<tr>
<td>Thyroid cancer</td>
<td>Rare</td>
</tr>
<tr>
<td>Acoustic neuroma</td>
<td>Rare</td>
</tr>
<tr>
<td>Hepatoma</td>
<td>Rare</td>
</tr>
</tbody>
</table>

In this condition, fibrous dysplasia and cutaneous pigmentation may be associated with pituitary, thyroid, adrenal, and gonadal hyperfunction. The commonest manifestation is with sexual precocity and GH excess, but Cushing's syndrome has been reported. The underlying abnormality is a somatic mutation in the subunit of the stimulatory G protein that is linked to adenyl cyclase. The mutation results in the G protein being constitutively activated, mimicking constant ACTH stimulation at the level of the adrenal. ACTH levels are suppressed, and adrenal adenomas may occur.

McCune-Albright Syndrome

Although macronodular hyperplasia commonly occurs in patients with ACTH-dependent Cushing's syndrome (see earlier), truly ACTH-independent macronodular disease is also recognized as a distinct entity. The nodules are nonpigmented and over 5 mm in diameter; occasionally the adenals may be massively enlarged. The pathogenesis is unknown in many cases, but in one kindred activating mutations of the ACTH receptor-coupled receptor may have caused the phenotype. It is likely that many cases may be explained on the basis of aberrant receptor expression within the adrenal cortex. Patients have been described with macronodular hyperplasia, ACTH-independent Cushing's syndrome, and enhanced adrenal responsiveness to gastric inhibitory polypeptide (GIP). Biochemically, plasma cortisol levels are subnormal in the morning and rise after food because of the normal increase in GIP after eating. The adenocortical tissue of these patients responded in vitro to low doses of GIP, whereas there was no such effect in normal adrenal cortex, suggesting that adrenal GIP receptors are linked to steroidogenesis in these patients. It remains to be seen whether abnormalities of adrenal sensitivity to GIP play a subtle role in other types of Cushing's syndrome. Similarly, Cushing's syndrome related to a cortisol-secreting adrenal adenoma has been attributed to aberrant expression of receptors for interleukin-1.

Iatrogenic Cushing's Syndrome

The basis for this condition is discussed under "Therapeutic Corticosteroids." Development of the features of Cushing's syndrome depends on the dose, duration, and potency of corticosteroid used in clinical practice. ACTH is rarely prescribed but chronically also results in cushingoid features. Some features such as increased intraocular pressure, cataracts, benign intracranial hypertension, aseptic necrosis of the femoral head, osteoporosis, and pancreatitis are commoner in iatrogenic compared with endogenous Cushing's syndrome, whereas other features, notably hypertension, hirsutism, and oligomenorrhea or amenorrhea, arerarer.

Special Features of Cushing's Syndrome

Cyclic Cushing's Syndrome
Of particular clinical interest has been a group of patients with cyclic Cushing's syndrome, characterized by periods of excess cortisol production interspersed with intervals of normal cortisol production. Some of these patients demonstrate a paradoxical rise in plasma ACTH and cortisol when treated with dexamethasone, and occasional patients show benefit with dopamine agonist (bromocriptine) or serotonin antagonist (cyproheptadine) therapy. The majority of cases have been thought to have pituitary-dependent disease, and in many of these patients basophil adenomas have been removed, some with long-term cure. However, cortisol secretion may show some evidence of cyclicity in patients with an ectopic source of ACTH syndrome.

Children

In children, in addition to the preceding features, growth arrest is almost invariably. The dissociation between height and weight on the growth chart is obvious. If the patient is growing along the same centile line, the diagnosis of Cushing's syndrome is highly unlikely. In addition to glucocorticoid-induced growth arrest, androgen excess may result in precocious puberty. Adrenal causes account for 65% of all cases.

Pregnancy

Pregnancy is rare in women with Cushing's syndrome because of associated amenorrhea related to androgen excess or hypercortisolism. However, approximately 100 such cases have been reported, 50% of which were due to adrenal adenomas. A few cases of true pregnancy-induced Cushing's syndrome have been described with regression post partum. In these cases, the etiology is unknown. Establishing a diagnosis and cause can be difficult; clinically, striae, hypertension, and gestational diabetes are common features in pregnancy, yet hypertension and diabetes are the commonest signs of Cushing's syndrome in pregnant women (70% and 30% of all cases, respectively). Furthermore, biochemically, normal pregnancy is associated with a threefold increase in plasma cortisol because of increased cortisol production rates and increases in CBG. Urinary free cortisol also rises, and dexamethasone does not suppress plasma cortisol to the same degree as in the nonpregnant state. Untreated, the condition results in high maternal and fetal morbidity and mortality. Adrenal or pituitary adenomas should be excised. Metyrapone, which is not teratogenic, has been effective in many cases in controlling the hypercortisolism.

Pseudo-Cushing's Syndromes

A pseudo-Cushing's state can be defined as some or all of the clinical features of Cushing's syndrome together with some evidence for hypercortisolism. Resolution of the underlying cause results in disappearance of the cushingoid state. Several causes have been described.

Alcohol

In the original description of this syndrome, urinary and plasma cortisol levels were elevated and not suppressed by dexamethasone. Plasma ACTH has been found to be normal or suppressed. The condition is rare but should be suspected in a patient with an ongoing history of heavy alcohol intake and biochemical or clinical evidence of chronic liver disease. The pathogenesis of this condition remains unknown, but a "two-hit" hypothesis has been put forward to explain its etiology. Chronic liver disease of any cause is associated with impaired cortisol metabolism, but in alcoholics this is associated with an increase in cortisol secretion rate rather than concomitant suppression in the presence of impaired metabolism. In some studies, alcohol has directly stimulated cortisol secretion; alternatively, vasopressin levels are elevated in patients with decompensated liver disease and may stimulate the HPA axis. With abstinence from alcohol the biochemical abnormalities rapidly revert to normal.

Depression

Although the cause is unknown, it is recognized that patients with depression may exhibit the hormonal abnormalities of patients with Cushing's syndrome. These abnormalities are reversible on correction of the psychiatric condition. Conversely, patients with Cushing's syndrome are frequently depressed, and a careful clinical and endocrinologic assessment is required.

Obesity

Although one of the commonest referrals to a clinical endocrinologist is to exclude an underlying endocrine cause in a patient with obesity, the diagnosis of Cushing's syndrome in such patients should not cause difficulties. Patients with obesity have mildly increased cortisol secretion rates, and the data suggest that this is due to activation of the hypothalamo-pituitary axis. However, circulating cortisol concentrations are invariably normal and urinary free cortisol concentrations are either normal or only slightly elevated. The stimulus for the increased secretion rate appears to be increased peripheral metabolism and hence clearance of cortisol (principally reduced hepatic conversion of cortisone to cortisol by 11-HSD type 1 and increased conversion of cortisol to 5-reduced derivatives).

Investigation of Suspected Cushing's Syndrome

There are two stages in the investigation of suspected Cushing's syndrome. (1) Does the patient have Cushing's syndrome? (2) If the answer is yes, then what is the cause? Unfortunately, many investigators fail to make this distinction and ill-advisedly use tests that are relevant to question 2 to try to answer question 1. In particular, it is essential that radiologic investigations not be undertaken until Cushing's syndrome has been confirmed biochemically. The major tests are listed in Table 14-15 and their application in Figure 14-19.

Circadian Rhythm of Plasma Cortisol

In normal subjects, plasma cortisol levels are at their highest first thing in the morning and reach a nadir at about midnight (<50 nmol/L (2 µg/dL) in a nonstressed subject). This circadian rhythm is lost in patients with Cushing's syndrome so that in the majority of patients the 9 AM plasma cortisol is normal but nocturnal levels are raised. Random morning plasma cortisol levels are therefore of little value in making the diagnosis, and a midnight cortisol level greater than

<table>
<thead>
<tr>
<th>TABLE 14-15 -- Tests Used in the Diagnosis and Differential Diagnosis of Cushing's Syndrome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
</tr>
<tr>
<td>------------------------------------------</td>
</tr>
<tr>
<td>ACTH stimulation test</td>
</tr>
<tr>
<td>Dexamethasone suppression test</td>
</tr>
<tr>
<td>24-hour urinary free cortisol</td>
</tr>
<tr>
<td>17-OHCS</td>
</tr>
<tr>
<td>17-ketosteroids</td>
</tr>
<tr>
<td>11-deoxycortisol</td>
</tr>
<tr>
<td>11-deoxycorticosterone</td>
</tr>
<tr>
<td>Salivary cortisol</td>
</tr>
<tr>
<td>Urinary catecholamines</td>
</tr>
<tr>
<td>Thyroid function tests</td>
</tr>
<tr>
<td>Bone density measurements</td>
</tr>
<tr>
<td>CT or MRI of head</td>
</tr>
<tr>
<td>CT or MRI of abdomen</td>
</tr>
<tr>
<td>Adrenal venous sampling</td>
</tr>
<tr>
<td>Dual energy X-ray absorptiometry (DEXA)</td>
</tr>
</tbody>
</table>

---

Note: The above information is a simplified representation of the content provided in the document. For a comprehensive understanding, please refer to the original source.
Certain drugs (phenytoin, rifampicin) may increase the metabolic clearance rate of dexamethasone and lead to false-positive results. Simultaneous measurement of true-positive rate and a false-positive rate less than 1% every 6 hours for 48 hours. Using a postdexamethasone plasma cortisol concentration of less than 50 nmol/L (2 µg/dL), this test is reported as having a 97% to 100% sensitivity.

In the 48-hour low-dose dexamethasone test, plasma cortisol is measured at 9 AM on day 0 and 48 hours later following dexamethasone given at a dose of 0.5 mg every 6 hours for 48 hours. Using a postdexamethasone plasma cortisol concentration of less than 50 nmol/L (2 µg/dL), this test is reported as having a 97% to 100% true-positive rate and a false-positive rate less than 1%. Sensitivity is higher if plasma rather than urinary cortisol is measured.

Corticotropin-releasing hormone
Metyrapone test
\[\text{Corticotropin-releasing hormone} \quad \text{Metyrapone test}\]

Urinary cortisol/creatinine ratio
Corticosteroid excretion
In normal subjects, the administration of a supraphysiologic dose of glucocorticoid results in suppression of ACTH and cortisol secretion. In Cushing's syndrome of any cause there is failure of this suppression when low doses of the synthetic glucocorticoid dexamethasone are given. Various doses of dexamethasone have been used, but 1 mg of dexamethasone is usually given at midnight. A normal response is a plasma cortisol less than 140 nmol/L (5 µg/dL) between 8 and 9 AM the following morning. A dose of 1.5 or 2 mg gives a 30% false-positive rate, whereas after 1 mg this is reduced to 12.5% with a false-negative rate less than 2%. In addition, sensitivity can be improved by reducing the plasma cortisol cutoff value; a postdexamethasone cortisol value of less than 50 nmol/L (2 µg/dL) effectively excludes Cushing's syndrome. Thus, the outpatient overnight test has high sensitivity (95%) but low specificity, and further investigation is often required.

In the 48-hour low-dose dexamethasone test, plasma cortisol is measured at 9 AM on day 0 and 48 hours later following dexamethasone given at a dose of 0.5 mg every 6 hours for 48 hours. Using a postdexamethasone plasma cortisol concentration of less than 50 nmol/L (2 µg/dL), this test is reported as having a 97% to 100% true-positive rate and a false-positive rate less than 1%. Sensitivity is higher if plasma rather than urinary cortisol is measured.

Corticosteroid excretion
Circadian rhythm of plasma cortisol

Low-Dose and Overnight Dexamethasone Suppression Tests

In normal subjects, the administration of a supraphysiologic dose of glucocorticoid results in suppression of ACTH and cortisol secretion. In Cushing's syndrome of any cause there is failure of this suppression when low doses of the synthetic glucocorticoid dexamethasone are given. Various doses of dexamethasone have been used, but 1 mg of dexamethasone is usually given at midnight. A normal response is a plasma cortisol less than 140 nmol/L (5 µg/dL) between 8 and 9 AM the following morning. A dose of 1.5 or 2 mg gives a 30% false-positive rate, whereas after 1 mg this is reduced to 12.5% with a false-negative rate less than 2%. In addition, sensitivity can be improved by reducing the plasma cortisol cutoff value; a postdexamethasone cortisol value of less than 50 nmol/L (2 µg/dL) effectively excludes Cushing's syndrome. Thus, the outpatient overnight test has high sensitivity (95%) but low specificity, and further investigation is often required.

In the 48-hour low-dose dexamethasone test, plasma cortisol is measured at 9 AM on day 0 and 48 hours later following dexamethasone given at a dose of 0.5 mg every 6 hours for 48 hours. Using a postdexamethasone plasma cortisol concentration of less than 50 nmol/L (2 µg/dL), this test is reported as having a 97% to 100% true-positive rate and a false-positive rate less than 1%. Sensitivity is higher if plasma rather than urinary cortisol is measured.

Corticosteroid excretion
Circadian rhythm of plasma cortisol

Low-Dose and Overnight Dexamethasone Suppression Tests

In normal subjects, the administration of a supraphysiologic dose of glucocorticoid results in suppression of ACTH and cortisol secretion. In Cushing's syndrome of any cause there is failure of this suppression when low doses of the synthetic glucocorticoid dexamethasone are given. Various doses of dexamethasone have been used, but 1 mg of dexamethasone is usually given at midnight. A normal response is a plasma cortisol less than 140 nmol/L (5 µg/dL) between 8 and 9 AM the following morning. A dose of 1.5 or 2 mg gives a 30% false-positive rate, whereas after 1 mg this is reduced to 12.5% with a false-negative rate less than 2%. In addition, sensitivity can be improved by reducing the plasma cortisol cutoff value; a postdexamethasone cortisol value of less than 50 nmol/L (2 µg/dL) effectively excludes Cushing's syndrome. Thus, the outpatient overnight test has high sensitivity (95%) but low specificity, and further investigation is often required.

In the 48-hour low-dose dexamethasone test, plasma cortisol is measured at 9 AM on day 0 and 48 hours later following dexamethasone given at a dose of 0.5 mg every 6 hours for 48 hours. Using a postdexamethasone plasma cortisol concentration of less than 50 nmol/L (2 µg/dL), this test is reported as having a 97% to 100% true-positive rate and a false-positive rate less than 1%. Sensitivity is higher if plasma rather than urinary cortisol is measured.
In patients with depression, urinary free cortisol concentrations may be elevated and overlap those seen in patients with true Cushing's syndrome. Compared with patients with Cushing's disease, depressed patients have greater suppressibility after dexamethasone and a reduced response to CRH, but neither of these tests is diagnostic. However, by performing a CRH test after the standard 2-day low-dose dexamethasone suppression test, separation of true versus pseudo-Cushing's syndrome has been possible. In normal subjects and in patients with endogenous depression, insulin-induced hypoglycemia results in a rise in ACTH and cortisol levels, a response that is usually not seen in Cushing's syndrome. Finally, loperamide lowers cortisol values in patients with pseudo-Cushing's but not in true Cushing's syndrome.

When the biochemical diagnosis has been made, a series of investigations are required to determine the cause of the Cushing's syndrome.

### Plasma ACTH

Ideally, ACTH should be measured using a modern two-site immunoradiometric assay, which differentiates ACTH-dependent from ACTH-independent causes. In Cushing's disease, 50\% of patients have a 9 AM ACTH within the normal reference range (2 to 12 pmol/L [9 to 54 pg/mL]); in the remainder it is modestly elevated. ACTH levels in the ectopic ACTH syndrome are high (usually > 20 pmol/L [90 pg/mL]) but nevertheless overlap values seen in Cushing's disease in 30\% of cases and cannot therefore be used to differentiate these two conditions. The most discriminatory time of day to measure ACTH is actually between 11 AM and 1 AM, when ACTH-cortisol secretion is at a nadir, and in our practice ACTH is usually measured with cortisol in the circadian rhythm studies. A midnight ACTH result greater than 5 pmol/L (23 pg/mL) in a patient with biochemical hypercortisolism confirms that the underlying disease is ACTH-dependent. The measurement of ACTH precursors (pro-ACTH, POMC) is not routinely available but may be more useful in detecting an ectopic source of ACTH; more data are required on patients with occult tumors causing the syndrome.

In patients with adrenal tumors, plasma ACTH is invariably undetectable (<1 pmol/L [4.5 pg/mL]). This can also occur with degradation of ACTH; as a result, nonhemolyzed blood samples should be taken on ice and immediately separated.

### Plasma Potassium

Hypokalemic alkalosis is present in more than 95\% of patients with the ectopic ACTH syndrome but is present in fewer than 10\% of patients with Cushing's disease. As discussed earlier ("Metabolic and Endocrine"), the etiology of this mineralocorticoid excess state is now established. Patients with the ectopic syndrome usually have higher cortisol secretion rates that saturate the renal protective 11-HSD type 2 enzyme, resulting in cortisol-induced, mineralocorticoid hypertension (see Chapter 15). In addition, these patients have higher levels of the ACTH-dependent mineralocorticoid DOC, which are not suppressed.

### High-Dose Dexamethasone Suppression Test

The rationale for this test is that in Cushing's disease there is a resetting of the negative feedback control of ACTH to a higher level than normal. Thus, cortisol levels are low normal or intermittently detectable. This may occur in macronodular hyperplasia. The danger is that in some patients the asymmetry of the nodular hyperplasia may lead to a diagnosis of adrenal adenoma, the plasma ACTH is ignored, and an inappropriate adrenalectomy is performed. Conversely, in some patients with this syndrome an autonomous adrenal tumor develops and, despite detectable ACTH, unilateral adrenalectomy is required.

### Metyrapone Test

Metyrapone blocks the conversion of 11-deoxycortisol to cortisol and DOC to corticosterone by inhibiting 11-hydroxylase. This effect lowers plasma cortisol and, through negative feedback control, increases plasma ACTH. This, in turn, stimulates an increase in the secretion of adrenal steroids proximal to the block. When metyrapone is given in doses of 750 mg every 4 hours for 24 hours, patients with Cushing's disease exhibit an exaggerated rise in plasma ACTH with 11-deoxycortisol levels at 24 hours exceeding 1000 pmol/L (35 pg/dL). In most patients with the ectopic ACTH syndrome there is little or no response, but occasional patients (possibly those producing both ACTH and CRH) have an 11-deoxycortisol response that may be similar to that observed in Cushing's disease.

The metyrapone test was originally used to distinguish patients with Cushing's disease from those with a primary adrenal cause. However, these can be more reliably distinguished by measuring plasma ACTH and subsequent computed tomographic (CT) scanning of the adrenals. As indicated, the test does not reliably distinguish between Cushing's disease and the ectopic ACTH syndrome, and the value of this test in modern endocrine practice has been questioned. It should be reserved for patients when the results of other tests are equivocal.

---

**Figure 14-20** Plasma adrenocorticotropic hormone (ACTH) concentrations in patients with Cushing's disease and Cushing's syndrome associated with adrenocortical tumors and ectopic ACTH syndrome. To convert values to pmol/L, multiply by 0.2202. (From Besser GM, Edwards CRW. Cushing's syndrome. Clin Endocrinol Metab 1972; 1:451-490.)
CRH is a 41-amino-acid peptide identified by Vale in 1981 from ovine hypothalamus. The ovine sequence differs by seven amino acid residues from that of the human but, despite this, is slightly more effective in stimulating the release of ACTH in humans. The test involves the intravenous injection of either ovine or human CRH in a dose of 1 µg/kg body weight or a single dose of 100 µg. In some centers CRH is combined with AVP, which results in an augmented ACTH response. The test can be performed in the morning or afternoon, and, after basal sampling, blood samples for ACTH and cortisol are taken every 15 minutes for 1 to 2 hours following the administration of CRH.

In normal subjects, CRH produces a rise in ACTH and cortisol (approximately 15% to 20%), but this response is exaggerated in Cushing’s disease, where typically an ACTH rise greater than 50% and a cortisol rise greater than 20% over baseline values are seen. No response is seen in the ectopic ACTH syndrome, but false-positive results have been reported. In distinguishing pituitary-dependent Cushing’s from the ectopic ACTH syndrome, the response of ACTH and cortisol to CRH has a specificity and sensitivity of approximately 90%. However, using an ACTH increase of 100% or a cortisol rise of 50% over baseline values, a positive response effectively eliminates a diagnosis of ectopic ACTH syndrome, and this is the real benefit of this test. Up to 10% of patients with Cushing’s disease do not respond to CRH.

**Inferior Petrosal Sinus Sampling and Selective Venous Catheterization**

The most robust test to distinguish Cushing’s disease from the ectopic ACTH syndrome is inferior petrosal sinus sampling (IPSS). As blood from each half of the pituitary drains into the ipsilateral inferior petrosal sinus, catheterization and venous sampling of both sinuses simultaneously can distinguish a pituitary from an ectopic source. In virtually all patients with the ectopic ACTH syndrome, the ratio of ACTH concentrations between the inferior petrosal sinus and simultaneously drawn peripheral venous level is less than 1.4:1. In contrast, in Cushing’s disease this ratio is elevated at greater than 2.0. However, because of the problem of intermittent ACTH secretion, it is useful to make measurements before and at intervals (for example, 2, 5, and 15 minutes) after intravenous injection of 100 µg of synthetic ovine CRH. Using this approach, an ACTH petrosal sinus/peripheral ratio greater than 3.0 after CRH has a sensitivity of 97% and specificity of 100% in diagnosing Cushing’s disease.

IPSS may also be of value in lateralizing a pituitary tumor in a patient in whom imaging techniques have failed to demonstrate a microadenoma, although other centers have found that this is of little value in predicting tumor location. Coadministration of desmopressin with CRH may help in localizing the tumor. However, it should be remembered that many tumors are central and may drain into both sinuses; current evidence suggests that it would unwise to base the surgical procedure on the results of IPSS studies alone.

IPSS is a useful technique for establishing the differential diagnosis of ACTH-dependent Cushing’s syndrome. However, it is technically demanding, has been associated with complications (referred aural pain, thomboembolism), and is expensive. In our practice, it is reserved for cases in which the differential diagnosis is still in doubt after conducting the preceding tests.

Rarely, selective catheterization of vascular beds may be required to identify the source of ectopic ACTH secretion, for example, from a small pulmonary carcinoid or thymic tumor.

**Tumor Markers**

Many tumors responsible for the ectopic ACTH syndrome also produce peptide hormones other than ACTH or its precursors.

**Imaging**

High-resolution, thin-section contrast-enhanced imaging using either CT or magnetic resonance imaging (MRI) has revolutionized the investigation of Cushing’s syndrome. However, the results of any imaging technique must always be interpreted alongside the biochemical results if mistakes are to be avoided. In imaging the adrenals, asymmetric nodular hyperplasia may lead to a false diagnosis of adrenal adenoma. Because of the presence of pituitary incidentalomas, pituitary CT or MRI scanning may produce false-positive results, particularly for lesions less than 5 mm in diameter.

Pituitary MRI is the investigation of choice when the biochemical tests suggest Cushing’s disease, with a sensitivity of 70% and specificity of 87%. About 90% of ACTH-secreting pituitary tumors are microadenomas (i.e., less than 10 mm in diameter). The classical features of a pituitary microadenoma are a hypodense lesion after contrast, associated with deviation of the pituitary stalk and a convex upper surface of the pituitary gland. With such small tumors it is not surprising that the sensitivity of CT scanning is relatively low (20% to 60%) with a similar specificity.

By contrast, for adrenal imaging, CT rather than MRI is the investigation of choice, offering better spatial resolution, but MRI scanning may provide diagnostic information in patients with suspected adrenal carcinoma. Once again, it is stressed that adrenal incidentalomas are present in up to 5% of normal subjects (see later), and thus adrenal imaging should not be performed unless biochemical investigation suggests a primary adrenal cause (undetectable ACTH concentrations). Adrenal carcinomas are large and often associated with metastatic spread at presentation.

In patients with occult ectopic ACTH syndrome, high-definition CT or MRI scanning of thorax, abdomen, and pelvis...
with images every 0.5 cm may be required to detect small ACTH-secreting carcinoid tumors (Fig. 14-26) (Figure Not Available).

### Scintigraphy Studies

Scintigraphy is of value in certain patients with primary adrenal pathology. The most commonly used agent is $^{131}$I-labeled 6-Iodomethyl-19-norcholesterol. This is a marker of adrenocortical cholesterol uptake. In patients with adenomas, the isotope is taken up by the

![Image](image_url)

### Treatment of Cushing's Syndrome

#### Adrenal Causes

Adrenal adenomas should be removed by unilateral adrenalectomy, which has a 100% cure rate. With the increasing experience of laparoscopic adrenalectomy in most tertiary centers, this has now become the surgical treatment of choice for unilateral tumors, reducing surgical morbidity and postoperative hospital stay compared with traditional open approaches. After operation it may take many months or even years for the contralateral suppressed adrenal to recover. It is wise, therefore, to give slightly suboptimal replacement therapy with dexamethasone at 0.5 mg in the morning, with intermittent measurement of morning plasma cortisol before taking dexamethasone. When the morning plasma cortisol is above 180 nmol/L (6 µg/dL), dexamethasone can be stopped. A subsequent negative dexamethasone suppression test confirms adrenal recovery. In the interim, all patients should carry a Steroid Alert card and increase their dose of replacement therapy in the event of an intermittent illness (see earlier).

### Adrenal Carcinomas

Adrenal carcinomas have a poor prognosis, and most patients are dead within 2 years of diagnosis. It is usual practice to try to remove the primary tumor even though metastases may be present so as to enhance the response to the adrenolytic agent $p,p'$-dichlorodiphenyldichloroethane ($p,p'$-DDD, mitotane, see later). Radiotherapy to the tumor bed and to some metastases, such as those in the spine, may be of limited value.

### Pituitary-Dependent Cushing's Syndrome

The treatment of Cushing's disease has been significantly enhanced through transsphenoidal surgery conducted by an experienced surgeon. Before the selective removal of a pituitary microadenoma, the treatment of choice was bilateral adrenalectomy. This had an appreciable mortality even in the best centers (up to 4%) and significant morbidity. The major risk was the subsequent development of Nelson's syndrome (postadrenalectomy hyperpigmentation with a locally aggressive pituitary tumor) [Fig. 14-27]. This is often attributed to loss of any negative feedback after adrenalectomy. In an attempt to avoid this, pituitary radiation was often carried out at the time of bilateral adrenalectomy. In addition, these patients required lifelong replacement therapy with dexamethasone at 0.5 mg in the morning, with intermittent measurement of morning plasma cortisol before taking dexamethasone. When the morning plasma cortisol is above 180 nmol/L (6 µg/dL), dexamethasone can be stopped. A subsequent negative dexamethasone suppression test may then demonstrate whether the response to stress is normal. In the interim, all patients should carry a Steroid Alert card and increase their dose of replacement therapy in the event of an intermittent illness (see earlier).

The surgical outcome for transsphenoidal hypophysectomy is center-dependent and related to surgical expertise. Because of the hazards of untreated Cushing's disease and potential complications of surgery, the endocrinologist should refer cases only to a recognized surgical specialist where outcome data have been established. In optimal centers, cure rates are 80% to 90% for microadenomas and 50% for macroadenomas. Rates for hypopituitarism and permanent diabetes insipidus postoperatively depend on how aggressive the surgeon has been in removing pituitary tissue. The ideal outcome is a cured patient with intact pituitary function, but this may not be possible in a patient with Cushing's disease in whom a pituitary adenoma was not identified preoperatively or during the
operation itself.

At the time of surgery, patients should be treated with corticosteroids as for any other potential or confirmed deficit of the HPA axis (see later). Postoperatively, hydrocortisone can be withdrawn to maintenance replacement doses, usually within 3 to 7 days. On day 5 postoperatively, plasma cortisol should be measured at 9 AM with the patient having omitted hydrocortisone for 24 hours. After selective removal of a microadenoma, the surrounding corticotrophs are normally suppressed (Fig. 14-28). In these cases plasma cortisol levels are less than 30 nmol/L (1 µg/dL) postoperatively and glucocorticoid replacement therapy is required. Using the dexamethasone regimen described earlier after removal of an adenoma, there is usually (but not invariably) gradual recovery of the HPA axis (Fig. 14-29). A nonsuppressed plasma cortisol postoperatively suggests that the patient is not cured even though cortisol secretion may have fallen to normal or subnormal values. The recurrence rate in patients with an established cure after pituitary surgery is 2%, but this value is higher in children (up to 40%). A detailed assessment of residual pituitary function is required in each case, and close follow-up of such individuals is warranted.

In the past, pituitary radiation was often used in the treatment of Cushing's disease. However, the improvements in pituitary surgery have resulted in far fewer patients being so treated. In children, pituitary radiation appears to be more effective. Radiotherapy is not recommended as a primary treatment but is reserved for patients who do not respond to pituitary microsurgery, those in whom bilateral adrenalectomy has been performed, or patients with established Nelson's syndrome.

Ectopic Adrenocorticotropic Hormone Syndrome

Treatment of the ectopic ACTH syndrome depends on the cause. If the tumor can be found and has not spread, then its removal can lead to cure (e.g., bronchial carcinoid or thymoma). However, the prognosis for small cell lung cancer associated with the ectopic ACTH syndrome is poor. The cortisol excess and associated hypokalemia and diabetes mellitus can be ameliorated by medical therapy. The treatment of the small cell tumor itself also, at least initially, produces improvement. Sometimes, if the ectopic source of ACTH cannot be found, it may be necessary to perform bilateral adrenalectomy and then observe the patient carefully (sometimes for several years) before the primary tumor becomes apparent.

Medical Treatment of Cushing's Syndrome

Several drugs have been used in the treatment of Cushing's syndrome. Metyrapone inhibits 11-hydroxylase and has been most commonly given, often to lower cortisol concentrations prior to definitive therapy or while awaiting benefit from pituitary radiation. The daily dose has to be determined by measuring either plasma or urinary free cortisol. The aim should be to achieve a mean plasma cortisol of about 300 nmol/L (11 µg/dL) during the day or a normal urinary free cortisol. The drug is usually given in doses ranging from 250 mg twice daily to 1.5 g every 6 hours. Nausea is a side effect that can be helped (if it is not due to adrenal insufficiency) by giving the drug with milk.

Aminoglutethimide is a more toxic drug that, in high doses, blocks earlier enzymes in the steroidogenic pathway and thus affects the secretion of steroids other than cortisol. In doses of 1.5 to 3 g daily (start with 250 mg every 8 hours) it commonly produces nausea, marked lethargy, and a high incidence of skin rash. It is commonly prescribed as combination therapy with metyrapone.

Trilostane, a 3-HSD inhibitor, is ineffective in Cushing's disease, as the block in steroidogenesis is overcome by the rise in ACTH. However, it can be effective in patients with adrenal adenomas.

Ketokonazole is an imidazole that has been widely used as an antifungal agent but causes abnormal liver function tests in about 15% of patients. Ketokonazole blocks a variety of steroidogenic cytochrome P450-dependent enzymes and thus lowers plasma cortisol levels. For effective control of Cushing's syndrome, 400 to 800 mg daily has been required.

Mitotane is an adrenolytic drug that is taken up by both normal and malignant adrenal tissue, causing adrenal atrophy and necrosis. Because of its toxicity, it has been used mainly in the management of adrenal carcinoma. Doses of up to 10 to 20 g/day are required to control glucocorticoid excess, although evidence that it causes tumor shrinkage or improves long-term survival is lacking. The drug also produces mineralocorticoid deficiency, and concomitant glucocorticoid and mineralocorticoid replacement therapy may be required. Side effects are common and include fatigue, skin rashes, and gastrointestinal disturbance.

Prognosis of Cushing's Syndrome

Studies carried out before the introduction of effective therapy indicated that 50% of patients with untreated Cushing's syndrome died within 5 years, principally from vascular disease. Even with modern management, an increased prevalence...
Figure 14-29 Gradual recovery of function of the hypothalamic-pituitary-adrenal axis after removal of a pituitary adrenocorticotropic hormone-secreting microadenoma. The insulin hypoglycemia test (I.H.T.) eventually demonstrated the return of a normal stress response.
Glucocorticoid Resistance

A small number of patients have been described who have increased cortisol secretion but without the stigmata of Cushing's syndrome. These patients are resistant to suppression of cortisol with low-dose dexamethasone but respond to high doses. ACTH levels are elevated and lead to increased adrenal production of androgens and DOC. Thus, patients may present with the features of androgen or mineralocorticoid excess, or both. Treatment with a dose of dexamethasone adequate to suppress ACTH (usually 3 mg/day) results in a fall in adrenal androgens and often return of plasma potassium and blood pressure to normal levels. Many of these patients have been found to have point mutations in the steroid-binding domain of the GR, with consequent reduction of glucocorticoid-binding affinity, but this is not invariable. A useful clinical discriminatory test to differentiate this condition from Cushing's syndrome is to measure bone mineral density; this is preserved in patients with glucocorticoid resistance or even increased in female patients because of the androgen excess. In addition, circadian rhythm for ACTH and cortisol is preserved in patients with glucocorticoid resistance.
Glucocorticoid Deficiency

Primary and Secondary Hypoadrenalism

Primary hypoadrenalism refers to glucocorticoid deficiency occurring in the setting of adrenal disease; secondary hypoadrenalism arises because of deficiency of ACTH (Table 14-16). A major distinction between these two is that mineralocorticoid deficiency invariably accompanies primary hypoadrenalism but does not occur in secondary hypoadrenalism because only ACTH is deficient; the renin-angiotensin-aldosterone axis is intact. A further important cause of adrenal insufficiency where there may be dissociation of glucocorticoid and mineralocorticoid secretion is congenital adrenal hyperplasia (CAH) (see later).

Primary Hypoadrenalism

Addison's Disease

Thomas Addison described this condition in his classical monograph published in 1855.

Etiology.

This is a rare condition with an estimated incidence in the developed world of 0.8 cases per 100,000 and prevalence of 4 to 11 cases per 100,000 population. It is associated with significant morbidity and mortality, but when the diagnosis is made it can be easily treated. The causes of Addison's disease are listed in Table 14-16.

Autoimmune Adenititis.

In the Western world, autoimmune adrenalitis accounts for over 70% of all cases. Pathologically, the adrenal glands are atrophic, with loss of most of the cortical cells, but the medulla is usually intact. In 75% of cases adrenal autoantibodies can be detected. Fifty percent of patients with this form of Addison's disease have an associated autoimmune disease (Table 14-17), thyroid disease being the commonest. Conversely, only 1% to 2% of patients with commoner autoimmune diseases such as insulin-dependent diabetes mellitus or thyrotoxicosis have antialdrenal autoantibodies and adrenal disease. This figure is higher in patients with autoimmune hypoparathyroidism (16%).

These autoimmune polyendocrine syndromes (APS) I and II have been classified into two distinct variants. APS type I is inherited as an autosomal recessive condition and comprises Addison's disease, chronic mucocutaneous candidiasis, and hypoparathyroidism. Also called autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy

---

**TABLE 14-16 -- Etiology of Adrenocortical Insufficiency (Excluding CAH)**

<table>
<thead>
<tr>
<th>Primary: Addison’s Disease</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Autoimmune</td>
<td></td>
</tr>
<tr>
<td>Sporadic</td>
<td></td>
</tr>
<tr>
<td>Autoimmune polyendocrine syndrome type I (Addison's disease, chronic mucocutaneous candidiasis, hypoparathyroidism, dental enamel hypoplasia, alopecia, primary gonadal failure, <a href="#">Chapter 37</a>)</td>
<td></td>
</tr>
<tr>
<td>Autoimmune polyendocrine syndrome type II (Schmidt's syndrome) (Addison's disease, primary hypothyroidism, primary hypogonadism, insulin-dependent diabetes, pernicious anaemia, vitiligo, <a href="#">Chapter 37</a>)</td>
<td></td>
</tr>
<tr>
<td>Infections</td>
<td></td>
</tr>
<tr>
<td>Tuberculosis</td>
<td></td>
</tr>
<tr>
<td>Fungal infections</td>
<td></td>
</tr>
<tr>
<td>Cytomegalovirus</td>
<td></td>
</tr>
<tr>
<td>HIV</td>
<td></td>
</tr>
<tr>
<td>Metastatic tumor</td>
<td></td>
</tr>
<tr>
<td>Infiltrations</td>
<td></td>
</tr>
<tr>
<td>Amyloid</td>
<td></td>
</tr>
<tr>
<td>Hemochromatosis</td>
<td></td>
</tr>
<tr>
<td>Intra-adrenal haemorrhage (Waterhouse-Fridrichsen syndrome) after meningococcal septicemia</td>
<td></td>
</tr>
<tr>
<td>Adrenoleukodystrophies</td>
<td></td>
</tr>
<tr>
<td>Congenital adrenal hypoplasia</td>
<td></td>
</tr>
<tr>
<td>DAX-1 mutations</td>
<td></td>
</tr>
<tr>
<td>SF-1 mutations</td>
<td></td>
</tr>
<tr>
<td>ACTH resistance syndromes</td>
<td></td>
</tr>
<tr>
<td>Mutations in MC2-R</td>
<td></td>
</tr>
<tr>
<td>Triple A syndrome</td>
<td></td>
</tr>
<tr>
<td>Bilateral adrenalectomy</td>
<td></td>
</tr>
</tbody>
</table>

**Secondary**

| Exogenous glucocorticoid therapy |          |
| Hypopituitarism                  |          |
| Selective removal of ACTH-secreting pituitary adenoma | 
| Pituitary tumors and pituitary surgery, craniopharyngiomas | 
| Pituitary apoplexy               |          |
| Granulomatous disease (tuberculosis, sarcoid, eosinophilic granuloma) |
Secondary tumor deposits (breast, bronchus)

Postpartum pituitary infarction (Sheehan’s syndrome)

Isolated ACTH deficiency

Idiopathic

Lymphocytic hypophysitis

POMC processing defect

POMC gene mutations

ACTH, adrenocorticotropic hormone; HIV, human immunodeficiency virus; POMC, pro-opiomelanocortin.

dysplasia (APECED) (Table 14-18), the condition is rare and is usually seen in childhood with either candidiasis or hypoparathyroidism. Other autoimmune conditions such as pernicious anemia, thyroid disease, chronic active hepatitis, and gonadal failure may occur but are rare (see also Chapter 37). The adrenal autoantibodies characterizing PGA type I are to the steroidogenic enzymes side-chain cleavage and 17-hydroxylase but not to 21-hydroxylase. The disease has been mapped to chromosome 21q22.3, and mutations in a transcription regulation gene designated AIRE (autoimmune regulator) have been defined.

APS II is commoner and comprises Addison’s disease, autoimmune thyroid disease, diabetes mellitus, and hypogonadism (see Table 14-18). The condition has an inherited basis with linkage to the human leukocyte antigen (HLA) major histocompatibility complex, notably HLA DR3 and HLA DR4. Autoantibodies to 21-hydroxylase are usually present and are predictive of the development of adrenal destruction.

Other autoimmune conditions such as pernicious anemia, thyroid disease, chronic active hepatitis, and gonadal failure may occur but are rare (see also Chapter 37). The condition has an inherited basis with linkage to the human leukocyte antigen (HLA) major histocompatibility complex, notably HLA DR3 and HLA DR4. Autoantibodies to 21-hydroxylase are usually present and are predictive of the development of adrenal destruction.

Other features may accompany APS I and APS II (see Table 14-18). Patients

Other features may accompany APS I and APS II (see Table 14-18).

TABLE 14-17 – Incidence of Other Endocrine and Autoimmune Diseases in Patients with Autoimmune Adrenal Insufficiency (N = 448)

<table>
<thead>
<tr>
<th>Disease</th>
<th>Incidence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thyroid disease</td>
<td></td>
</tr>
<tr>
<td>Hypothyroidism</td>
<td>8</td>
</tr>
<tr>
<td>Nontoxic goiter</td>
<td>7</td>
</tr>
<tr>
<td>Thyrotoxicosis</td>
<td>7</td>
</tr>
<tr>
<td>Gonadal failure</td>
<td></td>
</tr>
<tr>
<td>Ovarian</td>
<td>20</td>
</tr>
<tr>
<td>Testicular</td>
<td>2</td>
</tr>
<tr>
<td>Insulin-dependent diabetes mellitus</td>
<td>11</td>
</tr>
<tr>
<td>Hypoparathyroidism</td>
<td>10</td>
</tr>
<tr>
<td>Pernicious anemia</td>
<td>5</td>
</tr>
<tr>
<td>None</td>
<td>53</td>
</tr>
</tbody>
</table>

with APS are more likely to be female (70%); conversely, patients presenting with isolated autoimmune adrenalitis are usually male.

Infections.

Worldwide, infectious diseases are the commonest cause of primary adrenal insufficiency and comprise tuberculosis, fungal infections (histoplasmosis, cryptococcosis), and cytomegalovirus infection. Adrenal failure may also occur in the acquired immunodeficiency syndrome.

Tuberculous Addison’s disease results from hematogenous spread of the infection from elsewhere in the body, and extra-adrenal disease is usually evident. The adrenals are initially enlarged with extensive epithelioid granulomas and caseation, and both the cortex and the medulla are affected. Fibrosis ensues, and the adrenals become normal or smaller in size with calcification evident in 50% of cases.

The adrenals are frequently involved in patients with acquired

TABLE 14-18 – Clinical Manifestations of Autoimmune Polyendocrine Syndromes (APS) Associated with Adrenal Insufficiency

<table>
<thead>
<tr>
<th>Disorder</th>
<th>Prevalence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>APS, Type I</td>
<td></td>
</tr>
<tr>
<td>Endocrine</td>
<td></td>
</tr>
<tr>
<td>Hypoparathyroidism</td>
<td>89</td>
</tr>
<tr>
<td>Chronic mucocutaneous candidiasis</td>
<td>75</td>
</tr>
<tr>
<td>Adrenal insufficiency</td>
<td>60</td>
</tr>
<tr>
<td>Gonadal failure</td>
<td>45</td>
</tr>
<tr>
<td>Hypothyroidism</td>
<td>12</td>
</tr>
<tr>
<td>Insulin-dependent diabetes mellitus</td>
<td>1</td>
</tr>
<tr>
<td>Hypopituitarism</td>
<td>&lt;1</td>
</tr>
<tr>
<td>Diabetes insipidus</td>
<td>&lt;1</td>
</tr>
<tr>
<td>Nonendocrine</td>
<td></td>
</tr>
<tr>
<td>Malabsorption syndromes</td>
<td>25</td>
</tr>
<tr>
<td>Alopecia totalis or areata</td>
<td>20</td>
</tr>
<tr>
<td>Pernicious anemia</td>
<td>16</td>
</tr>
<tr>
<td>Chronic active hepatitis</td>
<td>9</td>
</tr>
<tr>
<td>Vitiligo</td>
<td>4</td>
</tr>
<tr>
<td>APS, Type II</td>
<td></td>
</tr>
<tr>
<td>Endocrine</td>
<td></td>
</tr>
<tr>
<td>Adrenal insufficiency</td>
<td>100</td>
</tr>
<tr>
<td>Autoimmune thyroid disease</td>
<td>70</td>
</tr>
<tr>
<td>Insulin-dependent diabetes mellitus</td>
<td>50</td>
</tr>
</tbody>
</table>
immunodeficiency syndrome (AIDS); adrenalitis may occur after infection with cytomegalovirus or atypical mycobacterium, and Kaposi's sarcoma may result in adrenal replacement. The onset is often insidious, but if tested, over 10% of patients with AIDS demonstrate a subnormal cortisol response following a short Synacthen test. Adrenal insufficiency may be precipitated through the concomitant administration of appropriate anti-infectives such as ketoconazole (inhibits cortisol synthesis) or rifampin (increases cortisol metabolism). Rarely, patients with AIDS and features of adrenal insufficiency are found to have elevated circulating ACTH and cortisol concentrations that are not suppressed normally by low-dose dexamethasone administration. This is thought to reflect an acquired form of glucocorticoid resistance related to reduced GR affinity, but the underlying cause remains unknown.

Miscellaneous.

With the exception of tuberculosis and autoimmune adrenal failure, other causes of Addison's disease are rare (see Table 14-16). Adrenal metastases (commonest primary being lung and breast) are often found at postmortem examination, but adrenal insufficiency resulting from these is uncommon, perhaps because over 90% of the adrenal cortex needs to be compromised before symptoms and signs become apparent. Necrosis of the adrenals related to intra-adrenal hemorrhage should be considered in any severely sick patient, particularly those with underlying infection, trauma, or coagulopathy. Intra-adrenal bleeding may be found with any cause of severe septicemia, particularly in children, in whom a common cause is infection with Pseudomonas aeruginosa. When Addison's disease is caused by meningococcus, the association with adrenal insufficiency is known as the Waterhouse-Friderichsen syndrome. Adrenal replacement may also occur with amyloidosis and hemochromatosis.

Adrenal hypoplasia congenita is an X-linked disorder comprising congenital adrenal insufficiency and hypogonadotropic hypogonadism. The condition is caused by mutations in the dosage-sensitive sex reversal, adrenal hypoplasia congenita, X-chromosome factor (DAX-1) gene, a member of the nuclear receptor family of unknown function that is expressed in the adrenal cortex, gonads, and hypothalamus. Mutations in another transcription factor, steroidogenic factor-1, also result in adrenal insufficiency related to lack of development of a functional adrenal cortex. The transcriptional regulation of many P450 steroidogenic enzymes is dependent upon steroidogenic factor-1. Congenital adrenal hypoplasia may also occur in association with glycogen kinase deficiency and muscular dystrophy.

Adrenoleukodystrophy has a prevalence of 1 in 20,000 and is a cause of childhood adrenal insufficiency in association with demyelination within the nervous system related to a failure of beta oxidation of fatty acids within peroxisomes because of reduced activity of very-long-chain acyl-CoA synthetase (VLCs). Increased accumulation of very-long-chain fatty acids occurs in many tissues, and serum assays can be used diagnostically. Only males have the fully expressed condition, and carrier females are usually normal. Several forms are recognized: a childhood cerebral form (30% to 40% of cases), adult adrenomyeloneuropathy (40% of cases), and Addison's disease only (7% of cases).

The childhood-onset form occurs at 5 to 10 years of age with eventual progression to a blind, mute, and severely spastic tetraplegic state. Adrenal insufficiency is usually present but does not appear to correlate with the neurologic deficit. Nevertheless, this is the commonest form of adrenal insufficiency in a child younger than 7 years. Adrenoleukomyelopathy, by contrast, arises later in life with the gradual development of spastic paresis and peripheral neuropathy. Both the childhood and adult conditions result from mutations in a gene on chromosome Xq28 that encodes a peroxisomal membrane protein with homology to the adenosine triphosphatase-binding membrane transporter proteins. At present, there are no genotype-phenotype correlations and it is uncertain how these mutations affect the activity of VLCs. Treatment is unsuccessful. Monounsaturated fatty acids that block the synthesis of the saturated very-long-chain fatty acids have been used. A combination of erucic acid and oleic acid (Lorenzo's oil) has led to normal levels of very-long-chain fatty acids, but this has not altered the rate of neurologic deterioration. However, more promising results have been obtained after bone marrow transplantation.

Familial glucocorticoid deficiency is a rare, autosomal recessive cause of hypothalamic adenohypophyseal dysfunction that is usually seen in childhood. The renin-angiotensin-aldosterone axis is intact, and children usually present either with neonatal hypoglycemia or later with increasing pigmentation, often with enhanced growth velocity. Patients have glucocorticoid deficiency with very high plasma ACTH levels; this occurs because of mutations in the melanocortin-2 or ACTH receptor (MC2R) on chromosome 18p11.2 in some but not all cases. A variant is called triple A or Allgrove's syndrome and refers to the triad of adrenal insufficiency related to ACTH resistance, achalasia, and alacrima. It is not caused by mutations in the MC2R; studies have mapped this disease to chromosome 12q13, and mutations in a novel gene have been reported.

Secondary Hypoadrenalism (Adrenocorticotropic Hormone Deficiency)

This is a common clinical problem and is most often due to sudden cessation of exogenous glucocorticoid therapy. Such therapy suppresses the hypothalamic-pituitary-adrenal axis with consequent adrenal atrophy, and this may last for months after stopping glucocorticoid treatment. Adrenal atrophy and subsequent deficiency should be anticipated in any subject who has taken more than the equivalent of 30 mg of hydrocortisone per day orally (7.5 mg/day prednisolone or 0.75 mg/day dexamethasone) for more than 3 weeks. In addition to the magnitude of the dose of glucocorticoid, the timing of administration of the dose may affect the degree of adrenal suppression. Thus, prednisolone in a dose of 5 mg given last thing at night and 2.5 mg in the morning produces more marked suppression of the hypothalamic-pituitary-adrenal axis than 2.5 mg at night and 5 mg in the morning because the larger evening dose blocks the early morning surge of ACTH. Secondary hypothalamic dysfunction may also occur after failure to give adequate glucocorticoid replacement therapy for intercurrent stress in a patient who has received long-term glucocorticoid therapy. Other causes of secondary adrenal insufficiency (see Table 14-16) reflect inadequate ACTH production from the anterior pituitary gland. In many of these, other pituitary hormones are deficient in addition to ACTH and the patient presents with partial or complete hypopituitarism. The clinical features of hypopituitarism make a relatively easy diagnosis (see also Chapter 8). Isolated ACTH deficiency is rare but a difficult diagnosis to make. It may occur in patients with lymphocytic hypophysitis. A rare but fascinating cause is related to a defect in the normal post-translational processing of POMC to ACTH by the prohormone convertase enzymes (PC1 and PC2). Such patients may have more generalized defects in peptide processing (e.g., cleavage of proinsulin to insulin) giving rise to diabetes mellitus. Patients have also been reported with mutations in the POMC gene that interrupt the synthesis of ACTH and causes ACTH deficiency. The elucidation of the phenotype of these cases, however, has uncovered a novel role for POMC peptides in regulating appetite and hair color. A central role for -MSH in regulating food intake through the hypothalamic

melandocortin-4 receptor has been established. Thus, in addition to adrenal insufficiency, mutations in the POMC gene result in severe obesity and red hair pigmentation. In recombinant mice lacking the POMC gene, the obese phenotype can be reversed by giving an -MSH agonist peripherally.

Secondary hypothalamic dysfunction is also observed in patients with Cushing's disease after successful and selective removal of the ACTH-secreting pituitary adenoma. The function of adjacent "normal" pituitary corticotrophs is suppressed and may remain so for many months after surgery.

Patients have primary adrenal failure usually have both glucocorticoid and mineralocorticoid deficiency. In contrast, those with secondary adrenal insufficiency have an intact renin-angiotensin-aldosterone system. This accounts for differences in salt and water balance in the two groups of patients, which in turn result in different

Clinical Features of Adrenal Insufficiency

Patients with primary adrenal failure usually have both glucocorticoid and mineralocorticoid deficiency. In contrast, those with secondary adrenal insufficiency have an intact renin-angiotensin-aldosterone system. This accounts for differences in salt and water balance in the two groups of patients, which in turn result in different

Gonadal failure | 550
Diabetes insipidus | <1
Nonendocrine | 
Villigo | 4
Alopecia, pemphigus anemia, myasthenia gravis, immune thrombocytopenia purpura, Sjögren's syndrome, rheumatoid arthritis | <1
clinical presentations. The most obvious feature that differentiates primary from secondary hypoadrenalism is skin pigmentation (Table 14-19 and Fig. 14-30), which is nearly always present in primary adrenal insufficiency (unless of short duration) and absent in secondary insufficiency. The pigmentation is seen in sun-exposed areas, recent rather than old scars, axillae, nipples, palmar creases, pressure points, and in mucous membranes (buccal, vaginal, vulval, anal). The cause of the pigmentation has long been debated, but it is thought to reflect increased stimulation of the melanocortin-2 receptor by ACTH itself. In autoimmune Addison’s disease there may be associated vitiligo (see Fig. 14-30).

The clinical features are related to the rate of onset and severity of adrenal deficiency. In many cases, the disease has an insidious onset and a diagnosis is made only when the patient presents with an acute crisis during an intercurrent illness. Acute adrenal insufficiency or an adrenal or addisonian crisis is a medical emergency manifesting as hypotension and

<table>
<thead>
<tr>
<th>TABLE 14-19 – Clinical Features of Primary Adrenal Insufficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Symptom, Sign, or Laboratory Finding</strong></td>
</tr>
<tr>
<td>Weakness, tiredness, fatigue</td>
</tr>
<tr>
<td>Anorexia</td>
</tr>
<tr>
<td>Gastrointestinal symptoms</td>
</tr>
<tr>
<td>Nausea</td>
</tr>
<tr>
<td>Vomiting</td>
</tr>
<tr>
<td>Constipation</td>
</tr>
<tr>
<td>Abdominal pain</td>
</tr>
<tr>
<td>Diarrhea</td>
</tr>
<tr>
<td>Salt craving</td>
</tr>
<tr>
<td>Postural dizziness</td>
</tr>
<tr>
<td>Muscle or joint pains</td>
</tr>
<tr>
<td>Sign</td>
</tr>
<tr>
<td>Weight loss</td>
</tr>
<tr>
<td>Hyperpigmentation</td>
</tr>
<tr>
<td>Hypotension (&lt;110 mm Hg systolic)</td>
</tr>
<tr>
<td>Vitiligo</td>
</tr>
<tr>
<td>Auricular calcification</td>
</tr>
<tr>
<td>Laboratory Finding</td>
</tr>
<tr>
<td>Electrolyte disturbances</td>
</tr>
<tr>
<td>Hyponatremia</td>
</tr>
<tr>
<td>Hyperkalemia</td>
</tr>
<tr>
<td>Hypercalcemia</td>
</tr>
<tr>
<td>Azotemia</td>
</tr>
<tr>
<td>Anemia</td>
</tr>
<tr>
<td>Eosinophilia</td>
</tr>
</tbody>
</table>

Acute circulatory failure (Table 14-20). Anorexia may be an early feature, which progresses to nausea, vomiting, diarrhea, and sometimes abdominal pain. Fever may be present, and hypoglycemia may occur. Patients presenting acutely with adrenal hemorrhage have hypotension; abdominal, flank, or lower chest pain; anorexia; and vomiting. The condition is difficult to diagnose, but evidence of occult hemorrhage (rapidly falling hemoglobin), progressive hyperkalemia, and shock should alert the clinician to the diagnosis.

Alternatively, the patient may present with vague features of chronic adrenal insufficiency: weakness, tiredness, weight loss, nausea, intermittent vomiting, abdominal pain, diarrhea or constipation, general malaise, muscle cramps, arthralgia, and symptoms suggestive of postural hypotension (see Table 14-19). Salt craving may be a feature, and there may be a low-grade fever. Supine blood pressure is usually normal, but almost invariably there is a fall in blood pressure on standing. Although adrenal androgen secretion is lost, this is clinically more apparent in women, who may complain of loss of axillary and pubic hair. Psychiatric symptoms may occur in long-standing cases and include memory impairment, depression, and psychosis. Patients may be inappropriately diagnosed as suffering from chronic fatigue syndrome or anorexia nervosa. These features regress upon treatment with replacement corticosteroids.

In secondary adrenal insufficiency associated with hypopituitarism, the presentation may be related to deficiency of hormones other than ACTH, notably LH or FSH (infertility, oligomenorrhea or amenorrhea, poor libido) and TSH (weight gain, cold intolerance). Fasting hypoglycemia occurs because of loss of the gluconeogenic effects of cortisol. It is rare in adults unless there is concomitant alcohol abuse or additional GH deficiency. However, hypoglycemia is a common presenting feature of ACTH or adrenal insufficiency in childhood. In addition, patients with ACTH deficiency present with malaise, weight loss, and other features of chronic adrenal insufficiency. Rarely, the presentation may be more acute in patients with pituitary apoplexy.

Investigation of Hypoadrenalism

Routine Biochemical Profile

In established primary adrenal insufficiency, hyponatremia is present in about 90% of cases and hyperkalemia in 65%. The blood urea concentration is usually elevated. Hyperkalemia occurs because of aldosterone deficiency and is therefore usually absent in patients with secondary adrenal failure. Hyponatremia may be depletional in an addisonian crisis, but in addition vasopressin levels are elevated, resulting in increased free water retention. Thus, in secondary adrenal insufficiency there may be a dilutional hyponatremia with normal or low blood urea. Reversible abnormalities in liver transaminases frequently occur. Hypercalcemia occurs in 6% of all cases and may be particularly marked in patients with coexisting thyrotoxicosis. However, free thyroxine concentrations are usually low or normal but TSH values are frequently moderately elevated. This is a direct effect of glucocorticoid deficiency and reverses with replacement therapy. Persistent elevation of TSH in association with positive thyroid autoantibodies suggests concomitant autoimmune thyroid disease.

Mineralocorticoid Status

In primary hypoadrenalism, mineralocorticoid deficiency usually occurs with elevated plasma renin activity and either low or low normal plasma aldosterone. The investigation of zona glomerulosa activity is frequently neglected in Addison’s...
Clinical suspicion of the diagnosis should be confirmed with definitive diagnostic tests. Basal plasma cortisol and urinary free cortisol levels are often in the low normal range and cannot be used to exclude the diagnosis. However, a basal cortisol value greater than 400 nmol/L (15 µg/dL) invariably indicates an intact HPA axis. In practice, rather than waiting for results of insensitive basal tests, all patients suspected of having adrenal insufficiency should have an ACTH stimulation test, although in patients with an Addisonian crisis treatment should be instigated immediately and stimulation tests conducted at a later stage.

The ACTH stimulation test involves intramuscular or intravenous administration of 250 µg of tetracosactrin (Synacthen), comprising the first 24 amino acids of normally secreted ACTH 1 to 39. Plasma cortisol levels are measured at 0 and 30 minutes after ACTH, and a normal response is defined by a peak plasma cortisol level greater than 525 nmol/L (19 µg/dL). This value equates to the fifth percentile response in normal subjects but is assay-dependent, with different cortisol radioimmunoassays giving different results. Incremental responses (i.e., the difference between peak and basal values) are of no value in defining a pass response and should not be used. Response is unaffected by the time of day of the test, and the test can be performed in patients who have commenced corticosteroid replacement therapy provided this is of short duration and does not include hydrocortisone (which would cross-react in the cortisol assay).

A prolonged ACTH stimulation test, involving the administration of depot or intravenous infusions of tetracosactrin for 24 to 48 hours, differentiates primary from secondary hypoadrenalism. In normal subjects the plasma cortisol at 4 hours is greater than 1000 nmol/L (36 µg/dL), and beyond this time there is no further increase. Patients with secondary hypoadrenalism show a delayed response with usually a much higher value at 24 and 48 hours than at 4 hours, but in primary hypoadrenalism there is no response at either time. However, the test is now rarely required if plasma ACTH has been appropriately measured at baseline. In primary adrenal insufficiency the ACTH level is disproportionately elevated in comparison with plasma cortisol.

Although there is agreement about the investigation of suspected primary adrenal failure, the diagnosis of secondary hypoadrenalism, notably in patients with existing hypothalamic or pituitary disease, is contentious. Based on correlations with the response of circulating cortisol to surgery, the insulin-induced hypoglycemia test or insulin tolerance test was introduced over 30 years ago as a laboratory test to assess the integrity of the HPA axis and should be considered the "gold standard" in this regard. It should not be performed in patients with ischemic heart disease (always check an electrocardiogram before the test), epilepsy, or severe hypopituitarism (that is, < 9 µg plasma cortisol less than 180 nmol/L [6 µg/dL]). The test involves intravenous administration of soluble insulin in a dose of 0.1 to 0.15 U/kg body weight with measurement of plasma glucose at 0, 30, 45, 60, 90, and 120 minutes. Adequate hypoglycemia (blood glucose less than 39 µg/dL [2.2 mmol/L] with signs of neuroglycopeniasweating and tachycardia) is essential. In normal subjects the peak plasma cortisol exceeds 500 nmol/L (19 µg/dL).

However, the cortisol response to hypoglycemia can be reliably predicted by the response to acute ACTH stimulation (see earlier), a safer, cheaper, and quicker test. This test relies on the principle that the cortisol response to an exogenous bolus of ACTH is determined by the endogenous ACTH trophic drive to the adrenal cortex, impaired ACTH secretion from the anterior pituitary resulting in an impaired cortisol response after Synacthen. However, the ACTH test should not be used to diagnose secondary hypoadrenalism in patients with a recent pituitary insult (surgery, apoplexy). Total hypophysectomy would result in a failed cortisol response to an insulin tolerance test immediately thereafter, but it takes 2 to 3 weeks for the adrenal cortex to readjust to the reduced level of ACTH secretion; in the interim, a false-positive cortisol response would be seen.

The short Synacthen test should also be avoided in patients with a primary diagnosis of Cushing's disease, in whom an exaggerated cortisol response to ACTH may persist. In clinical practice, if the ACTH test is normal, insulin hypoglycemia testing is not necessary in the vast majority of cases unless there is also a need to document endogenous GH reserve in a patient with pituitary disease. In our practice, an insulin tolerance test is performed if, in a patient with suspected hypopituitarism, there is a subnormal response to ACTH. Some patients have an inadequate response to ACTH but then respond normally to hypoglycemia. They do not require corticosteroid replacement therapy. This approach is open to debate, and even taking into account the preceding caveats, false-positive results have been reported for the short Synacthen test; although these are rare (<2% of cases) this should be noted, particularly in patients with ongoing symptoms and signs indicative of hypoadrenalism.

A low-dose ACTH stimulation test giving only 1 µg of ACTH has been proposed as a screen for adequacy of function of the HPA axis with the suggestion that it may be more sensitive than the conventional 250-µg test. Further validation of this test is required to support such a concept.

Two other tests have been advocated to assess adequacy of function of the HPA axis, but their use in modern clinical practice should be restricted to difficult diagnostic cases. In the overnight metyrapone test, metyrapone is given at 30 mg/kg (maximum 3 g) at midnight and plasma cortisol and 11-deoxycortisol are measured at 8 AM the following morning. In patients with an intact axis, ACTH levels rise after the blockade of cortisol synthesis by metyrapone and a normal result is signified by a peak 11-deoxycortisol value greater than 200 nmol/L (7 µg/dL). The CRH stimulation test has been used to diagnose adrenal insufficiency and, unlike the metyrapone test, differentiates primary from secondary causes. Patients with primary adrenal failure have high ACTH levels that rise further after CRH. Conversely, patients with secondary adrenal failure have low ACTH levels that fail to respond to CRH. Patients with hypothalamic disease show a steady rise in ACTH levels after CRH.

Other Tests

Radioimmunoassays to detect autoantibodies such as those against the 21-hydroxylase antigen are now available and should be analyzed in patients with primary adrenal failure. In autoimmune Addison's disease, it is also important to look for evidence of other organ-specific autoimmune disease. A CT scan may reveal enlarged or calcified adrenals, suggesting an infective, hemorrhagic, or malignant diagnosis.
For patients with both primary and secondary adrenal failure, beneficial effects of adrenal androgen replacement therapy with DHEA at 25 to 50 mg/day have been observed. Every patient receiving glucocorticoid therapy should be advised to register for a Medic Alert bracelet or necklace and must carry a Steroid Alert card.

For acute adrenal insufficiency, beneficial effects of adrenal androgen replacement therapy with DHEA at 25 to 50 mg/day have been observed. Every patient receiving glucocorticoid therapy should be advised to register for a Medic Alert bracelet or necklace and must carry a Steroid Alert card.

Chronic Replacement Therapy

The aim is to give replacement doses of hydrocortisone to mimic the normal cortisol secretion rate (Table 14-22). Initially, this was thought to be approximately 25 to 30 mg/day, but stable isotope studies indicate lower normal cortisol production rates of 22 to 41 µmol/day (8 to 15 mg/day). Increasingly, most patients can cope with less than 84 µmol/day (30 mg/day) (usually 15 to 25 mg/day in divided doses). Doses are usually given on waking, with a smaller dose in the late afternoon, but some patients may feel better with doses three times daily. In primary adrenal failure, cortisol day curves with simultaneous ACTH measurements may provide some insight into the adequacy of replacement therapy, but unfortunately there are no good objective tests in secondary adrenal failure. Decisions about doses of replacement therapy are largely based on crude but nevertheless important end points such as weight, well-being, and blood pressure. Bone mineral density may be reduced with conventional hydrocortisone doses of 30 mg/day, highlighting the need to strive for minimally effective but safe doses.

In primary adrenal failure, mineralocorticoid replacement is usually also required in the form of fludrocortisone (or 9-fluorinated hydrocortisone) at 0.05 to 0.2 mg/day. The mineralocorticoid activity of this is about 125 times that of hydrocortisone. After the acute phase has passed, the adequacy of mineralocorticoid replacement should be assessed by measuring electrolytes and supine and erect blood pressure and plasma renin activity; too little fludrocortisone may cause postural hypotension with elevated plasma renin activity and too much may cause the converse. Mineralocorticoid replacement therapy is all too frequently neglected in patients with adrenal failure.

Emergency Measures

1. Establish intravenous access with a large-gauge needle.
2. Draw blood for stat serum electrolytes and glucose and routine measurement of plasma cortisol and ACTH. Do not wait for laboratory results.
3. Infuse 2 to 3 L of 154 mmol/L NaCl (0.9% saline) solution or 50 g/L (5%) dextrose in 154 mmol/L NaCl (0.9% saline) solution as quickly as possible. Monitor for signs of fluid overload by measuring central or peripheral venous pressure and listening for pulmonary rales. Reduce infusion rate if indicated.
4. Inject intravenous hydrocortisone (100 mg immediately and every 6 hr)
5. Use supportive measures as needed.

TABLE 14-21 -- Treatment of Acute Adrenal Insufficiency (Adrenal Crisis)

<table>
<thead>
<tr>
<th>Subacute Measures After Stabilization of the Patient</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Continue intravenous 154 mmol/L NaCl (0.9% saline) solution at a slower rate for next 24 to 48 hr.</td>
</tr>
<tr>
<td>2. Search for and treat possible infectious precipitating causes of the adrenal crisis.</td>
</tr>
<tr>
<td>3. Perform a short ACTH stimulation test to confirm the diagnosis of adrenal insufficiency, if patient does not have known adrenal insufficiency.</td>
</tr>
<tr>
<td>4. Determine the type of adrenal insufficiency and its cause if not already known.</td>
</tr>
<tr>
<td>5. taper glucocorticoids to maintenance dosage over 1 to 3 days, if precipitating or complicating illness permits.</td>
</tr>
<tr>
<td>6. Begin mineralocorticoid replacement with fludrocortisone (0.1 mg by mouth daily) when saline infusion is stopped.</td>
</tr>
</tbody>
</table>

ACTH, adrenocorticotropic hormone.

Patients receiving glucocorticoid replacement therapy should be advised to double the daily dose in the event of intercurrent febrile illness, accident, or mental stress such as an important examination. If the patient is vomiting and cannot take medication by mouth, parenteral hydrocortisone must be given urgently, as indicated earlier. For minor surgery, 50 to 100 mg of hydrocortisone hemisuccinate is given with the premedication. For major operations, this is then followed by the same regimen as for acute adrenal insufficiency (see Table 14-22). Pregnancy proceeds normally in patients taking replacement therapy, but daily doses of hydrocortisone are usually increased modestly (5 to 10 mg/day) in the last trimester. During labor, patients should be well hydrated with a saline drip and receive hydrocortisone at 50 mg intramuscularly every 6 hours until delivery. Thereafter, doses can be rapidly tapered off to usual maintenance regimens.

For patients receiving glucocorticoid therapy who are not receiving corticosteroid therapy, beneficial effects of adrenal androgen replacement therapy with DHEA at 25 to 50 mg/day have been observed.
reported. To date, the reported benefit is principally confined to female patients and includes improvement in sexual function and well-being.
Congenital Adrenal Hyperplasia

These inherited syndromes are caused by deficient adrenal corticosteroid biosynthesis. In each case, there is reduced negative feedback inhibition of cortisol and, depending on the steroidogenic pathway involved, alteration in adrenal mineralocorticoid and androgen secretion (Table 14-23).

21-Hydroxylase Deficiency

Ninety percent of cases of CAH are due to 21-hydroxylase deficiency. In Western societies the incidence varies from 1 in 5000 to 1 in 15,000 live births, but in isolated communities the incidence may be much higher (1 in 300 in Alaskan Eskimos, for example). The condition arises because of defective conversion of 17-hydroxyprogesterone to 11-deoxycortisol. Reduced cortisol biosynthesis results in reduced negative feedback drive and increased ACTH secretion; as a consequence, adrenal androgens are produced in excess (Fig. 14-32). Seventy-five percent of cases have mineralocorticoid deficiency because of failure to convert progesterone to DOC in the zona glomerulosa. Clinically, several distinct variants of 21-hydroxylase deficiency have been recognized (Table 14-24).

Simple Virilizing Form

The enhanced ACTH drive to adrenal androgen secretion in utero leads to virilization of an affected female fetus. Depending on the severity, clitoral enlargement, labial fusion, and development of a urogenital sinus may occur, leading to sexual ambiguity at birth and even inappropriate sex assignment. Rarely, the diagnosis is not made in the neonatal period, especially in boys, who may be phenotypically normal at birth. Such

| Table 14-22 — Treatment of Chronic Primary Adrenal Insufficiency |
|------------------|------------------|
| **Maintenance Therapy** |
| **Glucocorticoid Replacement** |
| Hydrocortisone 1520 mg on awakening and 510 mg in early afternoon. |
| Monitor clinical symptoms and morning plasma ACTH. |
| **Mineralocorticoid Replacement** |
| Fludrocortisone 0.1 (0.050.2) mg orally. |
| Liberal salt intake. |
| Monitor lying and standing blood pressure and pulse, edema, serum potassium, and plasma renin activity. |
| Educate patient about the disease, how to manage minor illnesses and major stresses, and how to inject steroid intramuscularly. |
| Obtain Medic Alert bracelet/necklace, Emergency Medical Information Card. |
| **Treatment of Minor Febrile Illness or Stress** |
| Increase glucocorticoid dose twofold to threefold for the few days of illness; do not change mineralocorticoid dose. |
| Contact physician if illness worsens or persists for more than 3 days or if vomiting develops. |
| No extra supplementation is needed for most uncomplicated, outpatient dental procedures under local anesthesia. General anesthesia or intravenous sedation should not be used in the office. |
| **Emergent Treatment of Severe Stress or Trauma** |
| Inject contents of prefilled dexamethasone (4-mg) syringe intramuscularly. |
| Get to physician as quickly as possible. |
| **Steroid Coverage for Illness or Surgery in Hospital** |
| For moderate illness give hydrocortisone 50 mg twice a day orally or intravenously. Taper rapidly to maintenance dose as patient recovers. |
| For severe illness give hydrocortisone 100 mg intravenously every 8 hr. Taper dose to maintenance level by decreasing by half every day. Adjust dose according to course of illness. |
| For minor procedures under local anesthesia and most radiologic studies, no extra supplementation is needed. |
| For moderately stressful procedures, such as barium enema, endoscopy, or arteriography, give a single 100 mg intravenous dose of hydrocortisone just before the procedure. |
| For major surgery, give hydrocortisone 100 mg intravenously just before induction of anesthesia and continue every 8 hr for first 24 hr. Taper dose rapidly, decreasing by half per day, to maintenance level. |

| Table 14-23 — Congenital Adrenal Hyperplasia: Features for Each Enzyme Defect |
|--------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| **Feature** | **21-Hydroxylase Deficiency** | **11-Hydroxylase Deficiency** | **17-Hydroxylase Deficiency** | **3-Hydroxysteroid Deficiency** | **Lipoid Hyperplasia** | **Aldo-sterone Synthase Deficiency** |
| Defective gene | CYP21 | CYP11B1 | CYP17 | HSDB2 | SIAR | CYP11B2 |
| Chromosomal localization | 8p21.3 | 8q24.3 | 10q24.3 | 1p13.1 | 8p11.2 | 8q24.3 |
| Ambiguous genitalia | +(female) | +(female) | +(male) | +(male) | +(male) | No |
| Acute adrenal insufficiency | + | Rare | No | + | +++ | Salt wasting only |
| Incidence | 1:15,000 | 1:100,000 | Rare | Rare | Rare | Rare |
| Hormones | Glucocorticoids | Reduced | Reduced | Reduced | Corticosterone normal | Reduced | Normal |
| Mineralocorticoids | Reduced | Increased | Reduced | Increased | Reduced | Reduced |
| Androgens | Increased | Increased | Reduced | Reduced (male) | Reduced | Normal |
Increased (female) Elevated metabolite: 17-Hydroxy-progesterone, DOC, 11-deoxy cortisol, B, DOC, DHEA, 17\(^\beta\)-pregnenolone. None B, 18-OHB

Blood pressure, sodium balance: Decreased Increased Decreased Increased Decreased Decreased

Potassium: Increased Decreased Increased Decreased Increased Increased

B, corticosterone; DHEA, dehydroepiandrosterone; DOC, deoxycorticosterone; 18-OHB, 18-hydroxycorticosterone.

Figure 14-32 Congenital adrenal hyperplasia related to 21-hydroxylase deficiency. The normal synthesis of cortisol is impaired, and adrenocorticotropic hormone (ACTH) levels increase because of loss of normal negative feedback inhibition resulting in an increase in adrenal steroid precursors proximal to the block. The results are cortisol deficiency, variable mineralocorticoid deficiency, and excessive secretion of adrenal androgens. DHEA, dehydroepiandrosterone; DOC, deoxycorticosterone; HSD, hydroxysteroid dehydrogenase; StAR, steroidogenic acute regulatory protein.

Patients may present in early childhood with sexual precocity and pubic hair development. Initially, linear growth is accelerated because of premature androgen excess, but if left untreated this stimulates epiphyseal closure and final adult height is invariably diminished. \[472\] \[473\]

Salt-Wasting Form

Seventy-five percent of cases in both sexes also have concomitant aldosterone deficiency. In addition to the preceding features, neonates may present within the first week of life with a salt-wasting crisis and hypotension. Indeed, this may alert the clinician to the diagnosis in a male, but unfortunately the diagnosis is still delayed in many cases and the condition carries a significant neonatal mortality rate.

Cryptic or “Late-Onset” 21-Hydroxylase Deficiency

Patients present in childhood or early adulthood with premature pubarche or with a phenotype that may masquerade as polycystic ovary syndrome (PCOS). \[471\] \[474\] Indeed, late-onset CAH is a recognized secondary cause of PCOS and appears to be commoner than the classical variety. \[474\] In some series from tertiary referral centers, late-onset 21-hydroxylase deficiency may account for up to 12% of all patients with PCOS, but more realistic prevalence rates are probably 1% to 3%. Females present with hirsutism, primary or secondary amenorrhea or anovulatory infertility. \[472\] Androgenic alopecia and acne may be other presenting features. Males may develop enlargement of the testes related to adrenal rests, that is, ectopic adrenal tissue within the testes that regresses after glucocorticoid suppression of ACTH secretion. \[472\]

Heterozygote Deficiency

Salt-wasting, simple virilizing, and late-onset 21-hydroxylase deficiencies are all caused by homozygous or compound heterozygous mutations in the human \(CYP21A2\) gene, whereas in the carrier, heterozygote state, only one allele is mutated. The clinical significance of the heterozygote state is uncertain; it does not appear to affect reproductive capability but may cause signs of hyperandrogenism in women. \[471\]

Two \(CYP21A2\) genes are located within 50 kb of the short arm of chromosome 6 within the major histocompatibility locus, a 3' \(CYP21A2\) (Fig. 14-33). These two genes are closely homologous, and at least 25% of cases of 21-hydroxylase deficiency arise because of unequal crossover and genetic recombination of these two genes at meiosis. Although mutations have been identified within the \(CYP21A2\) gene in affected kindreds (point mutations, gene conversions, and deletions \[472\] \[479\]), the relationship between genotype and phenotype is complex. \[479\] Severe mutations within the gene do not correlate with a severe phenotype either within families or in individual

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Classical Salt Wasting</th>
<th>Simple Virilizing</th>
<th>Nonclassical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at diagnosis</td>
<td>Newborn 6 mo</td>
<td>Newborn 2 yr (female)</td>
<td>Child/Adult</td>
</tr>
<tr>
<td>24 yr (male)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genitalia</td>
<td>Males normal; females ambiguous</td>
<td>Males normal; females ambiguous</td>
<td>Males normal; females virilized</td>
</tr>
<tr>
<td>Incidence</td>
<td>1:20,000</td>
<td>1:60,000</td>
<td>1:1000</td>
</tr>
<tr>
<td>Hormones</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aldosterone</td>
<td>Reduced</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td>Renin</td>
<td>Increased</td>
<td>Normal or increased</td>
<td>Normal</td>
</tr>
<tr>
<td>Cortisol</td>
<td>Reduced</td>
<td>Reduced</td>
<td>Normal</td>
</tr>
<tr>
<td>17-Hydroxyprogesterone</td>
<td>&gt;5000 nmol/L</td>
<td>2500-5000 nmol/L</td>
<td>500-2500 nmol/L (ACTH stimulation)</td>
</tr>
<tr>
<td>Testosterone</td>
<td>Increased</td>
<td>Increased</td>
<td>Variable, increased</td>
</tr>
<tr>
<td>Growth</td>
<td>-23 SD</td>
<td>-12 SD</td>
<td>Probably normal</td>
</tr>
<tr>
<td>21-Hydroxylase activity (% of wild type)</td>
<td>0%</td>
<td>1%</td>
<td>20%-50%</td>
</tr>
<tr>
<td>Typical (CYP21A2) mutations</td>
<td>Deletions, conversions, nt656g</td>
<td>I172N</td>
<td>V281L</td>
</tr>
<tr>
<td>G1108nt, R356W</td>
<td>nt656g</td>
<td>P30L</td>
<td></td>
</tr>
<tr>
<td>I236N, V237E, M239K, Q318X</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ACTH, adrenocorticotropic hormone; SD, standard deviation.
A diagnosis of 21-hydroxylase deficiency should be considered in any newborn infant with genital ambiguity, salt wasting, or hypotension. Hypernatremia and hyperkalemia with raised plasma renin activity are found in salt wasters. In later life, adrenal androgen excess (DHEAS, androstenedione) is found in patients presenting with sexual precocity or a PCOS-like phenotype. 17-Hydroxyprogesterone (17-OHP) is invariably elevated, and clinically useful nomograms have been developed comparing circulating concentrations of 17-OHP before and 60 minutes after exogenous ACTH. This separates patients with classical and nonclassical 21-hydroxylase deficiency from heterozygote carriers and normal subjects, but there is some overlap between values seen in heterozygotes and normal people. 17-OHP is measured basally and then 60 minutes after 250 μg of Synacthen. Stimulated values are invariably grossly elevated in patients with classical and nonclassical varieties (in excess of 35 nmol/L [11 µg/dL]). Heterozygote patients usually have stimulated values between 10 and 30 nmol/L (3 to 9 µg/dL). 

Prenatal diagnosis of 21-hydroxylase deficiency has been advocated because treatment of an affected female may prevent masculinization in utero. 17-OHP can be assayed in amniotic fluid, but the most robust approach is the rapid genotyping of fetal cells obtained by chorionic villus sampling in early gestation. Unlike hydrocortisone, which is inactivated by placental 11-HSD, maternally administered dexamethasone can cross the placenta to suppress the fetal HPA axis. One approach is to advocate dexamethasone therapy as soon as pregnancy is confirmed in high-risk cases and to continue this until the diagnosis is excluded in the fetus. If the fetus is affected, only those of female sex require dexamethasone therapy during gestation. Therapy must be initiated before 8 to 10 weeks of gestation to be effective. However, because only one in eight cases treated in this way has an affected female fetus, the use of steroid therapy in this setting has been questioned. Dexamethasone can lead to maternal cushingoid effects in pregnancy and may in turn have long-term, deleterious effects on the fetus. In patients with known cases requesting fertility (be they male or female), determination of 17-OHP levels through a Synacthen test in the partner before conception uncovers late-onset or heterozygote cases and provides the endocrinologist or geneticist with some assignment of risk before pregnancy.

In the absence of any evidence-based data, there are no prescriptive steroid regimens to treat patients with CAH at any age, and as a result many individualized approaches are used. The overall goal is to replace glucocorticoid and mineralocorticoid, thereby preventing further salt-wasting crises, but also to suppress adrenal androgen secretion so that normal growth and skeletal maturation can proceed. Accurate replacement is essential; glucocorticoids in excess suppress growth, and inadequate replacement results initially in accelerated linear growth but ultimately in short stature because of premature epiphyseal closure. Response is best monitored through growth velocity and bone age, with biochemical markers (17-OHP, DHEAS, testosterone) being useful adjuncts. In difficult cases, a day curve study as described for patients with primary adrenal failure, but measuring the ACTH and 17-OHP response before and after corticosteroid replacement, may confirm overreplacement or underreplacement. Corrective surgery is frequently required (clitoral reduction, vaginoplasty) during childhood.

In late childhood and adolescence, appropriate replacement therapy is equally important. Overtreatment may result in obesity and delayed menarche or puberty with sexual infantilism, whereas underreplacement results in sexual precocity. Compliance with regular medication is often an issue through adolescence. Although much has been written about adequate control in childhood, adults with CAH often provide an ongoing dilemma for the endocrinologist. The follow-up of such patients should involve multidisciplinary clinics, initially with transition adolescence clinics to facilitate transfer from pediatric to adult care. Problems in adulthood are related to fertility concerns, hirsutism and menstrual irregularity in women, obesity and impact of short stature, sexual dysfunction, and psychological problems; counseling is often required in addition to endocrine support.

In women with hyperandrogenism and untreated late-onset CAH, there is no evidence that final height is affected. In this setting, glucocorticoid suppression in isolation rarely controls hirsutism and additional antiandrogen therapy is often required (cyproterone acetate, spironolactone, flutamide together with an oral estrogen contraceptive pill). However, ovulation induction rates with gonadotropin therapy are improved after suppression of nocturnal ACTH levels with 0.25 to 0.5 mg of dexamethasone. Once final height is achieved in adult males, strict control is required only for patients with adrenal rests within the testes or to ensure fertility; inadequate replacement therapy may result in adrenal androgen excess suppressing pituitary FSH secretion and lowering sperm counts.
reported for 21-hydroxylase deficiency, there remains a poor correlation between genotype and phenotype. There is loss of negative cortisol feedback and enhanced ACTH-mediated adrenal androgen excess (Fig. 14-35). Clinical features are therefore similar to those reported in the simple virilizing form of CAH (virilized female fetus, sexual ambiguity), and again milder cases can present later in childhood or even young adulthood. The principal difference compared with 21-hydroxylase deficiency is hypertension, and this is thought to be secondary to the mineralocorticoid effect of DOC excess (see Table 14-23). However, there is a poor correlation between DOC secretion and the presence of hypertension; furthermore, unexplained salt wasting has been reported in a few cases.

On this clinical background, the diagnosis can be made by demonstrating a plasma ACTH-stimulated 11-deoxycorticisol value that is more than three times the 95th percentile for an age-matched normal group. Although established heterozygotes do not demonstrate a rise in 11-deoxycorticisol above normal after Synacthen (unlike the 17-OHP response observed in heterozygote 21-hydroxylase patients), exaggerated ACTH-stimulated responses have been observed in patients with hirsutism and in patients with essential hypertension. Suggesting partial defects in 11-hydroxylase activity. As reported for 21-hydroxylase deficiency, treatment is with replacement glucocorticoid therapy; with suppression of DOC secretion, plasma renin activity (suppressed at baseline) increases into the normal range.

17-Hydroxylase Deficiency

Fewer than 150 cases of 17-hydroxylase deficiency have been reported. Mutations within the CYP17 gene result in the failure to synthesize cortisol (17-hydroxylase activity), adrenal androgens (17,20-lyase activity), and gonadal steroids (Fig. 14-36). Thus, in contrast to 21-hydroxylase and 11-hydroxylase deficiencies, 17-hydroxylase deficiency results in adrenal and gonadal insufficiency. A single enzyme is expressed in adrenal and gonadal tissue that is capable of producing both 17-hydroxylation and 17,20-lyase activities, but patients with isolated deficiency in the hydroxylation of 17-OHP or 17,20-lyase deficiency have rarely been reported. Loss of negative feedback results in increased secretion of steroids proximal to the block, and mineralocorticoid synthesis is enhanced. However, aldosterone levels are variable and the mineralocorticoid excess state that characterizes this condition is thought to be induced by DOC excess in over 80% of cases.

The genetic basis of the disease has been established in many cases, involving point mutations, gene deletions, and conversions in the CYP17 gene. Relative hydroxylase and lyase activities of mutant CYP17 complementary DNAs vary in in vitro transfection assays, but correlations with clinical phenotype are lacking. Thus, patients with clinically pure 17,20-lyase deficiency may have mutant CYP17 complementary DNAs that exhibit compromised 17-hydroxylase activity.

The diagnosis is usually made at the time of puberty when patients present with hypertension, hypokalemia, and hypogonadism, the latter occurring because of lack of CYP17 expression within the gonad and impaired gonadal steroidogenesis. As a result, LH and FSH levels are elevated. Female patients (XX) have primary amenorrhea with absent sexual characteristics, and males (46,XY) have complete pseudohermaphroditism with female external genitalia but absent uterus and fallopian tubes. The intra-abdominal testes should be removed, and such patients are usually reared as female.

Glucocorticoid replacement reverses the DOC-induced suppression of the renin-angiotensin system and lowers blood pressure. Additional sex steroid replacement is required from puberty onward.

3-Hydroxysteroid Dehydrogenase Deficiency

In this rare form of CAH, the secretion of all classes of adrenal and ovarian steroids is impaired because of mutations within the HSD3B2 gene encoding 3-HSDII. Patients usually present in early infancy with adrenal insufficiency. Loss of mineralocorticoid secretion results in salt wasting, although this is absent in 30% to 40% of cases (Fig. 14-37). As with 21-hydroxylase deficiency, absence of salt wasting may delay the presentation into childhood or puberty. The correlation between genotype and phenotype is once again poor; identical mutations have been found in the HSD3B2 gene in both salt wasters and nonsalt wasters. The spectrum of genital development is variable in both sexes. In males, because the 3-HSDII enzyme is also expressed within the gonad, male pseudohermaphroditism may occur with female external genitalia. In milder cases, hypospadias may be found or even normal male genitalia. In females, genital development can be normal but there is usually evidence of mild virilization, presumably because of enhanced adrenal secretion of DHEA, which is converted peripherally to testosterone. A late-onset form has been described in patients with premature pubarche and a PCOS-like phenotype (hirsutism, oligomenorrhea, amenorrhea). Because activity of the 3-HSDII enzyme present in skin and other peripheral tissues is intact, circulating 4 steroid levels (progesterone, 17-hydroxyprogesterone, androstenedione) may be normal (or even increased). However, a diagnosis is established by demonstrating an increased ratio of 5 steroids (pregnenolone, 17-hydroxyprogrenolone, DHEA) to 4 steroids in plasma or urine. ACTH stimulation may be required to detect a late-onset presentation. Treatment is with replacement glucocorticoids, fludrocortisone (if indicated), and sex steroids from puberty onward.

Steroidogenic Acute Regulatory Protein Deficiency

Mutations in the gene encoding STAR result in a failure of transport of cholesterol from the outer to the inner mitochondrial membrane.
membrane in steroidogenic tissues; as a result, there is deficiency of all adrenal and gonadal steroid hormones. Presentation is with acute adrenal insufficiency in the neonatal period, and males exhibit pseudohermaphroditism because of absent gonadal steroids. The condition is fatal in infancy in two thirds of all cases. The adrenal glands are often massively enlarged and full of lipid; prior to the characterization of STAR, the condition was termed congenital "lipoid" hyperplasia and the candidate gene was thought to be cholesterol side-chain cleavage (CYP11A1). In fact, to date, no mutations have been reported in the CYP11A1 gene; such mutations are thought to be lethal in utero. This clinical phenotype is endorsed by recombinant mouse models lacking the STAR gene.

**Apparent Cortisone Reductase Deficiency**

In this condition, adrenal glands become hyperplastic because of ACTH stimulation resulting from a defect in cortisol metabolism rather than an inherent defect within the gland itself. Patients with apparent cortisone reductase deficiency have a defect in the conversion of cortisone to cortisol, suggesting inhibition of 11-oxoreductase activity and, by implication, inhibition of the type 1 isozyme of 11-hydroxysteroid dehydrogenase (11-HSD1). Eight cases have been described; with one exception, all are female. Cortisol clearance is increased, and ACTH secretion is elevated to maintain normal circulating cortisol concentrations but at the expense of adrenal androgen excess. As a consequence, patients described are usually young women who present with hirsutism, menstrual irregularity, or androgenic alopecia.

Dexamethasone treatment to suppress ACTH has been used with some success to control the hyperandrogenism in these cases. Urinary tetrahydro metabolites of cortisol and cortisone show almost exclusively THE with little or no detectable THF or allo-THF (ratio of THF + allo-THF to THE less than 0.05, reference range 0.8 to 0.045).

**TABLE 14-25 -- Clinical and Biochemical Characteristics of Reported Cases of Apparent Cortisone Reductase Deficiency**

<table>
<thead>
<tr>
<th>Age</th>
<th>Sex</th>
<th>Clinical Features</th>
<th>Serum Androgens</th>
<th>THF + alloTHF:THE ratio</th>
<th>Comments</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>28 F</td>
<td>Hirsuitism</td>
<td>Testosterone</td>
<td>Marked fall in serum androgens on treatment with dexamethasone</td>
<td>0.51</td>
<td></td>
<td>506</td>
</tr>
<tr>
<td>17 F</td>
<td>Oligomenorrhea, hirsuitism, acne, obesity</td>
<td>Testosterone</td>
<td>Fall in androgens with dexamethasone treatment although developed cushingoid side effects.</td>
<td>0.045</td>
<td></td>
<td>508</td>
</tr>
<tr>
<td>18 F</td>
<td>Oligomenorrhea, hirsuitism, acne</td>
<td>Testosterone</td>
<td>Sibling of preceding patient. Fall in androgens with treatment.</td>
<td>0.03</td>
<td></td>
<td>509</td>
</tr>
<tr>
<td>30 F</td>
<td>Oligomenorrhea, hirsuitism, infertility</td>
<td>Testosterone</td>
<td>Fall in testosterone with treatment.</td>
<td>0.04</td>
<td></td>
<td>510</td>
</tr>
<tr>
<td>M</td>
<td>Excess body hair (sibling of preceding patient)</td>
<td>Testosterone</td>
<td>Sibling of preceding patient. No mutations on genetic sequence analysis of HSD11B1</td>
<td>0.51</td>
<td></td>
<td>511</td>
</tr>
<tr>
<td>37 F</td>
<td>Obesity, oligomenorrhea, hirsuitism</td>
<td>Testosterone</td>
<td>No mutations on genetic sequence analysis of HSD11B1</td>
<td>0.04</td>
<td></td>
<td>512</td>
</tr>
<tr>
<td>F</td>
<td>Congenital adrenal hyperplasia diagnosed shortly after birth (21-hydroxylase deficiency). 17-Hydroxyprogesterone levels unresponsive to cortisone acetate</td>
<td>Androstenedione</td>
<td>17-OHP levels suppressed completely with prednisolone, indicative of an inability to activate cortisone acetate. No mutations on genetic sequence analysis of HSD11B1</td>
<td>0.51</td>
<td></td>
<td>513</td>
</tr>
<tr>
<td>55 F</td>
<td>Androgenetic alopecia, mild hirsuitism</td>
<td>Testosterone</td>
<td>No mutations on genetic sequence analysis of HSD11B1</td>
<td>0.04</td>
<td></td>
<td>514</td>
</tr>
</tbody>
</table>

DHEAS, dehydroepiandrosterone sulfate; allo-THF, 5-tetrahydrocortisol; THE, tetrahydrocortisone; THF, 5-tetrahydrocortisol.

been described; with one exception, all are female (Table 14-25). Cortisol clearance is increased, and ACTH secretion is elevated to maintain normal circulating cortisol concentrations but at the expense of adrenal androgen excess. As a consequence, patients described are usually young women who present with hirsutism, menstrual irregularity, or androgenic alopecia.

Dexamethasone treatment to suppress ACTH has been used with some success to control the hyperandrogenism in these cases. Urinary tetrahydro metabolites of cortisol and cortisone show almost exclusively THE with little or no detectable THF or allo-THF (ratio of THF + allo-THF to THE less than 0.05, reference range 0.8 to 1.3). Further studies have also shown impaired plasma cortisol concentrations after an oral dose of cortisone acetate. Despite this biochemical evidence implicating a defect in 11-HSD1, investigations have revealed no mutations to date in the HSD11B1 gene in affected cases.

Patients with PCOS share many of the same clinical characteristics as those with apparent cortisone reductase deficiency. Although there is evidence to support increased cortisol secretion rates in PCOS, perhaps indicative of a defect in conversion of cortisone to cortisol, there remains to be a consensus with respect to THF + allo-THF/TESTH ratios. Both normal and reduced ratios have been reported in the literature.
Mineralocorticoid Deficiency

These syndromes are listed in Table 14-26. They can be divided into those that are congenital and others that are acquired. Mineralocorticoid deficiency may occur in some forms of CAH and with other causes of adrenal insufficiency (e.g., Addison’s disease and congenital adrenal hypoplasia) (see preceding).

Primary Defects in Aldosterone Biosynthesis: Aldosterone Synthase Deficiency

Failure of conversion of corticosterone to 18-hydroxycorticosterone or of 18-hydroxycorticosterone to aldosterone usually results in a salt-wasting crisis in neonatal life. Hyperkalemia, metabolic acidosis, dehydration, and hypotension are found. The condition has been called corticosterone methyl oxidase (CMO) deficiency, but this was before the final enzyme or enzymes involved in the conversion of DOC to aldosterone were characterized and cloned. In fact, a single enzyme, aldosterone synthase, carries a multistep reaction involving 11-hydroxylation of DOC to corticosterone and 18-hydroxylation of corticosterone to 18-hydroxycorticosterone followed by 18-dehydrogenation to aldosterone (see Fig. 14-3).

Two variants of CMO deficiency are described: CMO I is characterized by low 18-hydroxycorticosterone and aldosterone levels, whereas patients with CMO II deficiency have hypoaldosteronism but high 18-hydroxycorticosterone levels. In both cases, mutations in the gene encoding aldosterone synthase have been described and the discrepant 18-hydroxycorticosterone levels seem likely to be explained on the basis of variable 18-hydroxylase activity of the related CYP45011-hydroxylase enzyme.

CMO II is much more common in Iranian Jews than the white population.

Defects in Aldosterone Action: Pseudohypoaldosteronism

Pseudohypoaldosteronism type I may occur in neonatal life with respiratory difficulties but is usually found in infancy with severe salt wasting and failure to thrive, with very high plasma aldosterone and plasma renin activity levels and inappropriate urinary sodium loss. The MR appears to be defective, as judged by studies evaluating the binding of aldosterone to monocytes, but molecular studies have failed to show any abnormality in the MR itself. Rather, inactivating mutations in the , and subunits of the epithelial sodium channel have been shown to explain the condition. Acquired forms of pseudohypoaldosteronism can occur in patients after renal transplantation, following obstructive uropathy, and in premature infants.

Pseudohypoaldosteronism type II or Gordon’s syndrome is an autosomal dominant disorder characterized by hyperkalemia

<table>
<thead>
<tr>
<th>TABLE 14-26 -- Causes of Mineralocorticoid Deficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addison’s disease</td>
</tr>
<tr>
<td>Adrenal hypoplasia</td>
</tr>
<tr>
<td>Congenital adrenal hyperplasia (21-hydroxylase and 3-hydroxysteroid dehydrogenase deficiencies)</td>
</tr>
<tr>
<td>Pseudohypoaldosteronism types I and II</td>
</tr>
<tr>
<td>Hyporeninemic hypoaldosteronism</td>
</tr>
<tr>
<td>Aldosterone biosynthetic defects</td>
</tr>
<tr>
<td>Drug induced</td>
</tr>
</tbody>
</table>

but not salt wasting, in contrast to the type I condition. Patients have resistance to the mineralocorticoid effects of aldosterone on tubular potassium transport but not to those of sodium and chloride transport. As a result, affected individuals have hyperkalemia, hypertension, and suppression of plasma renin activity. Recently, deletions in the WNK4 gene (a member of the WNK family of serine-threonine kinases) have been described in affected cases.

Angiotensin II is a key stimulus to aldosterone secretion, and damage or blockade of the renin-angiotensin system may result in mineralocorticoid deficiency. Various renal conditions have been associated with damage to the juxtaglomerular apparatus and hence renin deficiency. These include systemic lupus erythematosus, myeloma, amyloid, AIDS, and use of nonsteroidal anti-inflammatory drugs, but the most common (greater than 75% cases) is diabetic nephropathy.

The usual picture is of an elderly patient with hyperkalemia, acidosis, and mild to moderate impairment of renal function. Plasma renin activity and aldosterone are low and fail to respond to sodium depletion, the erect posture, or furosemide administration. In contrast to those with adrenal insufficiency, patients have normal or elevated blood pressure and no postural hypotension. Muscle weakness and cardiac arrhythmias may also occur. Other factors may contribute to the hyperkalemia, including the use of potassium-sparing diuretics, potassium supplementation, insulin deficiency, and -adrenoceptor blocking drugs and prostaglandin synthetase inhibitors, which inhibit renin release.

Treatment of primary renin deficiency is with fludrocortisone in the first instance together with dietary potassium restriction. However, these patients are not salt depleted and may become hypertensive with fludrocortisone. In such a scenario, the addition of a loop-acting diuretic such as furosemide is appropriate. This increases acid excretion and improves the metabolic acidosis.
Adrenal Adenomas, Incidentalomas, and Carcinomas

Etiology of Adrenal Tumors

The underlying basis for adrenal tumorigenesis is unknown. Clonal analysis suggests progression from a normal to adenomatous to carcinomatous lesion, but the molecular pathways involved remain obscure. Several factors have been associated with malignant transformation including genes encoding p53, p7 cyclin-dependent kinase, menin, IGF-II, MC2R, and inhibin-βA. Mice lacking the inhibin-βA gene develop adrenal tumors through a process that is also gonadotropin-dependent.

Adenomas

Cortisol-secreting adrenal adenomas have been discussed in detail (“Cortisol-Secreting Adrenal Adenoma and Carcinoma”), and aldosterone-secreting adenomas (Conn’s syndrome) are discussed in Chapter 15.

Pure virilizing benign adrenal adenomas are rare, with approximately 50 cases reported in the literature. The majority of cases occur in women; male cases are restricted to childhood, when presentation is with sexual precocity and accelerated bone age. In females, the majority of cases arise before the menopause with marked hirsutism, deepening of the voice, and amenorrhea. Clitoromegaly is found in 80% of cases. Testosterone is usually strikingly elevated, but gonadotropin levels may not be suppressed. By definition, urinary free cortisol is normal. Tumors vary in size and should be treated surgically. Postoperatively, clinical features invariably improve and normal menses return.

Incidentalomas

Autopsy series have defined the prevalence of adrenal adenomas more than 1 cm in diameter to be between 1.5% and 7%. It is perhaps not surprising, therefore, that with the advent of high-resolution imaging procedures (CT, MRI), incidentally discovered adrenal masses have become a common clinical problem. An adrenal mass is uncovered in up to 4% of patients imaged for nonadrenal pathology. Incidentalomas are uncommon in patients younger than 30 years but increase in frequency with age; they occur equally in males and females. In more than 85% of cases these lesions are nonfunctioning, benign adenomas. Occasionally they may represent myelolipomas, hamartomas, or granulomatous infiltrations of the adrenal and result in a characteristic CT or MRI appearance (Fig. 14-38). Functioning tumors (pheochromocytomas or those secreting cortisol, aldosterone, or sex steroids) and carcinomas make up the remainder.

In addition, it is established that some incidentalomas may cause abnormal hormone secretion without obvious clinical manifestations of a hormone excess state; the best example of this is “preclinical” Cushing’s syndrome, which may occur in up to 20% of all cases. This may explain why incidentalomas appear to be common in patients with obesity and diabetes mellitus. As a result, all patients with incidentally discovered adrenal masses should undergo appropriate endocrine screening tests. These should comprise 24-hour urinary catecholamine collection, 24-hour urinary free cortisol, and overnight dexamethasone suppression tests. Because of the reported poor sensitivity of serum potassium measurements in detecting primary aldosteronism, our practice has been to measure supine circulating plasma renin activity and aldosterone levels. DHEAS should be measured as a marker of adrenal androgen secretion. Low levels may occur in patients with suppressed ACTH concentrations related to autonomous cortisol secretion from the adenoma, and it is important that DHEAS not be measured during the overnight dexamethasone study. Some studies have also documented high levels of 17-OHP after ACTH stimulation tests, suggesting partial defects in 21-hydroxylase in some tumors.

The possibility of malignancy should be considered in each case. In patients with a known extra-adrenal primary, the incidence of malignancy is obviously much higher (up to 20% of patients with lung cancer, for example, have adrenal metastases on CT scanning). In those with no evidence of malignancy, adrenal carcinoma is rare; in one study, only 26 of 630 incidentalomas were found to be adrenal carcinomas. In true incidentalomas, size appears to be predictive of malignancy; a lesion less than 5 cm in diameter is unlikely to be malignant. The majority of nonfunctioning lesions less than 5 cm can therefore be treated conservatively and patients followed up with annual imaging. Even incidentalomas larger than 5 cm are more likely to be benign than malignant, but because of an increased risk of malignancy many centers recommend removal of tumors more than 5 cm in diameter, preferably by laparoscopic adrenalectomy. Additional characteristic MRI appearances or scintigraphy studies may aid in differentiating malignant from nonmalignant lesions. CT-guided biopsy is useful in differentiating adrenal from nonadrenal tissue in the case of a suspected metastasis but is poor in differentiating benign adenomas from malignant adrenal lesions.

Carcinomas

Primary adrenal carcinoma is rare, with an incidence of 1 per million population per year. Women are more commonly affected than men (2:1); mean age of onset is 40 to 50 years, although men tend to be older at presentation. Eighty percent of tumors are functional, most commonly secreting glucocorticoids alone (45%), glucocorticoids and androgens (45%), or androgens alone (10%). Less than 1% of all cases secrete aldosterone. Patients present with features of the hormone excess state (glucocorticoid or androgen excess, or both), but abdominal pain, weight loss, anorexia, and fever occur in 25% of cases. An abdominal mass may be palpable.

Current treatments for what is often an aggressive tumor are poor. Surgery offers the only chance of cure for patients with local disease, but metastatic spread is evident in 75% of cases at presentation. Radiotherapy is ineffective, as are most chemotherapeutic regimens. Mitotane in high doses offers transient benefit in reducing tumor growth in 25% to 30% of cases and controlling hormonal hypersecretion in 75% of cases. Overall, the prognosis is poor, with 5-year survival rates of less than 20%.
References


4. Osler W. On six cases of Addison’s disease with the report of a case greatly benefited by the use of suprarenal extract. Int Med Mag 1896.


38. Lorence MC, Murray BA, Trant JM, Mason JI. Human 3-hydroxysteroid dehydrogenase\(^{3,4}\) isomerase from placenta: expression in nonsteroidogenic cells of a protein that catalyses the dehydrogenation/isomerization of C21 and C19 steroids. Endocrinology 1990;126:24932498.


35. Enyeart JJ, Mlinar B, Enyeart TA. Type Ca2+ channels are required for adrenocorticotropin-stimulated cortisol production by bovine adrenal zona fasciculata cells. Mol Endocrinol 1993; 7:10311040.
References

151. Kroczowski ZS, Funder JW. Renal mineralocorticoid receptors and hippocampal corticosterone-binding species have identical intrinsic steroid specificity. Proc Natl Acad Sci USA 1983; 80:66656665.


References


References


467. Fiad TM, Conway JD, Cunningham SK, McKenna TJ. The role of plasma renin activity in evaluating the adequacy of mineralocorticoid replacement in primary adrenal insufficiency. Clin Endocrinol (Oxf) 1996; 45:529534.


Hypertension is a common disorder, occurring in approximately 20% of the United States population. The great majority of hypertensive subjects have the diagnosis of essential or primary hypertension. Essential hypertension is a heritable syndrome reflecting a variety of pathophysiologic abnormalities that can lead, independently or together, to an elevated arterial blood pressure. Although secondary causes exist in a smaller percentage (10%) of hypertensive subjects, they still represent a large number of patients.

Broadly speaking, the secondary causes of hypertension can be divided into renal causes (e.g., parenchymal or renovascular disease) and endocrine causes. In some disorders, many cases can be diagnosed by an astute clinician because the signs and symptoms are often distinct (e.g., pheochromocytoma and Cushing's syndrome). In addition, hypertension refractory to antihypertensive treatment may prompt the physician to screen for secondary causes. The age and sex of the hypertensive patient may also be helpful in the diagnosis of disorders with the secondary etiologies. For example, fibromuscular hyperplasia and Cushing's syndrome are more commonly seen in younger females, whereas primary hypothyroidism most commonly occurs in older female patients. Finally, making a diagnosis of a secondary disorder is gratifying because it may lead to significant amelioration or in some instances cure of the elevated blood pressure.
PHYSIOLOGY OF THE SYMPATHOADRENAL SYSTEM AND PHEOCHROMOCYTOMA

The autonomic nervous system consists of the parasympathetic nervous system and the sympathoadrenal system. The neurotransmitter in the parasympathetic nervous system is primarily acetylcholine, and the neurotransmitters in the sympathoadrenal system include norepinephrine at sympathetic nerve endings in the periphery and central nervous system and epinephrine, which is secreted by the adrenal medulla into the systemic circulation. Another catecholamine, dopamine, acts primarily as a neurotransmitter in the central nervous system but is also secreted from peripheral sympathetic nerve endings. The sympathetic nervous system is under direct control of the central nervous system, allowing rapid onset of actions of short duration as a result of the abbreviated half-lives of catecholamines.

Structure and Organization of the Sympathoadrenal System

The sympathoadrenal system, which is composed of the ganglia of the sympathetic nervous system and the adrenal medulla, is embryologically derived from neural crest tissue. The precursor sympathogonia differentiate into neuroblasts, ultimately giving rise to the paravertebral and preaortic ganglion cells. Sympathetic preganglionic axons arise in large part from cells located in the thoracolumbar spinal cord. These preganglionic sympathetic neurons in turn have synapses with descending tracks from neurons in the pons, medulla, and hypothalamus, allowing regulation of sympathetic activity by the brain (Fig. 15-1; see also Fig. 15-16). Thus, the limbic system and cortex can also regulate sympathetic activity by connections with central nuclei in the hypothalamus and medulla. In turn, these central nervous system neurons that influence sympathetic activity are regulated by a variety of factors including substrates (glucose) and hormones (corticotropin-releasing hormone).

The axons of the preganglionic neurons synapse with postganglionic cell bodies located in the paravertebral and preaortic ganglia as well as neurons in the celiac and superior and inferior mesenteric ganglia (Fig. 15-2). Postganglionic axons from the cell bodies located in these ganglia in turn innervate the visceral organs. The splanchnic outflow of the lower thoracic and lumbar preganglionic axons also directly innervates the cells of the adrenal medulla (see Fig. 15-2). Acetylcholine is the neurotransmitter at the ganglionic synapses and at the adrenal medullary neurons. In different tissues, the postganglionic sympathetic innervation can be cholinergic or noradrenergic, or both. For example, at arteriolar synapitc clefts norepinephrine is released from postganglionic nerve terminals, and sweat glands have sympathetic cholinergic innervation. Adrenergic nerves also contain other mediators including the peptide substance P, neuropeptide Y, somatostatin, and chromogranin A. These substances, which are released in peripheral and central adrenergic nerves, may have synergistic or direct actions with norepinephrine on effector cells. On the other hand, the adrenal medullary cells that secrete epinephrine into the systemic circulation are innervated by the splanchnic outflow of cholinergic preganglionic neurons.

The arterial blood supply to the adrenal gland is derived from the aorta (middle adrenal artery), the inferior phrenic artery, and the renal artery. Adrenal blood flow drains centrally toward the medulla, eventually forming a single adrenal vein, which drains into the renal vein on the left and into the vena cava on the right. Although the existence of a corticomedullary portal system remains controversial, it is likely that local high concentrations of glucocorticoids influence the biosynthesis of epinephrine by induction of the enzyme phenylethanolamine N-methyltransferase (PNMT).
Catecholamines

All naturally occurring catecholamines contain a catechol nucleus (Fig. 15-3). Epinephrine is synthesized and stored in the adrenal medulla and released into the systemic circulation. Norepinephrine is synthesized and stored at peripheral nerve endings. Dopamine acts primarily as a neurotransmitter in the central nervous system, although epinephrine and norepinephrine also act as central nervous system neurotransmitters.

Catecholamines act widely in the body and affect many cardiovascular and metabolic processes. Specific catecholamine receptors mediate the biologic actions of these compounds. The amount of endogenous catecholamines released at peripheral nerve endings and the plasma concentration of epinephrine are the main determinants of the physiologic responses to activation of the sympathetic nervous system. Identification of -adrenergic and -adrenergic receptors and their receptor subtypes (1, 2, 3, and 4) in target tissue has led to an understanding of the physiologic responses to exogenous and endogenous administration of catecholamines.

Moreover, the pharmacologic development of selective - and -adrenergic antagonists has added a wide range of treatments for a variety of clinical disorders. For example, -agonists (terbutaline and albuterol), among their actions, can cause bronchial smooth muscle relaxation and are commonly prescribed in aerosol formulation for the treatment of bronchial asthma. On the other hand, -antagonists (such as atenolol and metoprolol) are considered standard therapies for angina pectoris, hypertension, and cardiac arrhythmias. The - and -adrenergic receptors on cell surfaces reciprocally increase or decrease in response to the receptor-specific agonist concentration. Inhibition or stimulation of the intracellular adenylate cyclase (cyclic adenosine monophosphate [cAMP]) system mediates the majority of responses to receptor subtypespecific agonists.

Catecholamine Synthesis

Catecholamines are formed from the amino acid tyrosine by a process of hydroxylation and decarboxylation (see Fig. 15-3). This process of amine precursor uptake and decarboxylation (APUD) is a feature of neuroendocrine tissues that have a common origin. Most reactions occur in the cytoplasm except for hydroxylation of dopamine into norepinephrine, which occurs in the secretory vesicle.

The rate-limiting step in catecholamine biosynthesis is the conversion of tyrosine to 3,4-dihydroxyphenylalanine (dopa) by the enzyme tyrosine hydroxylase (TH). The reaction requires tyrosine as substrate and oxygen, iron (Fe2+), and tetrahydrobiopterin as cofactors. Tyrosine hydroxylase is expressed only in neuronal tissues that synthesize catecholamines, and several factors regulate its activity. The intraneuronal or intracellular transport of tyrosine may be affected by other amino acids or drugs that compete for transport or act as competitive inhibitors of the transport system, such as -methylparatyrosine. Increased intracellular levels of catechols down-regulate the activity of the enzyme. As catechols are released from secretory granules in response to a stimulus, cytoplasmic catecholamines are depleted and the feedback inhibition of tyrosine hydroxylase is released. Four isoforms of tyrosine hydroxylase exist. Transcription is stimulated by glucocorticoids, cAMP-dependent protein kinases, Ca2+ dependent protein kinase, and Ca2+ -calmodulin-dependent protein kinase.

Aromatic L-amino acid decarboxylase (AADC) catalyzes the decarboxylation of dopa to dopamine, a process that can occur in any APUD tissue in which dopa is present. AADC is not specific for dopa. For example, decarboxylation of 5-hydroxytryptophan produces serotonin. Pyridoxal 5-phosphate is the cofactor required.

Catecholamine Uptake and Release

Catecholamines are taken up into storage vesicles by active transport using an H+ + ATP-driven proton pump and carrier proteins, vesicular monoamine transporters (VMATs). The ATP-driven pump maintains a steep electrical gradient. For every monoamine transported, ATP is hydrolyzed and two hydrogen ions are transported from the vesicle into the cytosol. Calcium is also maintained in high concentration within the vesicle. Labeled 131Iodoamphetamine (131I-MIBG) appears to be imported by VMATs into the storage vesicles in the adrenal medulla, which makes imaging with MIBG useful for evaluation of pheochromocytomas.
Catecholamine uptake, as well as MIBG, is inhibited by reserpine.  

Acetylcholine originating from preganglionic sympathetic fibers stimulates nicotinic cholinergic receptors and causes depolarization of the adrenomedullary chromaffin cell. Depolarization leads to activation of voltage-gated Ca\(^{2+}\) channels resulting in exocytosis of secretory vesicle contents. A Ca\(^{2+}\)-sensing receptor appears to be involved in the process of exocytosis. During exocytosis, all of the granular contents are released into the extracellular space.

**Catecholamine Metabolism**

Metabolism of catecholamines occurs through two enzyme pathways (Fig. 15-4). Catechol-O-methyltransferase (COMT) is found primarily outside neuronal tissue and converts epinephrine to metanephrine and norepinephrine to normetanephrine by meta-\(\text{O}\)-methylation. S-Adenosylmethionine is used as the methyl donor, and Ca\(^{2+}\) is required.

Metanephrine and normetanephrine are oxidized by monoamine oxidase (MAO) to vanillylmandelic acid (VMA) by oxidative deamination. MAO may also oxidize epinephrine and norepinephrine to 3,4-dihydroxymandelic acid, which is then converted by COMT to VMA. MAO is located on the outer membrane of mitochondria. In the storage vesicle, norepinephrine is protected from metabolism by MAO. MAO action may play an important role in regulating the metabolism of norepinephrine and dopamine. Intravesical stores of norepinephrine increase when MAO is inhibited.
Pheochromocytoma

Incidence and Importance

Pheochromocytomas are tumors of neuroectodermal origin arising from chromaffin cells. They are named for the dark staining reaction that is caused by the oxidation of intracellular catecholamine stores on exposure to dichromate salts. Although pheochromocytomas are a rare cause of hypertension, failure to recognize and treat a pheochromocytoma could prove a fatal oversight. Reportedly, less than 1% of patients who are evaluated for hypertension have pheochromocytomas, but this may be an underestimate. In one series, the incidence rate was calculated to be 0.8 per 100,000 person-years. In an autopsy series, approximately half of the pheochromocytomas were diagnosed at postmortem examination, demonstrating that this disorder is frequently not recognized.

Tumors that arise from chromaffin cells of the adrenal medulla are referred to as pheochromocytomas, and those that arise in paraganglia are termed paragangliomas or extra-adrenal pheochromocytomas. The paraganglia are collections of specialized neural crest cells that have migrated to their final destination throughout the body. Tumors can arise from sympathetic ganglia located from the neck to the bladder as well as the carotid body, vagal body, mediastinum, aorta, organs of Zuckerkandl, and pelvis (the most common site). It is commonly believed that the malignant potential is higher for extra-adrenal tumors than intra-adrenal tumors. However, some studies suggest that disease-free survival is similar for patients with extra-adrenal and intra-adrenal tumors.

Clinical Manifestations

Hypertension is the most common clinical manifestation of pheochromocytoma and is present in 90% to 100% of patients. Sustained hypertension is seen in approximately half, paroxysmal hypertension in a third, and normal blood pressure in less than a fifth of patients. In children, sustained hypertension occurs most frequently.

Patients with pheochromocytoma frequently present with paroxysmal episodes or spells that include the classic triad of severe headaches, palpitations, and diaphoresis. These episodes may occur daily or as frequently as every few months. More than 90% of patients present with at least two of the three symptoms in the classic triad. The headaches are typically abrupt in onset, throbbing, and bilateral and diminish within an hour. The headaches may be associated with pallor or nausea, may be brief, or may persist over a week. The presence of palpitations, anxiety, or tremulousness may suggest the predominant secretion of epinephrine. Less common symptoms include tremor, angina, nausea, Raynaud's phenomenon, livedo reticularis, and mass effect from the tumor.

Increased total peripheral resistance causes the hypertension in patients with pheochromocytoma, as in patients with essential hypertension. Heart rate is variably increased in patients with pheochromocytoma.

Normal cardiac output is maintained by a decreased stroke volume resulting from intravascular volume depletion.

Lability of blood pressure is caused by a combination of the following: (1) episodic catecholamine release, (2) impaired sympathetic reflexes, and (3) unrecognized chronic volume depletion. Altered sympathetic vascular regulation may underlie the orthostatic hypotension often seen in pheochromocytoma. In rare cases of pheochromocytoma with predominant secretion of epinephrine, dopamine, or dopamine, orthostatic hypotension may be the presenting symptom. Patients with pheochromocytoma who are asymptomatic despite high circulating levels of catecholamines may have adrenergic receptor desensitization related to chronic stimulation.

Other cardiovascular manifestations of pheochromocytomas include dilated cardiomyopathy resulting from catecholamine excess or hypertrophic cardiomyopathy. Both forms have been reported to be reversible with tumor resection. Myocarditis has also been described in patients with pheochromocytoma with a pathology characterized by infiltration of inflammatory cells, specifically perivascularly, and focal conduction band necrosis.

Patients may also present with features of acute myocardial infarction including chest pain or electrocardiographic abnormalities including ST segment elevation or depression or inversion of T waves, or both. Other electrocardiographic manifestations of pheochromocytoma include left ventricular hypertrophy, sinus tachycardia, T-wave inversion, and rhythm disturbances such as supraventricular tachycardia or supraventricular ectopic beats.

Hereditary Pheochromocytoma

The majority of pheochromocytomas are sporadic. However, approximately 10% or more occur in association with a familial disorder such as multiple endocrine neoplasia type 2A or 2B (MEN-2A or MEN-2B), von Hippel-Lindau (VHL) disease, or neurofibromatosis (see Chapter 36). The MEN-2 syndromes and VHL disease are inherited in an autosomal dominant pattern with age-related penetrance. MEN-2A (Siepe's syndrome) is characterized by pheochromocytoma, medullary carcinoma of the thyroid, and multiple mucosal neuromas, often in association with a marfanoid habitus. Germ line mutations of the RET proto-oncogene have been described in the MEN-2 syndromes, and the MEN-2A gene has been localized to chromosome 10q11.2. The mutation confers constitutive activation of the tyrosine kinase receptor leading to unregulated hyperplasia and increased susceptibility to malignant transformation.

The VHL disease phenotype includes pheochromocytoma, cerebellar and retinal hemangioblastomas, renal carcinoma, and renal and pancreatic cysts. The VHL disease suppressor gene has been cloned to chromosome 3p25-p26. A loss-of-function mutation leads to tumor formation with different kindreds demonstrating varied clinical manifestations related to different types of mutations. Missense mutations are thought to be more commonly associated with pheochromocytoma.

A low percentage (0.1% to 5.7%) of patients with von Recklinghausen's neurofibromatosis have pheochromocytoma. However, a much higher percentage (50%) of pheochromocytoma is seen in such patients who have hypertension. The majority of patients with von Recklinghausen's disease have solitary pheochromocytomas, whereas bilateral disease is often seen in other hereditary syndromes. Inactivating mutations in the neurofibromatosis F1 (NF1) gene, a tumor suppressor gene on
Hereditary pheochromocytomas are typically intra-adrenal and bilateral. In one series, up to 83% of the patients with familial pheochromocytoma had bilateral tumors. Patients with hereditary pheochromocytoma typically present at younger ages than those with sporadic pheochromocytoma. The mean age of diagnosis of familial pheochromocytoma was 38 ± 11 years, compared with 47 ± 16 years for patients with sporadic tumors. Sporadic pheochromocytoma cases usually present with hypertension; in contrast, many cases of the familial syndrome are diagnosed earlier as a result of biochemical surveillance or genetic testing, often before hypertension is detected.

The MEN-2 and VHL tumors make up most of the hereditary pheochromocytomas, but about 25% apparently sporadic patients may have germline mutations. It has been shown that MEN-2 tumors typically produce metanephrine, the metabolite of epinephrine, whereas tumors in patients with VHL disease produce normetanephrine, the metabolite of norepinephrine. These specific biochemical phenotypes demonstrate that unique mutation-dependent differential gene expression is probably involved in catecholamine synthesis. PNMT has been reported to be overexpressed in the MEN-2 tumors providing the epinephrine metabolism profile, whereas PNMT is under-expressed in VHL tumors providing the norepinephrine metabolism profile.

MEN-2 pheochromocytomas also appear to have increased tyrosine hydroxylase activity, which accounts for the greater concentration of catecholamine metabolites measured and clinical symptoms seen in MEN-2 patients compared with VHL patients. Patients with MEN-2 typically demonstrate episodic symptoms of hypertension. On the other hand, a pattern of sustained hypertension is seen in VHL patients. Thus, the biochemical phenotypes in these syndromes appear to be associated with particular patterns of catecholamine synthesis and release. Measurement of plasma-free metanephrines has been used to distinguish between MEN-2 and VHL disease and to reveal the presence of pheochromocytoma prior to clinical symptoms with greater sensitivity and specificity than urine testing (see later).

Hereditary paragangliomas of the neck (glomus tumors) are associated with germline mutations in a mitochondrial complex II gene, succinyl dehydrogenase subunit D (SDHD), which encodes an enzyme that is involved in oxidative phosphorylation. Somatic and germline mutations of the SDHD gene may also be associated with non-syndromic, sporadic, familial and familial pheochromocytoma.

**Diagnosis**

**Differential Diagnosis**

Pheochromocytoma may be suspected when a crisis, the physiologic consequence of abrupt catecholamine release, is precipitated by factors such as exertion, trauma, certain drugs, anesthesia, surgery, or surgical manipulation of the tumor. Tricyclic antidepressants, droperidol, glucagon, metoclopramide, phenothiazines, and naloxone have all been reported to induce hypertensive episodes. Foods or beverages, such as certain aged cheeses or red wine that contain tyramine, may precipitate a crisis. The β-blockers may cause a paradoxical rise in blood pressure. Several disorders may mimic the symptoms of pheochromocytoma and also cause elevations in catecholamines. Abrupt withdrawal from medications such as clonidine or from alcohol may produce such a picture. Cerebral events such as cerebral vasculitis, preeclampsia, subarachnoid hemorrhage, migraine, and intracranial lesions associated with increased intracranial pressure may mimic pheochromocytoma. Agents such as amphetamines, ephedrine, pseudoephedrine, isoproterenol, phenylpropanolamine, cocaine, phenylcyclidine (PCP), and lysergic acid diethylamide (LSD) also lead to excess catecholamine levels.

On the other hand, the symptoms of pheochromocytoma may be mistaken for those of panic attacks, hypoglycemic episodes, or accelerated hypertension of other etiologies. Lastly, disorders such as mastocytosis and the carcinoid syndrome, which are characterized by spells and episodic symptoms, may also mimic pheochromocytoma. However, hypertensive crises, which often occur with pheochromocytoma, are notably absent in these disorders. In fact, episodic hypotension may occur with mastocytosis or the carcinoid syndrome as a result of peripheral vasodilation.

**Indication for Screening**

Because pheochromocytomas do not occur frequently, physicians must appreciate when screening for the disorder is appropriate. The following are reasonable indications for screening:

1. Hypertension with episodic features suggesting pheochromocytoma (the classic triad of headaches, palpitations, and diaphoresis)
2. Refractory hypertension
3. Prominent lability of blood pressure
4. Severe pressor response during anesthesia, surgery, or angiography
5. Unexplained hypotension during anesthesia, surgery, or pregnancy
6. Family history of pheochromocytoma or a familial disorder such as MEN-2, VHL disease, neurofibromatosis, or glomus tumors
7. Incidentally discovered adrenal masses
8. Idiopathic dilated cardiomyopathy

**Biochemical Assessment**

In pheochromocytoma, enzyme activity involved in synthesis of catecholamines is augmented and enzyme activity involved in catabolism is decreased. Because the catecholamine excess cannot be effectively stored, the hormones spill into the peripheral circulation. Biochemical measurement of excessive catecholamine production by the tumor confirms the diagnosis of pheochromocytoma (see later). However, because catecholamines are normally constitutively produced by the sympathoadrenal system, it is the magnitude of the elevation that is diagnostic of pheochromocytoma.

**Screening**

The diagnosis is made with the demonstration of elevated circulating or urinary catecholamines or metabolites (see Fig. 15-3). Screening methods include (1) 24-hour urine collection for excretion of unmetabolized or so-called free catecholamines (epinephrine and norepinephrine) or catecholamine metabolites (metanephrine, normetanephrine, and vanillylmandelic acid) and (2) determination of plasma metanephrines and catecholamines (Fig. 15-5). Pheochromocytomas are heterogeneous in hormone metabolism and secretion. Therefore, there is no one optimal test for screening and there is disagreement about the preferred test for diagnosis. For example, as catecholamines have short half-lives and are secreted episodically, a
random plasma measurement may miss the peak catecholamine levels. On the other hand, plasma levels are particularly helpful when samples are collected during a paroxysm. Although the 24-hour urine collection has the advantage of integration of the catecholamine secretion over time, it is more cumbersome for patients and may yield a false-negative result if the collection is performed in absence of symptoms or hypertension in patients with episodic catecholamine secretion. The urinary methanephrine-to-creatinine ratio can be useful in compensating for overcollection (false-positives) or undercollection (false-negatives). 12 Overnight measurements of urine catecholamines have also been used to diagnose pheochromocytoma but involve an increased risk of both false-positive and false-negative results. 13

In the initial evaluation of a patient suspected of having a pheochromocytoma, we recommend a 24-hour urine collection for free or unmetabolized catecholamines (epinephrine and norepinephrine), total methanephrines, and creatinine (see Fig. 15-5). Of the different metabolites that can be detected in a 24-hour urine collection, methanephrines are the most sensitive and specific. 14 There are less data concerning the appropriate use of plasma measurements. However, more experience with the measurement of plasma methanephrines is accruing, and it is considered by some to be a highly sensitive method for biochemical diagnosis, especially for hereditary pheochromocytoma (sensitivity 97%, specificity 96%). 15 Because pheochromocytomas secrete primarily metabolized catecholamines, plasma methanephrines may be more useful than plasma catecholamines.

Measurement of plasma-free methanephrines as opposed to the conjugated or sulfated forms is particularly helpful because free fraction levels result from the actions of COMT on tumor catecholamine production. As a result, plasma-free methanephrines may show larger increases above normal than plasma catecholamines in pheochromocytoma. On the other hand, although plasma methanephrines are highly sensitive in detecting pheochromocytoma, a high number of false-positives may occur, particularly in older patients. Accordingly, the sensitivities and specificities of urine and plasma biochemical tests remain under investigation. As a result, some centers advocate urine testing as the initial screening test whereas others suggest plasma-free methanephrine levels. It is important to remember that the current technology for measuring plasma methanephrines requires that the patient abstain from acetaminophen for 3 to 5 days before testing 16 (see Fig. 15-5).

Typically, a measurement of urinary catecholamines or metabolites that is two or three times above the upper limit of normal is considered diagnostic of pheochromocytoma. For example, the upper limit of normal for total urinary catecholamines is approximately 100 µg per 24 hours, and a measurement above 250 µg per 24 hours is obtained in most patients with pheochromocytoma. Urine collections should include measurement of urinary creatinine to verify the adequacy of collection. A strong acid (such as 6 N HCl) is added to a sealed container. The optimal system for plasma catecholamine determination includes having the patient fast overnight and lie comfortably in a supine position with a heparin lock inserted 20 to 30 minutes before collection for withdrawing the blood.

Certain precautions should be taken in interpreting catecholamine or methanephrine values. For example, iodinated contrast dyes can interfere with some biochemical measurements. Labelato can give falsely elevated results when the following assays are employed: fluorometric methods of analysis used for catecholamine measurements, spectrophotometric methods used for methanephrine measurements, or radioenzymatic assays used for urinary-free catecholamine measurements. 17 Tricyclic antidepressants, prochlorperazine (Compazine), reserpine, clonidine, and clofibrate may interfere with urinary catecholamines and metabolite measurements. 18 Such medications should be discontinued, preferably 2 weeks before collection. Blood pressure should be controlled with agents such as dihydropyridine calcium channel blockers that do not interfere with the assays.

To provide better resolution against interfering substances, many laboratories use a reverse-phase high-performance liquid chromatography method with electrochemical detection. Measurement of these compounds by mass spectroscopy, which eliminates the problem of interfering substances, and use of immunoassay techniques are future directions. Stresses associated with serious illnesses, such as myocardial infarction, cerebral vascular accidents, or congestive heart failure, cause elevation in catecholamine levels. In renal insufficiency, plasma and urinary levels may be falsely elevated and urinary collections should be expressed in milligrams of creatinine. 19 In these circumstances, other diagnostic tests including imaging modalities are required for evaluation.

Chromogranin A is a soluble protein stored and secreted with catecholamines in chromaffin tissue. This biochemical marker is not specific for pheochromocytoma, and elevations may be seen with other neuroendocrine tumors. Plasma chromogranin A levels are elevated in more than 80% of patients with pheochromocytomas. 20 This assay is most often utilized in the postoperative surveillance of patients after resection of catecholamine-secreting tumors (see "Medical and Surgical Management"). Chronic renal failure is also associated with elevated chromogranin A levels.

Stimulation and Suppression Tests

The clonidine and glucagon tests are dynamic tests that are not routinely performed but are usually used when the suspicion of pheochromocytoma is high but the basal catecholamine levels are not diagnostic or are equivocal. Clonidine is a centrally acting α2-adrenoceptor agonist that normally suppresses the release of catecholamines from neurons but does not affect the autonomous release of catecholamines from a neoplasm. 21 In patients without pheochromocytoma, a decrease in basal plasma catecholamines by 50% or less than 3 nmol/L (500 pg/mL) is expected 2 to 3 hours after 0.3 mg of clonidine is administered. 22

Provocative testing is utilized when clinical suspicion of pheochromocytoma is not supported by the biochemical testing. Patients with pheochromocytomas typically demonstrate a threefold increase in plasma catecholamine levels or a concentration greater than 12 nmol/L (2000 pg/mL) 2 minutes after administration of 1.0 mg of intravenous glucagon. The glucagon provocative test is considered highly specific but poorly sensitive, whereas the clonidine test is considered highly sensitive with poor specificity. If both tests are negative, the diagnosis of pheochromocytoma may be reasonably excluded. 23 24

Imaging Techniques

After the diagnosis of pheochromocytoma is confirmed by biochemical testing, imaging techniques are employed for tumor location. Localization techniques include magnetic resonance imaging (MRI), computed tomography (CT), and MIBG or octreotide scintigraphy. Pheochromocytomas are typically large tumors (2 to 5 cm in diameter) and may contain areas of hemorrhage or necrosis. Pheochromocytomas in hereditary syndromes tend to be bilateral and smaller. 25 The latter feature is probably related to early detection as the result of periodic surveillance.

Approximately 98% of pheochromocytomas are intra-abdominal, and 90% originate within the adrenal gland. However,
Somatic serotonin receptor scintigraphy is another localization technique because somatostatin receptors are normally expressed in adrenomedullary and paraganglionic tissues. The receptor density is increased on pheochromocytoma tissue, which enables imaging with the somatostatin analogue octreotide. Octreotide, which binds to somatostatin receptor subtypes 2 and 5, is labeled with 111 In-DTPA. As with MIBG, octreotide scanning is best employed for detection of pheochromocytomas of extra-adrenal origin and of metastases from malignant pheochromocytoma. Some malignant tumors down-regulate the expression of somatostatin receptors. As a result, lack of uptake by octreotide may be a poor prognostic indicator in such patients.

In contrast to MIBG or octreotide scintigraphy, which requires 24 to 48 hours for optimal visualization, a promising new technology that can immediately image a pheochromocytoma is 6-[18 F]Fluorodopamine positron emission tomography. The uptake and retention of this radiopharmaceutical in chromaffin cells with subsequent imaging by emission scanning may provide a useful diagnostic test in patients with pheochromocytoma.

Finally, venous sampling has been utilized in the past to confirm or rule out the diagnosis of pheochromocytoma. The adrenal sampling effluent has a norepinephrine/epinephrine ratio less than 1. Higher ratios suggest the presence of pheochromocytoma.

Medical and Surgical Management

Preoperative Management

Surgical excision of a pheochromocytoma is the treatment of choice, but it involves a risk of morbidity as high as 40% and a risk of mortality of 2% to 4%. Surgical outcomes have improved with preoperative treatment such as -blocker blockade and volume expansion. Volume expansion is initially achieved with a high-sodium diet (150 to 200 mEq/L [150 to 200 mmol/dl]) unless contraindicated by congestive heart failure or renal insufficiency. Phenoxbenzamine, a noncompetitive -blocker, has traditionally been used for preoperative preparation. Phenoxbenzamine is titrated to reduce blood pressure to normal levels or orthostasis, or both. The starting dose, 10 mg/day by mouth, is titrated upward every 2 days. Most patients require 80 to 100 mg daily given in divided doses.

When -blockade is established, -blockade may be initiated if the patient is tachycardic or has arrhythmias. Without prior -blockade, -blockade alone can lead to unopposed -receptor stimulation and further elevation of blood pressure. Because phenoxbenzamine blocks catecholamine binding to receptors, it minimizes the risk of a hypertensive crisis during intubation, during induction with anesthesia, or during exploration and tumor manipulation. A noncompetitive -blocker is theoretically preferred to a competitive inhibitor because catecholamine levels, which can increase 500-fold, may overcome a competitive inhibitor. However, complete -blockade can mask the dramatic fall in blood pressure seen after tumor resection that signals to the surgeon that the pheochromocytoma is resected. Phenoxbenzamine may also lead to postoperative hypotension because of its prolonged half-life of 24 hours.

Calcium channel blockers, in particular nicardipine from the dhydropyridine class, are increasingly popular. These agents improve intraoperative systemic vascular resistance by blunting catecholamine-mediated arterial vasoconstriction during tumor manipulation, and they have few side effects. The selective, -inhibitor doxazosin has been effective in preoperative management without causing tachycardia or other serious side effects. The oral formulation of labetalol, an -blocker, might be ideal for preparation for surgery because the -blockade is weaker than that of phenoxbenzamine. As a result, additional vasodilators may be required intraoperatively in labetalol-treated subjects. Finally, metyrosine, which inhibits catecholamine synthesis, has been used in combination with -blockade for preoperative management.

Tradition and experience have guided the length of preoperative therapy with volume expansion and -blockade. Patients are typically treated for 10 to 14 days, although this time course has not been consistently associated with better operative and postoperative outcomes. Unfortunately, there are no reliable features that predict a smooth surgical course. Thus, each patient must be evaluated on an individual basis when selecting antihypertensive medicines.

Surgical Management

The surgical approach is dictated by the clinical situation. In patients with familial pheochromocytoma, a transabdominal incision allows adequate visualization and bilateral adrenalectomy if required. The flank approach offers better exposure and reduced blood loss for the patient with a solitary tumor. Surgeons are gaining experience with laparoscopic adrenalectomy performed when the tumor is smaller than 6 cm. The patient with pheochromocytoma should be referred to a surgeon who has experience in the management of pheochromocytoma and who collaborates with an experienced anesthesiologist to establish a smooth team effort.

Intraoperative hypotension is managed initially with volume expansion and then with intravenous pressor agents if necessary. Postoperative hypoglycemia, which may be due to reactive hyperinsulinemia, should be anticipated and warrants routine screening of glucose monitoring in the early postoperative hours. Surgical outcomes have been improved by intraoperative hemodynamic monitoring, the combination of fast-acting intravenous vasodilators and -blockers (sodium nitroprusside and esmolol, respectively), and the use of intravenous vasoconstrictors (norepinephrine or epinephrine).

The following is an approach followed in our institution to prepare the patient for surgery. As soon as the diagnosis of pheochromocytoma is established, -blockade is started and titrated upward (see earlier). Five days before surgery, if not earlier, a high-sodium (150 to 200 mEq/L [150 to 200 mmol]) diet is initiated; therapy with -blockade is continued, and daily weights and vital signs are monitored. Admission for volume expansion with intravenous saline is considered, depending on the patient's status. One day before surgery, the patient is transferred to a monitored intensive care setting with intravenous arterial and Swan-Ganz catheters. Isotonic saline is administered to achieve a pulmonary capillary wedge pressure greater than 10 mm Hg. If systemic vascular resistance is elevated (>1000 dyne second/cm 2/m2), intravenous sodium nitroprusside at 0.5 to 2 µg/kg per minute is initiated and increased as needed (500 to 1000 mg/minute may be required). On the day of surgery, the patient is given two separate intravenous lines, one for administration of pressors and the other for administration of vasodilators. Sodium nitroprusside, esmolol, epinephrine, and norepinephrine infusions are on standby in the event that they are required.

When surgery is performed, the catecholamine levels usually return to normal in approximately 2 weeks. If hypertension persists despite normal catecholamine levels and patient condition is good, reexploration should be considered.

Pregnancy

Management of pheochromocytoma in pregnancy is especially challenging. The mortality rate for mother and fetus is reported to be approximately 50%. Diagnosis before term improves these rates considerably. Clinical symptoms are similar to those in nonpregnant individuals, but unique features can occur. For example, the gravid uterus may compress the pheochromocytoma, causing paroxysms in the supine position with normal blood pressure in the erect posture.

Pheochromocytomas may also be easily misdiagnosed as preeclampsia, especially later in pregnancy. The diagnosis is typically made by evaluation of the urinary collection of catecholamines and metanephrines. Methyldopa should be discontinued before collection because of interference in catecholamine measurements. MRI is the preferred imaging modality because there is no ionizing radiation, and MIBG is contraindicated. Surgery is typically performed before 20 to 24 weeks of gestation. Thereafter, medical therapy is attempted, depending on the maternal status, and cesarean section is planned followed by tumor resection. Phenoxbenzamine has been used during pregnancy, but it does cross the placental barrier; as a result, calcium channel blockers may be preferable to control blood pressure.
Malignant Pheochromocytoma

Malignant pheochromocytoma occurs in 3% to 13% of all cases. The 5-year survival rate is 23% to 44%, compared with 97% 5-year survival in benign pheochromocytoma. These tumors typically grow slowly, and evidence of malignancy may not be seen for several years. Malignancy is defined by direct local invasion of sites that do not typically have chromaffin tissue. Malignant tumors most commonly metastasize to the lungs, bone, liver, or lymph nodes or may recur locally. Surgical removal or debulking is the treatment of choice.

After surgery, treatment goals are palliative to control the symptoms related to excess catecholamines. To achieve this end, -blockade followed by -blockade has most often been used (as discussed earlier). However, use of other antihypertensive treatments, such as dihydropyridine calcium channel blockers, is increasing. Unfortunately, the response to chemotherapeutic agents has been disappointing, but they may be tried in combination with antihypertensive treatment. Chemotherapeutic agents such as vincristine, cyclophosphamide, and dacarbazine have been used, often in combination. Although the response rate is suboptimal, the experience with high-specific-activity 131I-labeled MIBG is increasing, sometimes in combination with chemotherapy. External beam radiation has been attempted for palliation of bone metastases. Tumor embolization is another approach when surgery is not possible.
RENIN-ANGIOTENSIN-ALDOSTERONE AXIS

Several different mechanisms can lead to an increase in blood pressure in patients with essential hypertension and are similar to the mechanisms that cause an increase in blood pressure in individuals with secondary forms. A leading candidate for one of these mechanisms is a derangement in the renin-angiotensin system.\[1\]

Components

Components of the renin-angiotensin system are shown in Figure 15-7.\[1\]

Re nib

Renin is an enzyme produced in a number of cells in the body, principally in the juxtaglomerular apparatus of the kidney. In tissues that produce renin, it is stored in granules and released in response to specific secretagogues. It is a member of the aspartyl proteinase family of enzymes and is synthesized as a pre-proprotein. In humans, the gene that encodes renin is located on the short arm of chromosome 1 (1q21q22). In the rat the gene is located on chromosome 13, and in the mouse it is located on chromosome 1\[1\] (the mouse has two renin genes). In each species, the nucleotide sequence is approximately 12 kb, with 10 exons and 9 introns. The transcript production is a 1.5-kb messenger ribonucleic acid, and the initial protein consists of 340 amino acids, of which the first 43 are a prosegment cleaved to produce the active enzyme.

Renin is termed a double-domain enzyme because the N-terminal and C-terminal halves are similar.\[1\] Each domain contains a single aspartic acid residue critical for its catalytic activity. The three-dimensional structure of the enzyme has been characterized. A number of factors can regulate the transcription of the renin gene; consensus elements are present in the 5-flanking region of the gene, including those for cAMP and a number of steroid receptors (estrogen, progesterone, and glucocorticoids).\[1\] [1]

Angi tensinogen

Angiotensinogen is the only known substrate for renin and is catabolized to angiotensin peptides. The interaction between enzyme and substrate appears to be species specific because minor structural variations in the substrate render it relatively inactive in different species.\[1\] Human angiotensinogen belongs to the serpin superfamily of proteins and is encoded by a gene on chromosome 1q24.3 near the renin gene.\[1\] The angiotensinogen gene consists of five exons and four introns and is approximately 13 kb long. The transcript encodes a protein of 485 amino acids, 33 of which constitute a presegment that is cleaved after secretion. Angiotensin I is composed of the first 10-amino-acid sequence following the presegment. The 5 promotor region has consensus sequences for control by glucocorticoids, estrogens, and cytokines.\[1\] [1]

Angiotensin-Converting Enzyme

Angiotensin-converting enzyme (ACE), a second enzyme involved in the final production of angiotensin II (see Fig. 15-7), is a dipeptidyl carboxyl zinc metallopeptidase usually found bound to cell membranes.\[1\] It is also present in intracellular granules in certain tissues that produce angiotensin II. Its molecular weight is considerably greater than that of renin and it consists of two homologous domains, suggesting that there are two active sites in each molecule. In humans, the ACE gene is located on chromosome 17q23 and consists of 26 exons and 25 introns. Two molecular forms of ACE are products of a single gene but have separate promoter regions. One product is a somatic, or endothelial, ACE that consists of 1306 amino acids, and the second is a germinal ACE with a promoter region upstream from the 13th exon.\[1\]

Angiotensin Receptors

In humans, the two primary forms of the angiotensin receptor are termed AT\[1\] and AT\[2\]. A single gene on chromosome 3 encodes the angiotensin receptor in humans; rats have two genes. The S-flanking region contains three putative glucocorticoid response elements. The receptor has seven transmembrane regions, with a disulphide bridge linking the first and fourth extracellular segments. The principal signaling mechanism involved in the AT\[1\] receptor operates through a \(G_\alpha\) protein-mediated activation of phospholipase C.\[1\] However, some data suggest a linkage to protein tyrosine kinase.\[1\] [1] \[1\] The AT\[2\] receptor gene has three exons and two introns and a seven-transmembrane-domain structure.\[1\] [1]

Angiotensin Peptides

At least four angiotensin-like peptides have biologic activity (Table 15-1).\[1\] [1] The action of renin on angiotensinogen produces angiotensin I, a decapeptide that does not appear to have biologic activity. Angiotensin II is formed by cleavage of the two carboxyl-terminal peptides by ACE.\[1\] [1] and has full biologic activity. Amino peptidase A can remove the aminoterminal aspartic acid to produce the heptapeptide, angiotensin III. Angiotensin II and angiotensin III have equivalent efficacy in promoting aldosterone secretion and modifying renal blood flow. However, angiotensin III has less pressor activity. Amino peptidase B can cleave an additional amino acid from angiotensin III to form angiotensin IV (angiotensin 38).\[1\] The function of this peptide is not clear, but it may be involved in the regulation of cerebral circulation and may produce vasodilation rather than vasoconstriction. A fourth biologically active compound is produced from angiotensin I by the action of a propyl...
endopeptidase to form angiotensin 17, whose function is unclear.
Functions of Angiotensin II

The effects of the renin-angiotensin system can be mediated by local paracrine effects or through endocrine action. The endocrine system primarily involves renin from the juxtaglomerular apparatus of the kidney and angiotensinogen from the liver. In the circulation, the concentrations of each are such that variations in the angiotensinogen levels can modify angiotensin I generation. The half-life in the circulation of angiotensin II is short (probably less than a minute). Although circulating levels of angiotensin II are in the picomolar range, its affinity for its receptor is in the nanomolar range, suggesting that some angiotensin II effects may actually be mediated not by the circulating peptide but by its local generation.

Elements of the renin-angiotensin system are present in the adrenal, the kidneys, the heart, and the brain. For example, the adrenal glomerulosa cells contain the proteins needed to produce and secrete angiotensin II. Other tissues contain one or more components of the renin-angiotensin system and require other cells or circulating components, or both, to generate angiotensin II. For example, fat cells synthesize angiotensinogen but not renin or ACE, but they can generate angiotensin II locally. An increasing body of evidence suggests that many of the functions of angiotensin II are mediated by these paracrine effects. In some tissues, such as the heart, the angiotensin II may be generated by a nonrenin system the chymase system.

Angiotensin II functions through the AT₁ receptor to maintain normal extracellular volume and blood pressure in five ways: (1) constriction of vascular smooth muscle, thereby increasing blood pressure and reducing renal blood flow; (2) release of norepinephrine and epinephrine from the adrenal medulla; (3) enhancement of the activity of the sympathetic nervous system by increasing central sympathetic outflow, thereby increasing norepinephrine discharge from sympathetic nerve terminals; (4) promotion of the release of vasopressin; and (5) increasing aldosterone secretion.

Other functions of angiotensin II mediated through the AT₁ receptor include (1) central nervous system effects, including modification of thirst or the sense of well-being, or both; (2) modification of the release of corticotropin from the pituitary gland; (3) possible effects on placental and ovarian function; (4) activation of plasminogen activator inhibitor type 1, thereby contributing to the coagulation cascade; and (5) modification of growth of the heart, kidneys, and vascular smooth muscle.

In many respects, the action of angiotensin II through the AT₂ receptor antagonizes its effects through the AT₁ receptor. Thus, AT₂-receptor activation results in vasodilation, renal sodium loss, and apoptosis (thereby antagonizing the growth-promoting effects of AT₁ receptor activation). AT₂ receptors are highly expressed in fetal compared with adult tissue unless the adult tissue is damaged.
Functions of Aldosterone

Aldosterone's classical functions are twofold: regulation of extracellular volume and control of potassium homeostasis. These effects are mediated by binding to the mineralocorticoid receptor in the cytosol of epithelial cells, principally in the renal collecting duct. Transport to the nucleus and binding to specific binding domains on targeted genes lead to their increased expression. Although not all the genes have been identified, serum and glucocorticoid-induced kinase appears to be a key intermediary. Its increased expression leads to modification of the apical sodium channel and the basal lateral Na⁺,K⁺-adenosine triphosphatase (ATPase), resulting in increased sodium ion transport across the cell membrane (see Chapter 14). Glucocorticoids and mineralocorticoids bind equally to the mineralocorticoid receptor. Specificity of action is provided in many tissues by the presence of a glucocorticoid-degrading enzyme, 11-hydroxysteroid dehydrogenase, which prevents glucocorticoids from interacting with the receptor (see Chapter 14).

A second protective mechanism for untoward mineralocorticoid action is "escape" from its renal sodium-retaining effect. This usual occurs within 3 to 5 days of continued administration. Several mechanisms contribute to this escape, including renal hemodynamic factors and an increase in atrial natriuretic peptide.

In addition to these classical genomic actions, mediated by aldosterone binding to cytosolic receptors, an increasing body of data suggests that mineralocorticoids have acute, nongenomic actions secondary to activation of an unidentified cell surface receptor. This action involves a G protein signaling pathway and probably a modification of the sodium-hydrogen exchange activity. In both epithelial and nonepithelial cells (e.g., myocytes and leukocytes) this effect has been demonstrated.

There are additional nonclassical effects of aldosterone primarily on nonepithelial cells. These actions, although probably genomic and therefore mediated by activation of the cytosolic mineralocorticoid receptor, do not include modification of sodium-potassium balance. Aldosterone-mediated actions include the expression of several collagen genes; genes controlling tissue growth factors, such as transforming growth factor \( eta \); and plasminogen activator inhibitor type 1; or genes mediating inflammation. The resultant actions lead to microangiopathy, necrosis (acutely), and fibrosis in a variety of tissues, such as heart, the vasculature, and kidney. Increased levels of aldosterone are not necessary to cause this damage. Rather, an imbalance between the volume or sodium balance state and the level of aldosterone appears to be the critical factor.
Regulation

Renin

The release of renin into the circulation from the kidneys is controlled by four factors: (1) the macula densa, a specialized group of distal convoluted tubular cells that function as chemoreceptors for monitoring the sodium and chloride loads present in the distal tubule; (2) juxtaglomerular cells acting as miniature pressure transducers that sense renal perfusion pressure; (3) the sympathetic nervous system, which modifies the release of renin, particularly in response to upright posture in humans; and (4) humoral factors including potassium, angiotensin II, and atrial natriuretic peptides. The tissue renin-angiotensin systems are not necessarily regulated in the same manner as the circulating renin-angiotensin system. For example, a high potassium intake reduces renal renin release and increases adrenal renin secretion.

Aldosterone

The action of angiotensin II on aldosterone involves a negative feedback loop that also includes extracellular fluid volume (Fig. 15-8) (Figure Not Available). The major function of this feedback loop is to modify sodium homeostasis and, secondarily, to regulate arterial pressure. Thus, sodium restriction activates the renin-angiotensin-aldosterone axis. The effects of angiotensin II on both the adrenal cortex and the renal vasculature promote renal sodium conservation. Conversely, with suppression of renin release and suppression of the level of circulating angiotensin, aldosterone secretion is reduced and renal blood flow is increased, thereby promoting sodium loss.

In addition to the usual internal regulation of this negative feedback loop, a secondary fine-tuning component is related to the level of dietary sodium intake. Most endocrine negative feedback loops are not particularly sensitive to environmental factors. In contrast, the renin-angiotensin-aldosterone loop is exquisitely sensitive to dietary sodium intake. Sodium excess enhances the renal and peripheral vasculature responsiveness and reduces the adrenal responsiveness to angiotensin II (Fig. 15-9). Sodium restriction has the opposite effect. Thus, sodium intake modifies, or modulates, target tissue responsiveness to angiotensin II, a fine tuning that appears to be critical to maintaining normal sodium homeostasis without modifying blood pressure, particularly chronically. The mechanism or mechanisms by which dietary sodium intake induces these changes in the adrenal is unclear, but in the vascular system the effects are a consequence of angiotensin II down-regulation of the target tissue responsiveness to its agonists.
ESSENTIAL HYPERTENSION

The renin-angiotensin system has a powerful influence on both vasoconstrictor activity and volume regulation. Thus, defects in its regulation could lead to a rise in blood pressure by either or both of these mechanisms. Two other hormonal systems are implicated in the pathogenesis of essential hypertension: insulin (either directly or mediated by selective insulin resistance) and the calcium regulating systems.

Role of the Renin-Angiotensin System in the Pathogenesis of Essential Hypertension

In the late 1960s and early 1970s, Laragh and colleagues developed a classification of hypertension based on the level of circulating renin activity. By controlling dietary sodium and potassium intake, they classified patients into those whose values were low, normal, or high and used this information to define whether an individual case of hypertension was more volume-dependent or vasoconstrictor-dependent. The model predicted that individuals with low plasma renin activity (PRA) levels would have a volume-sensitive form of hypertension and those with high plasma levels would have a vasoconstrictor form of hypertension. It was presumed that classifying patients in this manner would lead to a more rational treatment program.

However, several concerns have been raised. First, age modifies the level of renin activity, older subjects having lower renin levels regardless of volume status. Second, race modifies the level of renin activity (whites, in general, have higher levels than blacks). Finally, in individuals who consume a relatively large amount of sodium (more than 175 mmol/day) low, normal, and high PRA levels are difficult to distinguish from each other. However, the concept of subclassification of hypertensive patients on the basis of the level of renin activity was useful in the development of better approaches to subclassifying patients.
Pathophysiologic Mechanisms in Low-Renin Essential Hypertension

Several mechanisms are thought to cause volume expansion and suppress renin activity in some patients with essential hypertension. Adrenal mechanisms may be involved in some subjects with low-renin hypertension because spironolactone (a mineralocorticoid antagonist) and aminoglutethimide (an inhibitor of steroid hormone biosynthesis) substantially reduce their blood pressure. Wisgerhof and Brown reported that the adrenal response to angiotensin II is enhanced in some patients with low-renin essential hypertension and that the enhanced responsiveness alters the renin-angiotensin-aldosterone negative feedback loop, allowing restoration of normal sodium homeostasis with decreased PRA and angiotensin II levels. However, with normal to high sodium intake, this enhanced adrenal response could result in a scenario in which aldosterone secretion would not be suppressed adequately, promoting sodium retention and an increase in blood pressure. The frequency of this abnormality is unclear because no population-based studies have been reported. However, even some patients with so-called normal renin essential hypertension appear to have a similar defect.
Nonmodulating Hypertension: Salt Sensitivity and Normal to High Renin Levels

Some patients with normal or high renin levels have a peculiar form of salt-sensitive hypertension in which increased sodium intake fails to change the vascular and the adrenal response to angiotensin II. These patients, termed nonmodulators, appear to be a subset of the essential hypertensive population, as documented by a bimodal distribution of several of their biochemical features. Patients with these features have been reported from Argentina, Brazil, Japan, The Netherlands, France, Italy, and the United States. In whites, between 25% and 30% of hypertensive subjects are nonmodulators, and in black hypertensives the frequency is likely to be greater.

Nonmodulators share several features with low-renin essential hypertensive patients: (1) they both have salt-sensitive hypertension, and (2) they tend to be older than the rest of the hypertensive population. However, nonmodulators have several features that are not similar to those of low-renin hypertension, including (1) fasting hyperinsulinemia, (2) a positive family history of hypertension and myocardial disease, (3) elevated levels of cholesterol and triglycerides, and (4) a decreased adrenal response to angiotensin II as assessed with a sodium-restricted intake. Finally, and perhaps most important, the characteristics associated with nonmodulation distribute in a bimodal fashion in the hypertensive population, suggesting a discrete subgroup.

In nonmodulators, target tissue responsiveness to angiotensin II does not change when sodium intake is modified. Two functional tests have been used to distinguish them from the rest of the hypertensive population. One measures the aldosterone response to an angiotensin II infusion of 3 ng/kg per minute with a low-salt (10 mEq) diet. The other approach is to measure the renal blood flow response to the same dose of angiotensin II with a high-salt (200 mEq) diet. Unless the dietary sodium intake is precisely controlled, a hypertensive subject can be misclassified. The correlation between these two criteria is 70% to 80%. Thus, if feasible, the best approach is to require both criteria to be positive in defining a nonmodulator. Other characteristics of this subset include failure of renal blood flow to increase when dietary sodium intake is changed from low to high and an enhanced response of atrial natriuretic peptide to infused angiotensin II.

Nonmodulators appear to have an inherited form of hypertension, as evidenced by (1) bimodality of the distribution of the nonmodulating characteristic in the hypertensive population, (2) the presence of the nonmodulating characteristic in normotensive subjects, (3) a strong family history of hypertension in nonmodulators (approximately 80%, compared with about 30% for the rest of the hypertensive population), (4) familial aggregation of nonmodulating characteristics with hypertension, and (5) the association of the nonmodulating phenotype with individuals who are homozygous for the angiotensinogen 235T genotype (Fig. 15-11). Data also support the involvement of the ACE and aldosterone synthase (CYP11B2) genes. Nonmodulators are twice as likely as the rest of the hypertensive population to be homozygous for the angiotensinogen 235T genotype, four times as likely if they have both the angiotensinogen and ACE dd genotypes, and nearly six times as likely if they have the angiotensinogen, ACE, and CYP11B2-344T genotypes.

A defect in the renin-angiotensin system is likely to underlie nonmodulating hypertension. It is probably a defect in the local renal and adrenal renin-angiotensin systems, as evidenced by the following: (1) the previously cited genetic data involving genes of the renin-angiotensin-aldosterone system; (2) low renal blood flow with a high-sodium diet and a reduced renal vascular response to infused angiotensin II, suggesting inappropriately high local renal angiotensin II levels; (3) correction of the renal blood flow defect by administration of a converting enzyme inhibitor; and (4) correction of the nonmodulating adrenal defect by a converting enzyme inhibitor.

The effect of sodium intake on blood pressure in nonmodulators has been extensively evaluated. Either short-term (3 days) or chronic (2 weeks) salt loading increases blood pressure in nonmodulators but not in other normal or high-renin hypertensive patients. The salt sensitivity of the hypertensive is due to the tendency for nonmodulators to retain more of a salt load both acutely and chronically. The abnormality in sodium handling is probably due to the alteration in renal hemodynamics with salt loading described previously and secondary to an inappropriately high local angiotensin II level. Support for this conclusion comes from correction of salt-sensitive hypertension in nonmodulators by converting enzyme inhibitors.

In summary, nonmodulators are a distinct subgroup of the essential hypertensive population and may constitute as much as 30% of that population. They have a sodium-sensitive form of hypertension, probably owing to a derangement of the local renin-angiotensin system in the kidney and the adrenal. These patients also have insulin resistance, hypercholesterolemia, a family history positive for myocardial infarction.
and an association with a specific allelic variant of the angiotensinogen gene. Finally, the defect appears to be correctable by the administration of converting enzyme inhibitors.
Insulin Resistance and Hypertension

Noninsulin-dependent diabetes mellitus (NIDDM), hypertension, and obesity are commonly associated, and the frequency of this association may be greater than their occurrence in the general population (see the review by Hopkins and colleagues). There is a common etiology. In support of this possibility is the fact that insulin resistance and hypertension can coexist without obesity or other stigmata of NIDDM. There may be a genetic component of this interaction. For example, in whites of European descent there is a strong relation between insulin resistance and blood pressure, whereas in normotensive blacks or Pima Indians there is no such relationship. However, most hypertensive blacks are insulin resistant.

Causative Role of Insulin in Hypertension

Several mechanisms have been proposed to explain the insulin-resistant state, including abnormalities in insulin binding to its receptor, defects in glucose transport, changes in the signal transduction pathway within insulin-sensitive cells, and metabolic abnormalities in glycolysis, glucose oxidation, or glucagon synthesis. Yet little is known concerning the cause of insulin resistance in essential hypertension. (Fig. 15-14) Several features are relevant. First, insulin resistance is common in essential hypertension whether defined by fasting or postglucose load insulin levels or by euglycemic, hyperinsulinemic clamps. Second, obesity cannot explain all cases of insulin resistance. Third, insulin directly stimulates the calcium pump in insulin-sensitive tissues and promotes calcium loss from the cell, and raising cytosolic calcium levels in an adipocyte can induce insulin resistance. If a cell is resistant to insulin, the insulin-induced calcium loss from cells would be decreased, and in vascular smooth muscle cells the resultant increase in intracellular calcium would enhance responsiveness to vasoconstrictors and increase blood pressure.

Two other mechanisms have been proposed to explain the linkage between insulin resistance and hypertension: increased activity of the adrenergic nervous system and increased renal sodium retention. Underlying both these hypotheses is the assumption that insulin resistance in a hypertensive subject may be selective. Accordingly, insulin resistance in the skeletal muscle or liver, or both, would induce a rise in circulating insulin levels. However, there would be little, if any, resistance at the renal tubule or adrenergic nervous system. Finally, for the vasoconstrictor hypothesis to be correct, there would have to be an imbalance between insulin's direct vasodilator effect and the vasoconstriction induced by activation of the adrenergic nervous system.

A hyperinsulinemic response to glucose loading has been described in salt-sensitive but not in salt-resistant normotensive subjects. This salt sensitivity also extends to metabolic abnormalities that are often associated with insulin resistance, such as increased levels of circulating low-density lipoprotein cholesterol in salt-sensitive hypertensives compared with salt-resistant hypertensives. Thus, salt sensitivity of blood pressure is associated with lipid and glucose metabolic abnormalities and increased cardiovascular risk. Impaired insulin sensitivity and hyperlipidemia have been described in healthy volunteers with normal to high plasma renin levels compared with those with low renin levels. Thus, the data derived from these sources are inconsistent. Individuals who are salt-sensitive, as noted earlier, are likely to have low PRA. Yet salt-sensitive subjects as a group also have an increased risk of carbohydrate and lipid abnormalities.

In summary, insulin resistance occurs in some patients with essential hypertension who do not have obesity or NIDDM. Several lines of evidence suggest that insulin resistance per se or hyperinsulinemia, or both, could result in increased sodium reabsorption, enhanced vascular tone, and activation of the adrenergic nervous system. Alternatively, this state could be associated with abnormal regulation of the renin-angiotensin system. Environmental factors can exacerbate this defect. For example, if patients with insulin resistance gain weight or receive drugs that increase insulin resistance, such as diuretics or beta-blockers, the hypertension may be worsened. However, it is still unclear whether the insulin resistance is a marker for some other abnormality or a primary defect in these patients.
Calcium and Hypertension

In 1982 McCarron and co-workers\(^2\) reported that dietary calcium intake in humans with hypertension was lower than in normotensive control subjects, and the inverse relation between blood pressure and calcium intake has been confirmed in epidemiologic studies. \(^3\) These studies also suggest an association between the level of calcium intake and the degree of sensitivity of the blood pressure to sodium intake. In part, this may not be surprising given the known relationship between the reabsorption of calcium and that of sodium by the proximal tubule of the kidney. Blood pressure, in part, may also correlate with magnesium intake, at least in women.\(^4\) Indeed, the relative risk of developing hypertension was 0.65 when both magnesium (<200 mg/day) and calcium (<400 mg/day) intakes were lower. There appears to be a critical threshold for the effect of calcium intake on blood pressure, the effect not being evident unless calcium intake is less than 700 to 800 mg/day.\(^5\) Thus, increasing calcium intake above this threshold may not modify blood pressure.

Clinical trials designed to evaluate the validity of these observational data have provided equivocal results. However, a meta-analysis of these trials suggests that high calcium intake causes a minimal reduction of blood pressure in the general population and a modest reduction in individuals who already have hypertension. In one meta-analysis\(^6\) of results for 2412 subjects, high calcium intake reduced systolic blood pressure \(1.33\) mm Hg in the general population and \(4.3\) mm Hg in hypertensive patients. Hypertensive subjects also had a significant reduction in diastolic blood pressure \(1.5\) mm Hg. A second meta-analysis involving \(1231\) individuals did not demonstrate as large an effect.\(^7\)

Pathophysiologic Mechanisms

The mechanisms responsible for the impact of calcium intake on blood pressure are uncertain. Some studies have shown that parathyroid hormone (PTH) levels, on average, are higher in hypertensive subjects compared with normal control subjects, suggesting a potential role for PTH in mediating hypertension. However, PTH levels in hypertensive patients are still within the normal range and are not inappropriate for the level of ionized calcium. Thus, it is unclear whether the elevated PTH causes an increase in blood pressure or is simply a reflection of a modest change in calcium homeostasis. Furthermore, when infused, PTH is a vasodilator.\(^8\) and PTH inhibits contraction of vascular smooth muscle, presumably by inhibiting calcium entry.\(^9\)

Pang and Lewanczuk\(^10\) have suggested that there is a specific hypertensive factor from the parathyroid gland, distinct from PTH, that is increased in some patients with essential hypertension. Several studies in experimental hypertension suggest that the plasma level of the factor is elevated and that it can modify vascular smooth muscle function by increasing cytosolic calcium levels.\(^11\) In some patients with hyperparathyroidism and hypertension, parathyroid hypertensive factor is said to be elevated and becomes undetectable after parathyroidectomy as blood pressure decreases.\(^12\) However, only a single group has reported on the presence of such a factor.

PTH-related protein (PTHrP), a peptide with a structure similar to that of PTH but derived from a different gene, appears to share with PTH an ability to produce vasodilation. It has been suggested that a deficiency in PTHrP could lead to an elevated blood pressure, as this substance may be produced to counteract the effect of vasoconstrictors.\(^13\)

Finally, the active metabolite of vitamin D \((1,25\text{-dihydroxycholecalciferol})\) increases calcium uptake in cardiac and vascular smooth muscle cells, induces vascular contractions, and exerts a myotropic effect on vascular smooth muscle.\(^14\) Patients with low-renin or salt-sensitive hypertension tend to have an increase in both PTH and \(1,25\text{-dihydroxycholecalciferol}\) levels. With sodium loading, levels of both hormones increase further as the blood pressure increases. However, as with the changes in PTH levels, it is difficult to determine a cause-and-effect relationship. For example, the increase in \(1,25\text{-dihydroxycholecalciferol}\) levels in these patients may be secondary to the higher PTH level because PTH stimulates 1-hydroxylase activity.

In brief, epidemiologic and experimental evidence suggests that calcium and calcium-regulating hormones play a role in the control of vascular tone. Meta-analyses of clinical trials support a blood pressure-lowering effect of a high calcium intake, particularly in patients with essential hypertension. However, the mechanism or mechanisms involved and the relationship of the effects of calcium intake to the effects of sodium intake on blood pressure are still unclear. Finally, it is uncertain how many of these associations are primary versus secondary events.
Summary

Several lines of evidence suggest that many patients with essential hypertension have an endocrine basis for elevated blood pressure. Increased circulating hormone levels, changes in the responsiveness of target tissues to these hormones, and abnormalities in vascular tone can all contribute to the pathogenesis of the hypertension. Whether these endocrine abnormalities are primary or secondary events is unclear. Intriguingly, most do not fit the classical endocrine pattern for disease because hormonal overproduction is rare. Rather, there is a change in the response of target tissues to specific hormones, with associated adaptive responses probably the major contributor to the hypertensive process. Finally, a number of the abnormalities appear to have a major genetic component. This fact makes possible a more precise dissection of subgroups or phenotypes of hypertensive patients with the use of genetic markers.
RENIN-ANGIOTENSIN SYSTEM AND SECONDARY HYPERTENSION

In addition to its involvement in primary hypertension, the renin-angiotensin system is a major factor in the most common cause of secondary hypertension: renal disease. Indeed, many insights into the renin-angiotensin system have come from the study of patients with renal disease.

Renal Vascular Hypertension

Goldblatt and colleagues described the pathologic role of excess renin production in the hypertension associated with constriction of a renal artery, and 4 years after their observations, a nephrectomy in a hypertensive patient with a small kidney led to correction of the hypertension. Thus, renal vascular hypertension is defined as hypertension associated with either unilateral or bilateral ischemia. Unilateral renal vascular disease is likely to be the cause of elevated blood pressure in approximately 1% of the hypertensive population, and bilateral renal parenchymal disease is causative in another 2% to 4%.

It is important to distinguish between renal vascular disease and renal vascular hypertension. Perhaps 50% or more of subjects older than 60 years have renal vascular disease, only a minority of whom also have hypertension. This discrepancy is not surprising when one considers Goldblatt's original experiment, which documented that the lumen of the renal artery needs to be reduced to less than 30% of its original size before hypertension develops. Documentation of a functional abnormality in association with a radiologically defined renal arterial lesion is a critical diagnostic maneuver before therapeutic intervention.

Blacks and patients with diabetes mellitus seem to have a lower frequency of renal vascular hypertension, even though in both groups the incidence of renal vascular disease is higher than in the nondiabetic white population. Most patients with renal vascular disease have either atherosclerotic plaques or fibromuscular disease, and in 10% of the cases the lesion may not be in the main renal artery but in a segmental or branch artery.

Pathophysiology

The initiating event in the hypertension in subjects with renal disease is a reduction in perfusion pressure to the affected kidney, which stimulates the release of renin. The increased production of renin leads to increased angiotensin II levels and increased aldosterone secretion (secondary aldosteronism; see Fig. 15-6, Figure Not Available). As a consequence of the hyper-aldosteronism, sodium is retained and potassium is lost. The combination of angiotensin II-induced vasoconstriction and sodium retention increases blood pressure and leads to a natriuresis through the contralateral kidney. The increased levels of renin, elevated blood pressure, and sodium retention all act in concert to suppress renin production from the contralateral kidney, an important feature in the diagnostic evaluation of these patients. With bilateral renal arterial disease, the vasculature to both kidneys is compromised. Thus, the ability of the kidneys to excrete sodium is reduced, a gradual volume expansion occurs, and circulating renin levels are suppressed.

In long-standing unilateral renal artery stenosis, the contralateral kidney may become damaged secondary to the elevated blood pressure. Thus, the affected kidney is protected by the stenotic lesion and may ultimately suffer less damage, except for the ischemia induced by the stenosis. In these cases correction of the renal artery stenosis may not correct the elevated blood pressure, which can be sustained by the diffusely damaged contralateral kidney. Paradoxically, in this circumstance repair of the renal artery stenosis and removal of the contralateral kidney may normalize blood pressure.

Although renal vascular hypertension occurs in all age groups, the etiology varies. In individuals younger than 50 years, renal vascular hypertension is more common in women and is usually secondary to fibromuscular dysplasia of the renal artery. After the age of 50 it is more likely to be secondary to atherosclerosis and therefore more common in men. In addition to unilateral or bilateral vascular insufficiency, generalized renal ischemia can result from renal compression, such as in hydrocephalus. Initially the kidney is enlarged, but over time cortical atrophy develops.

Diagnosis

Renal vascular hypertension should be suspected in a normotensive individual of any age who has sudden onset of hypertension, in a known hypertensive subject with an acute acceleration of blood pressure, or in an individual younger than 30 years who has significant hypertension. Other clinical features are suggestive of this condition (Fig. 15-15). In general, if the hypertension is mild to moderate (diastolic blood pressure 105 mm Hg) with an onset between ages 30 and 60 years, the probability of renal vascular hypertension is low. In these patients, a detailed search for renal vascular hypertension is probably not warranted.

The diagnosis of renal vascular hypertension rests on two criteria: (1) the identification of a significant arterial obstruction and (2) evidence of excess renin secretion by one or both kidneys. The following tests are particularly useful in patients with fibromuscular dysplasia whose hypertension is renin-dependent and often cured by revascularization. In elderly individuals with renal artery stenosis that is probably secondary to atherosclerosis, consideration of the therapeutic options may modify the extent of testing.

Noninvasive Testing

Gadolinium-Enhanced Magnetic Resonance Angiography

This procedure is useful for defining anatomic lesions of one or both renal arteries. In some centers it is becoming the screening procedure of choice. Because gadolinium is not nephrotoxic, it may be particularly useful for individuals with renal insufficiency. The reported sensitivity (90% to 100%) and specificity (90% to 95%) are excellent for detecting a stenosis greater than 50%. Limitations of this technique include the requirement for breath holding, overestimation of stenosis,
inadequate visualization of segmental and accessory renal arteries, low availability of MR angiography, and cost. With future development, rapid imaging techniques may permit the acquisition of hemodynamic data by visualization of the vascular phases.

**Captopril Renography**

This is a functional test. Captopril, a converting enzyme inhibitor, is administered 30 to 60 minutes before the renogram. This test takes advantage of the dependence of the hemodynamics of the renal vasculature on angiotensin II because both the efferent and afferent arterioles are highly sensitive to the vasoconstrictor actions of angiotensin II. Indeed, balanced constriction of these two vessels maintains a normal glomerular filtration rate with a variety of changes in renal blood flow, so-called renal autoregulation. An increase in angiotensin II levels secondary to unilateral stenosis of a renal artery leads to vasoconstriction of the afferent and efferent arterioles in both kidneys. Because of the stenotic lesion on the affected side, blood flow to the glomerulus, and therefore glomerular pressure, on this side is determined primarily by the structural lesion.

When a converting enzyme inhibitor (which reduces intrarenal angiotensin II levels) is given, angiotensin II-induced afferent arteriolar constriction is reduced in both kidneys. In the unaffected contralateral kidney there is a concomitant reduction in efferent arteriolar tone, an increase in blood flow, and no change in glomerular filtration rate. However, in the stenotic kidney, where perfusion of the glomeruli is restricted by the stenosis, a reduction in efferent arteriolar tone results in a fall in glomerular filtration rate and reduced uptake and delayed excretion of the isotopic tracer as assessed by the renogram. The sensitivity (80% to 90%) and specificity (65% to 80%) of the captopril renogram are somewhat less than those of the MR angiogram, particularly in patients with renal insufficiency.

**Digital Subtraction Angiography**

This procedure is used in some centers, but because of the cost and the need for arterial rather than venous injection, its role as a screening test is limited.

**Renal Duplex Ultrasonography**

This test provides both a functional (Doppler) and an anatomic (B-mode imaging) evaluation of the renal artery and renal perfusion. Intrarenal ultrasonography measures the pulsatility index and blood flow acceleration during early systole. Its sensitivity (85% to 95%) and specificity (70% to 85%) fall between those of the captopril renogram and the MR angiogram. Although in theory this may be an ideal screening test, in practice its validity is dependent on the skills of the radiologist. Depending on operator availability, it is the procedure of choice in some centers.

**Invasive Testing**

The definitive test for a correctable renal vascular lesion is the combination of bilateral renal vein renin sampling and a selective renal arteriogram. As noted earlier, the renal arteriogram alone defines structural lesions but does not provide insight into their functional significance. Simultaneous bilateral renal vein renin measurement provides a valuable adjunct for predicting whether therapeutic intervention will modify the hypertension. When one kidney is ischemic and the other is normal, nearly all the circulating renin is produced by the affected side. As a result, the venous concentration of renin from the ischemic kidney is at least 1.5 times greater than that from the contralateral kidney. Theoretically, the renin concentration from the contralateral kidney should be the same as that in the peripheral circulation. Unfortunately, in many circumstances there is a varying degree of damage to the contralateral kidney, and therefore total suppression of renin from that kidney does not occur.

Some administer captopril before renal vein sampling to exaggerate the renin release from the stenotic side; all agents that suppress renin secretion, such as β-blockers, should be withheld before the study. If feasible, the patient should also have a low sodium intake to enhance renin release from the affected kidney. If these procedures are followed, approximately 80% of subjects with unilateral renin elevations that fulfill the preceding criteria have a beneficial response to therapeutic intervention.

**Treatment**

The treatment of choice in renal artery stenosis secondary to fibromuscular disease is renal angioplasty with or without insertion of a stent. Lesions at the ostium of the renal artery often do not respond well to angioplasty alone, but the addition of a stent may improve the outcome. Surgical revascularization, previously the primary approach, is used less frequently in older adults with atherosclerotic lesions because the surgical risk is high in these individuals. The same reservation may apply to angioplasty or stenting, or both, in comparison with medical treatment. Thus, surgery is usually reserved for individuals in whom angioplasty has proved unsuccessful.

Medical management to control blood pressure may be appropriate in elderly individuals, those who are not candidates for definitive corrective procedures, or those in whom these procedures have failed. Converting enzyme inhibitors or an angiotensin II receptor antagonist would be the treatment of choice, given the pathophysiology of the disease. However, for the reasons outlined earlier, these agents may worsen ischemia in the affected kidney. Converting enzyme inhibitors should be used with great caution in individuals who may have bilateral renal artery disease or arterial disease in a solitary kidney because these agents may reduce the glomerular filtration rate, cause renal hypoxia, and precipitate renal failure. Finally, a more aggressive therapeutic approach to preserve renal function may require invasive techniques.
Primary Reninism

Rarely, juxtaglomerular cell tumors of the kidney or ectopic tumors secrete renin. Individuals with such tumors have typical features of renal vascular hypertension: hypertension, elevated renin levels, hypokalemia, and hyponatremia. Most, however, are young and have high circulating levels of renin in the blood and also severe hypertension at the time of diagnosis. When a mass lesion is discovered in the kidney, there is unilateral renin secretion but no evidence of renal artery stenosis. Radiologic evaluation with CT scanning is invaluable. Surgical removal of the tumor cures the hypertension and the hyperreninemia.
Renal Parenchymal Disease

A variety of conditions can cause hypertension associated with renal parenchymal disease, the most common of which are hypertension per se, diabetes mellitus, and autoimmune disease (e.g., systemic lupus erythematosus). In these patients there is the potential, depending on the level of sodium intake, for a shift from a vasoconstrictor (angiotensin II) form of hypertension to a volume-sensitive hypertension secondary to decreased capacity for renal sodium excretion. The local (intrarenal) renin-angiotensin system is activated to a variable degree in these subjects. This activation results in an elevation of hydraulic pressure in the glomeruli (so-called glomerular hypertension) secondary to the vasoconstrictor effect of angiotensin II on the efferent arteriole as noted earlier.慢性 elevation of glomerular pressure leads to glomerular sclerosis and progressive loss of functioning nephron units.

The administration of converting enzyme inhibitors can slow the progression of renal damage in both diabetic and nondiabetic renal parenchymal disease. This effect appears to be an added action of converting enzyme inhibitors beyond their ability to lower systemic blood pressure, probably because they also selectively reduce renal glomerular pressure. Thus, agents that produce a decrease in systemic blood pressure equivalent to that accomplished by converting enzyme inhibitors do not afford the same degree of protection of renal function.

However, caution should be exercised in administering converting enzyme inhibitors to patients who may have an increased risk of bilateral renal artery stenosis. In such circumstances, instead of improvement, converting enzyme inhibitors would cause a sudden deterioration of renal function. Interestingly, the renal protective actions of converting enzyme inhibitors work in advanced diabetic nephropathy, a condition usually associated with hyporeninemic hypoaldosteronism. Because of the possible further reduction of aldosterone production with converting enzyme inhibitors in such patients, frequent measurement of serum potassium levels is mandatory, with institution of appropriate measures to reduce potassium levels if hyperkalemia occurs (e.g., low-potassium diet, potassium-wasting diuretics). In some instances, the converting enzyme inhibitor may have to be discontinued.
Hypertension during Pregnancy

Pregnancy-induced hypertension (PIH), or gestational hypertension, is defined as de novo hypertension arising during the second half of pregnancy. The literature regarding the prevalence and pathophysiological abnormalities in this disorder is clouded by studies that have often included pregnant patients with chronic hypertension. The classification of hypertensive disorders of pregnancy developed by the American College of Obstetricians and Gynecologists has been adopted by the National Institutes of Health:

1. Chronic hypertension: blood pressure greater than 140/90 mm Hg before pregnancy or before the 20th week of pregnancy
2. Preeclampsia: a systolic blood pressure increase of at least 30 mm Hg or a diastolic blood pressure increase of at least 15 mm Hg over prepregnancy or early pregnancy values combined with proteinuria (300 mg per 24 hours) or edema, or both
3. Preeclampsia superimposed on chronic hypertension: the same criteria as for item 2 but occurring in women with preexisting hypertension
4.Transient hypertension: a blood pressure increase similar to that in preeclampsia but without proteinuria or edema

Preeclampsia may also progress to eclampsia, defined as the occurrence of convulsions. As viewed according to the preceding classification, PIH includes preeclampsia and transient hypertension of pregnancy. However, the prognosis for each is different. The preeclampsia form of PIH is self-limited and usually does not occur in subsequent pregnancies, whereas...

### Cardiovascular Measurements

<table>
<thead>
<tr>
<th>Reduced circulating plasma volume</th>
<th>Increased systemic vascular resistance</th>
<th>Decreased cardiac index</th>
</tr>
</thead>
</table>

### Hormonal Indicators of Volume

<table>
<thead>
<tr>
<th>Reduced circulating plasma renin activity (PRA), angiotensin II, and aldosterone</th>
<th>Enhanced angiotensin II pressor responsiveness</th>
<th>Increased levels of atrial natriuretic peptide (ANP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increased levels of digitalis-like factor (DLF)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Other Hormonal Alterations

<table>
<thead>
<tr>
<th>Hyperinsulinemia and insulin resistance</th>
<th>Decreased prostaclin and/or elevated thrombocytopenic production</th>
</tr>
</thead>
</table>

### Risk Factors for Pregnancy-Induced Hypertension

PIH arises in approximately 5% of nulliparous pregnant women, and the prevalence appears to be related to both environmental and genetic factors. For example, the prevalence of PIH is double in indigent, inner-city pregnant women. Other risk factors include nulliparity, low dietary calcium intake, multiple gestations, black race, chronic hypertension, increasing age, inherited and acquired coagulation disorders such as protein S and protein C deficiencies and antiphospholipid antibodies, and a history of a mother who had this syndrome.

### Pathophysiology

Pathophysiological abnormalities in PIH should be divided into those in preeclampsia and those in transient hypertension of pregnancy. In normal pregnancy, plasma volume increases by 40%, associated with a reduction in peripheral vascular resistance (40% to 80%) and a rise in cardiac output, renal blood flow, and glomerular filtration rate. The renin-angiotensin-aldosterone system is activated despite the increase in plasma volume. This activation is believed to be related to prostaglandins, for example, prostacyclin and prostaglandin E₂, direct effects of estrogen; or an antiinflammatory action of progesterone.

A number of alterations have been reported in subjects with PIH, although many investigations do not clearly distinguish between patients with preeclampsia and those with transient hypertension of pregnancy. Table 15-3 presents the findings in preeclampsia compared with normotensive pregnancy. Plasma volume is reduced in preeclampsia and systemic vascular resistance is increased. Paradoxically, the hormonal markers of volume are consistent with a volume-expanded state, relative suppression of the renin-angiotensin-aldosterone system, and increased levels of atrial natriuretic peptide and digitalis-like factor. This apparent paradox is unexplained unless these hormonal changes induce the volume changes or there is a misperception of extracellular fluid volume in this disorder. On the other hand, elevated levels of digitalis-like factor, an inhibitor of the Na⁺,K⁺-ATPase pump, could increase intracellular sodium levels, as reported in red blood cells in patients with PIH. The role of digitalis-like factor is controversial, however, because of disparate findings for digoxin-like immunoreactivity versus measurement by bioassay, that is, inhibition of Na⁺,K⁺-ATPase.

Enhanced maternal vascular reactivity is also seen in PIH with increased pressor sensitivity to infused angiotensin II even in normotensive phases of pregnancies, before patients progress to PIH. In contrast, in normal pregnancies pressor responsiveness to angiotensin II is blunted. One important hypothesis that could explain these observations includes decreased production of vasodilatory prostaglandins, such as prostaglandin E₁, and prostacyclin; in fact, decreased prostacyclin production may precede the appearance of hypertension in PIH. Others propose a relative increase in the vasocostriction prostaglandin, thromboxane, and subnormal nitric oxide. Reported cation abnormalities include increased intracellular calcium and enhanced responses of intracellular calcium to vasopressin in platelets of patients with PIH. These findings suggest that parallel changes might occur in vascular smooth muscle, where intracellular calcium is a major determinant of peripheral vascular resistance. It is unclear whether the abnormalities in PIH reflect initiating events or are secondary alterations that sustain the elevation of blood pressure.

One unifying hypothesis to explain PIH is uteroplacental hyperperfusion. Perhaps secondary to structural abnormalities in the spiral arteries supplying the uterus, uterine blood flow is impaired. Through unknown mechanisms, prostaglandin production is reduced and there is generalized endothelial damage (endothelin levels are elevated in PIH). Subsequently, platelet aggregation is enhanced, fibrin is deposited in the glomeruli, and proteinuria occurs. Although this is an attractive hypothesis, a number of links need to be established.

### Treatment

There is little evidence that sodium restriction improves the outcome of pregnancy in women with PIH or is effective for prophylaxis against PIH. On the other hand, some studies suggest that volume expansion may improve both blood pressure and outcome in preeclampsia. It now appears that a moderate to liberal...
sodium diet should be recommended to pregnant women and that the diet should be adequate in calcium. Bed rest with lateral recumbency is commonly recommended for women with PIH and is believed to result in hemodynamic improvements, including increased renal and uterine blood flow and reduced peripheral vascular resistance with lowering of systemic blood pressure.

Low-dose aspirin (50 to 100 mg/day) may reduce the incidence of PIH, although the results of several large trials remain controversial. The rationale for low-dose aspirin is to restore the balance between vasodilator and vasoconstrictor prostaglandins (primarily by reducing thromboxane A₂ but without affecting prostacyclin levels). Antihypertensive agents are usually given when diastolic blood pressure exceeds 100 mm Hg. Traditional agents include methyldopa and hydralazine. β-Blockers, such as prazosin, and calcium channel blockers, primarily nifedipine, are being used with increased frequency because of their greater efficacy. These agents also appear to have acceptable safety profiles, although they have not been used as long as the traditional drugs. ACE inhibitors and angiotensin receptor antagonists are contraindicated in PIH because of adverse effects on kidney development in the fetus.
PRIMARY MINERALOCORTICOID EXCESS STATES

Primary mineralocorticoid excess states are characterized by suppressed PRA and hypokalemia and include primary aldosteronism.

### TABLE 15-4 -- Mineralocorticoid Excess States Associated with Low Plasma Renin Levels

<table>
<thead>
<tr>
<th>Syndrome</th>
<th>Proportion of Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary aldosteronism</td>
<td>65%</td>
</tr>
<tr>
<td>Aldosterone-producing adenoma (APA)</td>
<td></td>
</tr>
<tr>
<td>Idiopathic hyperplasia (idiopathic hyperaldosteronism)</td>
<td></td>
</tr>
<tr>
<td>Adrenocortical carcinoma</td>
<td></td>
</tr>
<tr>
<td>Glucocorticoid-remediable aldosteronism (GRA)</td>
<td></td>
</tr>
<tr>
<td>Congenital adrenal hyperplasia</td>
<td></td>
</tr>
<tr>
<td>11-Hydroxylase deficiency</td>
<td></td>
</tr>
<tr>
<td>17-Hydroxylase deficiency</td>
<td></td>
</tr>
</tbody>
</table>

### Increased Mineralocorticoid Action

- Apparent mineralocorticoid excess (AME)
- Congenital
- Licorice ingestion
- Ectopic corticotropin production
- Liddle's syndrome

Deoxycorticosterone (DOC)-secreting tumors, and inherited diseases (Table 15-4).

**Primary Aldosteronism**

Primary aldosteronism, the cause of approximately 0.05% to 2.2% of all unselected cases of hypertension, was first described in 1955 by Conn in conjunction with an aldosterone-producing adrenal adenoma (APA). Other etiologies include idiopathic bilateral hyperplasia (idiopathic hyperaldosteronism) and inherited entities. Its prevalence may be as high as 10% in hypertensive patients, depending on the study population and the criteria used. The prevalence of tumors remains fairly constant. However, the prevalence of bilateral hyperplasia can vary greatly.

#### Clinical Features

The clinical symptoms of primary aldosteronism are nonspecific and result from potassium depletion. Neuromuscular symptoms (weakness, periodic paralysis, cramps, or tetany), fatigue, and paresthesias are not uncommon; polyuria and nocturia probably result from a hypokalemia-induced renal concentrating defect. Despite the continuous high levels of aldosterone, patients rarely exhibit edema, presumably owing to escape, in which the sodium-retaining effects of chronic mineralocorticoid excess are lost. Intracellular potassium depletion can also impair insulin secretion and cause glucose intolerance or overt diabetes mellitus. Resetting of the osmostat can occur in primary aldosteronism, as evidenced by slightly higher than normal serum sodium levels. This is a useful clinical point because there is a tendency for a reduced serum sodium level in states of secondary aldosteronism.

The hypertension associated with primary aldosteronism is usually moderate to severe with mean blood pressures (± standard deviation) of 184 ± 8/112 ± 16 mm Hg. However, some patients have malignant hypertension, and others have normal or only mildly elevated blood pressure. Individuals with APA tend to have higher blood pressures than those with idiopathic hyperaldosteronism (IIA). Patients with primary aldosteronism may be refractory to conventional antihypertensive agents and may experience severe hypokalemia after institution of potassium-wasting diuretics such as hydrochlorothiazide. Although it would be expected that the hypertension is related to volume expansion, measurements of extracellular sodium spaces in patients with APA are usually normal while peripheral resistance is increased.

End-organ damage with primary aldosteronism is variable. In general, the left ventricular hypertrophy is disproportionate to the level of blood pressure when compared with the situation in essential hypertension.

**TABLE 15-5 -- Causes of Primary Aldosteronism and Their Frequencies**

<table>
<thead>
<tr>
<th>Syndrome</th>
<th>Proportion of Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aldosterone-producing adenoma, including renin-responsive adenoma</td>
<td>65%</td>
</tr>
<tr>
<td>Idiopathic hyperaldosteronism, including primary adrenal hyperplasia</td>
<td>30%-40%</td>
</tr>
<tr>
<td>Glucocorticoid-remediable aldosteronism (GRA)</td>
<td>1%-3%</td>
</tr>
</tbody>
</table>

With removal of an aldosteronoma, its regression occurs even if blood pressure does not become normal. In addition, structural damage to the kidney, cerebral circulation, and retinal vasculature occurs more frequently than would be assumed on the basis of the duration and level of blood pressure. As many as 50% of patients with primary aldosteronism may have proteinuria, and renal failure may occur in as many as 15%. Thus, it is probable that, independent of aldosterone's effect on blood pressure, its excess production induces cardiovascular damage.

Mineralocorticoid excess states are associated with vascular remodeling that results in perivascular fibrosis and vascular wall thickening. Cardiac fibrosis has also been reported in postmortem studies of patients with adrenal adenoma with mineralocorticoid hypertension. An excessive number of renal cysts has also been noted in association with hypokalemia and hyperaldosteronism (in as many as 60% of cases). However, other studies reported a prevalence of renal cysts similar to that in the normal population.

Etiologies of primary aldosteronism are shown in Table 15-5.

**Aldosterone-Producing Adenoma**

The solitary APA is the most common cause of primary aldosteronism and accounts for approximately 65% of cases. These lesions are usually less than 2 cm in
Idiopathic Hyperaldosteronism

Idiopathic hyperplasia (bilateral adrenal hyperplasia) is the cause of approximately 30% of cases of primary aldosteronism. Microscopically, the glands show hyperplasia of the zona glomerulosa accompanied by adrenocortical nodules. The aldosterone excess in IHA is usually milder than in APA; as a result, the biochemical abnormalities such as hypokalemia and suppression of PRA are usually less severe than in APA.

Because the plasma aldosterone response to infused angiotensin II is exaggerated compared with that in normal individuals or patients with APA, IHA has been hypothesized to represent a syndrome of enhanced responsiveness to angiotensin II. Others suggest that IHA is a form of essential hypertension representing one end of the distribution of aldosterone production in essential hypertension and is therefore related to low-renin essential hypertension, in which an enhanced aldosterone response to angiotensin II is seen (see earlier).

Glucocorticoid-Remediable Aldosteronism (GRA)

See later.

Regulation of Aldosterone Secretion

The renin-angiotensin system is suppressed in primary aldosteronism and does not contribute to the regulation of aldosterone production. In this sense, aldosterone production is "autonomous." However, studies in patients with APA, for example, demonstrate that the adenomas are regulated by corticotropin and potassium. Thus, autonomy is defined by the failure of aldosterone production to respond to maneuvers that normally activate (upright posture) or suppress (sodium loading) the renin-angiotensin system. A variant form of APA has been described in which the adenomas are renin responsive (see later). Moreover, in IHA aldosterone production is usually responsive to stimuli that activate the renin-angiotensin system. On the other hand, adrenal carcinomas are truly resistant to all secretagogues.

Diagnosis

Screening Tests

Spontaneous hypokalemia is commonly present. Some patients may have normal potassium levels, possibly because of self-selected dietary sodium restriction, because aldosterone-induced renal potassium wasting is diminished by decreased sodium delivery to the distal nephron. When potassium-wasting diuretics are given as antihypertensive agents, hypokalemia is more frequent and more severe. Serum potassium measurement is not a good screening test in GRA because many patients have normal potassium levels.

PRA is suppressed in almost all patients (<1.0 ng/mL per hour [<0.8 nmol/L per hour]) and does not increase appropriately (>2 ng/mL per hour [>1.6 nmol/L per hour]) after dietary sodium restriction or after acute diuretic administration with furosemide followed by 90 to 120 minutes of upright posture. Although a subset of patients with essential hypertension (25%) also have low PRA, documentation of a normal or high stimulated PRA level excludes primary aldosteronism.

Given the high prevalence of suppressed PRA levels in essential hypertension, documentation of a concomitant elevation of plasma aldosterone (PA) makes the diagnosis of primary aldosteronism more likely (Fig. 15-16). Thus, a PA:PRA ratio greater than 30 is suggestive and a ratio of 50 or more is virtually diagnostic of primary aldosteronism (when PRA is expressed as ng/mL per hour and PA as ng/dL). To improve the accuracy of this test, salt intake should not be restricted and serum aldosterone levels should be greater than 500 pmol/L (>15 ng/dL). Because a number of drugs (ACE inhibitors, β-blockers, and spironolactone) alter PRA levels, such antihypertensives should be withdrawn for 2 to 4 weeks if possible (6 to 8 weeks for spironolactone) before determining the PA:PRA ratio. Hypokalemia reduces aldosterone levels, and diagnostic studies should be performed with patients in a potassium-repleted state. Specimens should be obtained after 2 hours of upright posture because stimulated ratios have been shown to have better diagnostic accuracy than supine values.

The captoril test can also be used to diagnose primary aldosteronism. One protocol is to administer 50 mg orally at 9:00 AM; blood samples are obtained before and 90 minutes later. In normotensive individuals or essential hypertensives, acute inhibition of ACE decreases angiotensin-regulated aldosterone production. However, in primary aldosteronism aldosterone levels fail to decline because the renin-angiotensin system is suppressed and aldosterone production is autonomous. A postcaptopril aldosterone reduction of more than 20%, usually to less than 410 pmol/L (<15 ng/dL), is considered a normal response. Although the sensitivity of this test ranges from 90% to 100%, the specificity is significantly less (50% to 80%).

Diagnosis of Autonomus Aldosterone Production

Oral sodium loading for 3 days followed by a 24-hour urine collection for determination of aldosterone excretion can discriminate primary aldosteronism from essential hypertension with excellent sensitivity and specificity (96% and 93%, respectively). A 24-hour urinary aldosterone excretion rate greater than 28 to 39 nmol/day (10 to 14 µg/day) in the presence of urinay sodium excretion greater than 250 mmol/day is considered diagnostic of primary aldosteronism. Autonomous aldosterone production can also be demonstrated by acute intravenous volume expansion with isotonic saline. Isotonic saline is administered intravenously at a rate of 500 mL/hour for 4 to 6 hours. Postsaline plasma aldosterone levels greater than 280 pmol/L (>10 ng/dL) or more stringent value of greater than 140 to 220 pmol/L (>5 to 8 ng/dL) has also been proposed to confirm the diagnosis of autonomous aldosterone production.
Hormonal testing provides supportive data, especially if radiography fails to show a solitary tumor. The posture test is the most common study used in the differential diagnosis of primary aldosteronism. Samples are collected for PRA and plasma aldosterone when the subject is recumbent and after 2 to 4 hours of standing or walking. The response in normal subjects and those with essential hypertension is an increase in plasma aldosterone levels of at least 50% compared with recumbent levels. Before testing, hypokalemia should be corrected with oral potassium supplementation. The accuracy of the test is enhanced by simultaneous measurement of recumbent and upright cortisol levels.  

In patients with APA, aldosterone levels generally decline in parallel with the circadian secretion of cortisol, the so-called anomalous postural response. This is because the renin-angiotensin system is suppressed in patients with APA, and therefore changes in posture do not stimulate increased production of aldosterone. In contrast, in IHA there is usually an increase in renin and aldosterone levels in response to assumption of the upright posture. The predictive value of the posture test in distinguishing between APA and IHA approaches 90%; however, the specificity of this test is reduced in variant forms of APA and IHA (see later). A postural decline in plasma aldosterone is also seen in GRA because aldosterone secretion is regulated solely by corticotropin in this disorder. Blood levels of 18-OH-corticosterone, an intermediate of the aldosterone biosynthetic pathway, are generally greater than 2800 nmol/L (100 ng/dL) in subjects with APA, whereas patients with IHA have lower levels.  

Documentation of the unique hybrid 18-oxygenated cortisol compounds 18-oxocortisol and 18-OH-cortisol in a 24-hour urine collection can be used to make the diagnosis of GRA and also differentiate APA from IHA (see later). In contrast to modest elevations in APA and normal levels in IHA, levels of these compounds are 10-fold above normal in GRA. A major drawback is the lack of general availability of assays for these compounds.

Radiologic Studies

Tomography using spiral CT techniques is the imaging modality of first choice. It is critical to make a biochemical diagnosis before imaging the adrenal because of the 2% to 10% incidence of nonfunctioning adrenal masses in CT studies of the abdomen. The use of adrenal scintigraphy with NP-59 (I-131 I-iodomethyl-19-norcholesterol) can differentiate APA from IHA because lateralization is seen in the former disorder. However, a lateralizing scan lacks specificity because it may also be seen with adrenal adenomas that do not produce aldosterone.

Adrenal venous sampling, which should be reserved for cases in which diagnostic imaging and biochemical studies are inconclusive, is the most sensitive means of differentiating APA from IHA. In most cases of APA, the ratio of the ipsilateral to the contralateral aldosterone concentration is greater than 10:1. This procedure is diagnostic in more than 95% of cases when catheterization of the right adrenal vein is successful. However, the incidence of unsuccessful procedures can be as high as 25%. Furthermore, adrenal venous sampling is invasive and is associated with a small but significant risk of venous thrombosis, adrenal hemorrhage, or adrenal insufficiency.

Nonclassical Variants of Primary Aldosteronism

The aldosterone-producing renin-responsive adenoma (APRA) and primary adrenal hyperplasia are variants of primary aldosteronism that represent important exceptions in terms of both diagnosis and treatment. In contrast to findings in APA, the changes in plasma renin (features characteristically considered diagnostic of IHA) after upright posture in subjects with APRA cause an increase in aldosterone levels. Thus, in patients diagnosed with primary aldosteronism the importance of adrenal imaging in documenting a solitary lesion is emphasized by this unusual entity.

Primary unilateral adrenal hyperplasia is another variant form of primary aldosteronism. The biochemical features resemble those in APA: no increase or a decline in aldosterone levels in response to upright posture and elevated levels of urinary 18-OH-cortisol and 18-oxocortisol. The syndrome can be ameliorated by unilateral adrenalectomy or a reduction in adrenal mass.

Therapy

Surgery is the treatment of choice for patients with APA, APRA, and primary adrenal hyperplasia. Cure rates (defined as blood pressure less than 140/90 without medications 6 to 12 months after surgery) vary between 35% and 50%. All patients should receive medical treatment before surgery to control blood pressure and replete potassium stores (see later). Persistent postoperative hypertension may be related to the chronicity or severity of hypertension, the presence of endorgan changes, or concurrent essential hypertension. On the other hand, the hypertension in IHA responds poorly to bilateral adrenalectomy, although the potassium-wasting state is reversed. Pharmacologic treatment is the therapy of choice for IHA, for preoperative management of APA, or when the patient is not a surgical candidate. A sodium-restricted diet (sodium < 2 g/day) is also prescribed in conjunction with pharmacologic treatment to minimize potassium wasting and lower the blood pressure.

Spironolactone, a competitive antagonist of aldosterone, has traditionally been the drug of first choice, with doses of 100 to 500 mg/day usually being required. Spironolactone also blocks testosterone biosynthesis and action, resulting in erectile dysfunction, decreased libido, and gynecomastia in men; menstrual irregularities are seen in women. Amiloride blocks the apical sodium channel in the distal nephron and is an alternative to spironolactone; it is given in divided doses starting at 5 mg twice daily with a maximal dose of 15 mg twice daily.

The sustained-release formulation of the calcium channel blocker nifedipine (dose range 30 to 90 mg/day) has also been used in the medical management of primary aldosteronism because this compound inhibits aldosterone biosynthesis in vitro. However, the antihypertensive response to this agent alone in primary aldosteronism is disappointing, and nifedipine should be viewed as a second-line agent. ACE inhibitors and angiotensin II receptor antagonists may also have a role in the medical management of IHA because the response to angiotensin II may be exaggerated in this disorder.
Genetic Basis for Mineralocorticoid Excess States

Glucocorticoid-remediable aldosteronism (GRA) is the commonest heritable form of hyperaldosteronism. In the following syndromes, the steroids responsible for the mineralocorticoid excess states include DOC and cortisol. As a result of the suppression of the renin-angiotensin system, the aldosterone levels are lownot elevated as in the previously described syndromes. Some forms of congenital adrenal hyperplasia have a mineralocorticoid component.

Other genetic causes include a mutation in enzymes not in the biosynthetic pathway or ion channels important in mediating or mimicking the action of aldosterone.

Hypermineralocorticoidism/Suppressed Plasma Renin Activity

Congenital Adrenal Hyperplasia

DOC excess is seen in several hypertensive forms of congenital adrenal hyperplasia or rarely in neoplasms (adenoma or carcinoma) that overproduce DOC. Congenital adrenal hyperplasia results from a deficiency in cortisol biosynthesis (see Chapter 14). Deficiencies in both 11-hydroxylase (CYP11B) and 17-hydroxylase (CYP17) are associated with hypertension and hypokalemia. In 11-hydroxylase deficiency, the impaired conversion of 11-DOC to corticosterone results in accumulation of DOC, a potent mineralocorticoid. Mineralization in females, usually occurring in children, results from shunting into the androgen pathway. Blood levels of DOC, 11-deoxycorticol, and adrenal androgens are characteristically elevated. This form of congenital adrenal hyperplasia is more prevalent in Middle Eastern Moslems and Jews.

17-Hydroxylase deficiency is characterized by hypogonadism, hypokalemia, and hypertension. As with 11-hydroxylase deficiency, this disorder is a result of decreased production of cortisol with shunting into the unblocked mineralocorticoid pathway. Because 17-hydroxylase is required for the biosynthesis of gonadal testosterone and estrogen, a defect in this enzyme in both sexes is associated with sexual immaturity, high gonadotropin levels, and low urinary 17-ketosteroid excretion. Females have primary amenorrhea and lack of development of secondary sexual characteristics. Males may have ambiguous external genitalia or a female phenotype (male pseudohermaphroditism). Blood levels of 17-oxo-progesterone are low, and corticosterone and DOC levels are elevated.

The elevated blood pressure and hypokalemia in both syndromes result from elevated levels of DOC, a potent mineralocorticoid; excessive sodium retention and hypertension result and lead to suppression of PRA and low levels of aldosterone. The genetic lesions causing 11-hydroxylase deficiency are in the gene that encodes CYP11B. A large number of mutations in the CYP17 gene can cause 17-hydroxylase deficiency.

Glucocorticoid suppression with dexamethasone or prednisone restores normal levels of DOC and reverses the mineralocorticoid excess state in these adrenal hyperplasia syndromes. Caution must be exercised to avoid overdosing and induction of Cushing’s syndrome.

Glucocorticoid-Remediable Aldosteronism

This syndrome is inherited in an autosomal dominant fashion and is probably responsible for fewer than 3% of cases of primary aldosteronism. GRA is characterized by hypertension of early onset that is usually severe and refractory to conventional antihypertensive therapies. Prospective screening of GRA pedigrees has revealed that many affected individuals are not hypokalemic. The sine qua non of GRA is aldosterone production that is solely under the control of corticotropin. As a result, the syndrome can be mitigated by exogenous glucocorticoid therapy.

GRA is caused by a chimeric gene duplication that results from unequal crossing over between the highly homologous 11-hydroxylase (CYP11B) and aldosterone synthase (CYP17) genes. This chimeric gene contains 3 corticotropic-responsive portion of the promoter from the 11-hydroxylase gene fused to the 5 coding sequence of the aldosterone synthase gene. The result is ectopic expression of aldosterone synthase activity in the cortisol-producing zona fasciculata. Thus, mineralocorticoid production is regulated by corticotropin instead of the normal secretagogue, angiotensin II. This mutation results in overproduction of aldosterone and also the characteristic hybrid steroids 18-oxocortisol and 18-OH-cortisol, which can be measured in the urine to make the diagnosis.

Genetic Testing

The diagnosis of GRA was initially based on the family history and the clinical response to dexamethasone suppression. Subsequently, GRA was diagnosed by demonstrating markedly elevated levels of 18-oxocortisol and 18-OH-cortisol in a 24-hour urine collection. However, the discovery of the genetic basis of GRA by Lifton and co-workers made it possible to make a genetic diagnosis from a peripheral blood sample with the use of Southern blotting techniques. Genetic testing is a sensitive and specific means of diagnosing GRA and obviates the need to measure the urinary levels of 18-oxocortisol and 18-OH-cortisol or to perform dexamethasone suppression testing. Genetic analysis can be arranged by calling the International GRA Registry (617-732-5011). Because some individuals with GRA do not have hypokalemia and have only mild hypertension, other clues that may indicate the need for a genetic test include suppressed plasma renin levels and juvenile-onset hypertension, a history of early-onset hypertension in first-degree relatives, and a history of early hemorrhagic stroke in the subject or relative.

Therapy

GRA is unique among the syndromes of primary aldosteronism in that the underlying pathophysiologic abnormality is the regulation of aldosterone production solely by corticotropin. As a result, glucocorticoid treatment usually reverses the syndrome. Of great importance is an awareness of the potential toxicity (Cushing’s syndrome) of excessive doses of glucocorticoids, especially with the use of dexamethasone in children. When a decision to use glucocorticoids is made, the smallest effective dose of shorter-acting agents such as prednisone or hydrocortisone should be prescribed in relation to body surface area (hydrocortisone, 10 to 12 mg/m² per day). Target blood pressure in children should be guided by age-specific blood pressure percentiles. Children should be monitored by pediatricians with expertise in glucocorticoid therapy, with careful attention paid to preventing retardation of linear growth by overtreatment. Therapeutic alternatives in treating hypertension in GRA are mineralocorticoid antagonists, which also avoid the adverse effects of chronic glucocorticoid therapy. Amiloride and spironolactone are effective as monotherapies in most patients with GRA.

Increased Mineralocorticoid Action/Low Plasma Renin Activity

Apparent Mineralocorticoid Excess

This syndrome is the result of impaired activity of the enzyme 11-hydroxysteroid dehydrogenase (11-HSD), which normally inactivates cortisol in the kidney by converting it to cortisone. As a result of the enzyme deficiency, high levels of cortisol accumulate in the kidney. The characteristic abnormal urinary cortisol metabolite profile seen in apparent mineralocorticoid excess also reflects decreased 11-HSD activity (ratio of cortisol to cortisone increased 10-fold compared with the normal ratio of approximately 1). As a result of elevated intrarenal levels, cortisol binds to mineralocorticoid receptors in the distal tubule, which are normally sites of aldosterone binding.

Underlying the pathogenesis of apparent mineralocorticoid excess is the nonselectivity of renal mineralocorticoid receptors, which in vitro bind cortisol with affinity equal to that of aldosterone. Thus, 11-HSD normally excludes physiologic glucocorticoids from nonselective mineralocorticoid receptors by converting them to the inactive 11-keto compound, cortisone. 11-HSD has bidirectional activity in different tissues, acting primarily as a reductase in the liver and a dehydrogenase in the
kidney. These different activities are the consequence of two isoenzymes that are expressed in liver and kidney, respectively.

Decreased 11-HSD activity may be hereditary or secondary to pharmacologic inhibition of enzyme activity by glycyrrhetinic acid, the active principle of licorice root and some chewing tobaccos. The hereditary form contains mutations in the gene coding for isoenzyme 2. The phenotype of patients with apparent mineralocorticoid excess includes hypertension, low PRA levels, hypokalemia, normal plasma cortisol levels, and low plasma aldosterone levels.

Treatment has been difficult, although some success has been achieved with use of a high-potency glucocorticoid to suppress endogenous cortisol production. Although the synthetic steroid can also bind to the mineralocorticoid receptor, its concentration is far less than that of the endogenous cortisol. Alternatively, the mineralocorticoid receptor antagonist spironolactone could be given. However, its antiandrogenic and progestational side effects limit its long-term use, particularly in children.

The mineralocorticoid excess state commonly seen in patients with the ectopic corticotropin syndrome is believed to be related to the high rates of cortisol production that cause a relative deficiency of 11-HSD activity. However, DOC levels are high and could account for the hypokalemia in this disorder.

**Liddle’s Syndrome**

This syndrome is inherited as an autosomal dominant disorder in which affected subjects present with hypertension, suppressed PRA, low aldosterone levels, and usually hypokalemia. The hypokalemic state cannot be corrected by administration of the antimineralocorticoid spironolactone but is ameliorated by triamterene or amiloride, agents that block renal sodium reabsorption and potassium secretion by mineralocorticoid receptor-independent mechanisms. This disorder is caused by mutations in the subunits of the renal sodium epithelial channel. The amiloride-sensitive epithelial channel, considered the rate-limiting step for sodium absorption in the distal nephron, is composed of three subunits \( \alpha \), \( \beta \), \( \gamma \); mutations have been found in two of these subunits. As a result of these mutations, constitutive activity of the epithelial channel leads to increased sodium absorption and volume expansion.

**Hypermineralocorticoidism**

Occasionally, hyperaldosteronism occurs without edema or hypertension (Bartter’s and Gitelman’s syndromes). Bartter’s syndrome is usually characterized by severe hyperaldosteronism with hypokalemic alkalosis, substantial increases in PRA, hypercalcuria, no edema, normal blood pressure, and onset usually in childhood. The pathogenesis results from a defect in renal sodium or chloride conservation, or both. The hyperaldosteronism-induced hypokalemia can be accentuated by an additional defect in renal potassium conservation in some individuals. In many cases, Bartter’s syndrome is caused by a mutation in the renal Na-K-2Cl cotransporter gene.

Gitelman’s syndrome is an autosomal recessive trait also characterized by renal sodium wasting. As a result, as with Bartter’s syndrome, there is activation of the renin-angiotensin-aldosterone system. Affected subjects therefore have low serum potassium and magnesium levels, high serum bicarbonate, and low blood pressure. In contrast to that in Bartter’s syndrome, urinary calcium excretion is reduced. Gitelman’s syndrome results from mutations that lead to loss of function of the renal thiazide-sensitive Na-Cl cotransporter.
OTHER ENDOCRINE DISORDERS ASSOCIATED WITH HYPERTENSION

Glucocorticoid Excess (Cushing's Syndrome)

Cushing's syndrome, characterized by hypersecretion of cortisol, is associated with elevations in blood pressure in more than 80% of cases (see Chapter 14). Diastolic blood pressure exceeds 100 mm Hg in more than 50% of patients with endogenous Cushing's syndrome. On the other hand, the incidence of hypertension is lower and more variable in patients treated with exogenous glucocorticoids. Nevertheless, hypertension and associated metabolic abnormalities (diabetes mellitus and hyperlipidemia) probably account for the atherosclerotic cardiovascular morbidity and mortality seen in spontaneous Cushing's syndrome.

Mineralocorticoid production is usually normal in endogenous Cushing's syndrome. In Cushing's disease, commonly secondary to corticotropin hypersecretion from a pituitary microadenoma, aldosterone and renin levels are usually normal and DOC levels are normal or increased modestly. On the other hand, in ectopic corticotropin syndrome, increased mineralocorticoid activity and hypokalemia are the rule as a result of elevated levels of DOC and mineralocorticoid effects of high levels of cortisol. In adrenal carcinomas, DOC and aldosterone may also be elevated. Thus, in adrenal carcinomas and in ectopic corticotropin secretion, mineralocorticoid production may contribute to the hypertension; in such situations, PRA is usually suppressed.

The elevation of blood pressure by cortisol and synthetic glucocorticoids (which have minimal mineralocorticoid activity) is mediated by multiple mechanisms. Glucocorticoids increase cardiac output and activate the renin-angiotensin system by increasing the hepatic production of angiotensinogen. Other actions of glucocorticoids include reduction of the synthesis of vasodilatory prostaglandins secondary to inhibition of phospholipase A₂, thus blocking the release of arachidonic acid from phospholipids. There is also evidence for reduction of the components of the kallikrein-kinin system as well as enhanced pressor sensitivity to endogenous vasoconstrictors (epinephrine and angiotensin II). Glucocorticoids may also promote sodium influx into vascular smooth muscle cells.

The screening for endogenous cortisol excess is accomplished by measuring the response of plasma cortisol to the 1-mg dexamethasone suppression test or by the measurement of elevated levels of free cortisol in a 24-hour urine collection. Further studies to determine the etiology of the cortisol excess state are outlined in Chapter 14.
Thyroid Disease

Hypothyroidism

Hypothyroidism may account for 1% to 2% of cases of diastolic hypertension in the general population. In a large series of patients screened for the secondary forms of hypertension by age, the prevalence of hypothyroidism was 3%. In that study, hypothyroidism was thought to be a significant cause of secondary hypertension, especially in women older than 70 years. That hypothyroidism can actually cause hypertension was shown by Streeten and colleagues in another study in which 32% of hypertensive hypothyroid patients had a fall in diastolic blood pressure to 90 mm Hg or less after replacement levothyroxine treatment and withdrawal of all hypertensive drugs. Postulated mechanisms for the elevation of blood pressure include extracellular volume expansion and elevation in systemic vascular resistance.

Hyperthyroidism

In contrast to the diastolic hypertension associated with hypothyroidism, hyperthyroidism usually causes elevated systolic blood pressure. Thyrotoxic patients usually have tachycardia, high cardiac output, increased stroke volume, and decreased peripheral vascular resistance (see Chapter 11). These hemodynamic alterations are usually ameliorated by -blocker therapy. In elderly patients, atrial fibrillation may be the sole manifestation of thyrotoxicosis.
Acromegaly

Hypertension occurs in one third of patients with acromegaly (see Chapter 8), presumably owing to sodium retention with resultant extracellular volume expansion. This retention of sodium, in the context of an increase in glomerular filtration rate and low PRA, is the consequence of uncharacterized antinatriuretic actions of growth hormone. Sodium retention is also a complication of exogenous administration of growth hormone. The prevalence of primary aldosteronism in acromegaly appears to be increased. As previously discussed, in primary aldosteronism, plasma renin levels are suppressed and aldosterone levels are increased. In acromegaly, plasma renin levels also are suppressed owing to volume expansion but aldosterone levels are not increased.
Hyperparathyroidism

In contrast to the overall incidence of 0.1% of primary hyperparathyroidism in the general population, this disorder occurs in approximately 1% of hypertensive patients (see Chapter 26). Conversely, approximately, 30% to 40% of individuals with hyperparathyroidism are hypertensive. The mechanisms are unclear because there is no direct correlation with the elevated PTH or calcium levels. Hypertension may or may not remit after successful parathyroidectomy.

---

**TABLE 15-6 -- Exogenous Causes of Secondary Hypertension**

<table>
<thead>
<tr>
<th>Cause</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mineralocorticoids (licorice and licorice-containing chewing tobacco; fludrocortisone)</td>
</tr>
<tr>
<td>Growth hormone</td>
</tr>
<tr>
<td>Glucocorticoids</td>
</tr>
<tr>
<td>Gonadal steroids</td>
</tr>
<tr>
<td>Oral contraceptives</td>
</tr>
<tr>
<td>Androgens</td>
</tr>
<tr>
<td>Sympathomimetic amines (amphetamines, cocaine)</td>
</tr>
<tr>
<td>Cyclosporine</td>
</tr>
</tbody>
</table>

Because the blood pressure response to correction of primary hyperparathyroidism is variable and hypertension is not a clear-cut manifestation of the hyperparathyroid state, hypertension is considered a minor criterion for recommending surgery to patients with mild asymptomatic primary hyperparathyroidism. On the other hand, surgery may cause regression of myocardial hypertrophy in normotensive patients with hyperparathyroidism and may also reverse the increased mortality associated with hyperparathyroidism in subjects younger than 70 years. The hypertension associated with hyperparathyroidism can also result as a complication of hypercalcemia-induced renal impairment or when this disorder is part of a MEN syndrome that includes pheochromocytoma or primary aldosteronism.
Exogenous Treatments

Treatment with certain hormones or pharmacologic agents can elevate blood pressure (Table 15-6). Fludrocortisone, a potent mineralocorticoid used in the treatment of primary adrenal insufficiency, can cause hypertension when administered in supraphysiologic doses, as to patients with orthostatic hypotension, to cause volume expansion. Licorice and chewing tobacco abuse can result in hypokalemia, sodium and water retention, and blood pressure elevation, as noted earlier. The administration of growth hormone in pharmacologic doses to patients who have received transplants or to severely catabolic hospitalized subjects for its protein-sparing actions can raise blood pressure.

The hypertension associated with cyclosporine treatment appears to be related to renal vasospasm and secondary volume expansion. Glucocorticoid treatment in supraphysiologic doses for inflammatory and allergic disorders frequently elevates blood pressure, as noted earlier, but less commonly than in endogenous Cushing’s syndrome. Oral contraceptive preparations containing higher dose estrogen-progesterone formulations are known to induce hypertension. It is not known whether current lower dose formulations cause hypertension. However, postmenopausal estrogen replacement therapy does not elevate blood pressure. Androgens in pharmacologic doses can also produce volume expansion and arterial hypertension.

Finally, as with the hypertension in pheochromocytoma, ingestion of sympathomimetic amines or substances that potentiate endogenous sympathetic nervous system activity (e.g., cocaine inhibits reuptake of catecholamines at adrenergic nerve endings) produces increased cardiac output, increased peripheral arteriolar vasoconstriction, and hypertension.
NORMOTENSIVE SYNDROMES

A diverse group of loss-of-function mutations in the sodium epithelial channel (ENaC), mineralocorticoid receptor (MR),

<table>
<thead>
<tr>
<th>Disorder</th>
<th>Genes Mutated</th>
<th>Potassium Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autosomal dominant pseudohypoaldosteronism type I (PHAI)</td>
<td>MR</td>
<td>Hyperkalemia*</td>
</tr>
<tr>
<td>Autosomal recessive (AR) PHAI</td>
<td>ENaC</td>
<td>Hyperkalemia</td>
</tr>
<tr>
<td>Gitelman’s syndrome (AR)</td>
<td>Thiazide-sensitive Na-Cl cotransporter in DCT</td>
<td>Hypokalemia</td>
</tr>
<tr>
<td>Bartter’s syndrome (AR)</td>
<td>Ion transporters in thick ascending loop of Henle</td>
<td>Hypokalemia</td>
</tr>
</tbody>
</table>

Key: ENaC, sodium epithelial channel; MR, mineralocorticoid receptor; DCT, distal convoluted tubule; AR, autosomal recessive inheritance.

*In all of these disorders aldosterone levels are markedly elevated as a result of activation of the renin-angiotensin system due to salt wasting, but in the PHAI syndromes increased mineralocorticoid action is blocked due to loss-of-function mutations of MR and ENaC.
References


References


Section 5 - Reproduction

Chapter 16 - The Physiology and Pathology of the Female Reproductive Axis

Serdar E. Bulun
Eli Y. Adashi

REPRODUCTIVE PHYSIOLOGY

Tightly coordinated functions of the hypothalamus, pituitary gland, ovaries, and endometrium give rise to cyclic, predictable menses that indicate regular ovulation. Regular ovulation also requires normal functioning of other endocrine units such as the thyroid and adrenal glands. For example, patients with hypothyroidism or hyperthyroidism, Cushing's syndrome, or glucocorticoid resistance may present with anovulation. Thus, it is imperative for the clinician to have a thorough knowledge of the functions of the hypothalamus, pituitary, ovaries, and uterus as well as interactions of these tissues with other systems in order to diagnose correctly reproductive disorders and generate treatment strategies.

The most obvious reproductive function of the hypothalamus is the pulsatile secretion of luteinizing hormonereleasing hormone (LHRH). Negative feedback effects of a number of factors including ovarian steroids (Fig. 16-1) regulate hypothalamic LHRH secretion into the portal vessels. Dopamine, norepinephrine, serotonin, and opioids produced in the brain may mediate the regulation of LHRH secretion by ovarian hormones or other stimuli. In response to LHRH, the anterior pituitary cells secrete follicle-stimulating hormone (FSH) and LH. Ovarian steroids and peptides of ovarian origin (e.g., inhibin) modify secretion of FSH and LH (see Fig. 16-1). LH stimulates androstenedione production in theca cells of the ovary, whereas FSH regulates estradiol production in the granulosa cells and follicular growth. Release of an egg from the mature follicle is dependent on a sudden rise in LH levels in midcycle. After ovulation, the follicle transforms into a corpus luteum that secretes both estradiol and progesterone under the control of LH and FSH. Endometrium, the mucosal lining of the uterine cavity, has extremely high levels of nuclear receptors for estrogen and progesterone and is extremely sensitive to these hormones. Estrogen induces the growth of endometrium, whereas progesterone limits this estrogenic effect and enhances differentiation. Sloughing off the functional portion (functionalis) of the endometrium follows withdrawal of estrogen or progesterone. The remaining basal layer (basalis) is capable of full regeneration in response to estrogen.

Ovaries remain quiescent until puberty because the hypothalamus is immature in prepubertal children and FSH and LH do not stimulate the ovaries. The entire reproductive function and most of the endocrine function of the ovaries cease after menopause because ovaries have lost all oocytes and surrounding stromal cells at this time. These prepubertal and postmenopausal states, characterized by the absence of ovarian function, are associated with the lack of menses.

In summary, the female reproductive function from puberty to menopause can be viewed as an extremely delicate ticking clock. The normal function of this apparatus is dependent on coordinate actions of the hypothalamus, pituitary, ovaries, and endometrium. The end result is regular menses every 24 to 35 days. Any disorder of these tissues or disorders of other systems that affect these reproductive units secondarily may result in anovulation and consequent irregular uterine bleeding.

Reproductive Functions of the Hypothalamus

Luteinizing Hormone-Releasing Hormone

LHRH is a 10-amino-acid peptide that is synthesized primarily in specialized neuronal bodies of the arcuate nucleus of the medial basal hypothalamus. Axons from LHRH neurons project to the median eminence and terminate in the capillaries that drain into the portal vessels. The portal vein is a low-flow transport system that descends along the pituitary stalk and connects the hypothalamus to the anterior pituitary. The direction of the blood flow in this hypothyseal portal circulation is from the hypothalamus to the pituitary. Thus, LHRH originating in the neurons of the arcuate nucleus is secreted at the median eminence into the portal circulation, which delivers this hormone to the anterior pituitary (Fig. 16-2). The mature decapeptide LHRH is derived from the posttranslational processing of a large precursor molecule, pre-pro-LHRH (see Fig. 16-2). This precursor peptide is the product of a gene located in the short arm of chromosome 8. The pre-pro-LHRH consists of 92 amino acids and contains four parts (from the N-terminal to the C-terminal): (1) a 23-amino-acid signal domain, (2) the LHRH decapeptide, (3) a 3-amino-acid proteolytic processing site, and (4) a 56-amino-acid domain called LHRH-associated peptide (GAP). The cleavage products of this precursor, LHRH and GAP, are transported to the nerve terminals and secreted into the portal circulation (see Fig. 16-2). A physiologic role of GAP has not been established.

In humans, LHRH neurons are located primarily in the arcuate nucleus of the medial basal hypothalamus and the preoptic area of the anterior hypothalamus. The population of LHRH-producing neurons is relatively limited and is in the range of 1000 to 3000. The neurons that produce LHRH originate from the olfactory area during embryogenesis. These cells migrate during embryogenesis along cranial nerves connecting the nose and forebrain to the hypothalamus. A neuronal cell-surface glycoprotein that probably mediates cell-to-cell adhesion appears to be an important migratory determinant. Mutation of the gene that encodes this adhesion molecule is associated with X-linked isolated gonadotropin deficiency (Kallmann's syndrome) characterized by anosmia, LHRH deficiency, and hypogonadotropic hypogonadism.
A long-acting depot formulation of an LHRH agonist gives rise eventually to down-regulation of the gonadotropin-gonadal axis within 1 to 3 weeks. The initial create a large reserve pool of gonadotropins. The most prominent agonistic response is observed during the early follicular phase, when the combined effects of LHRH agonist and elevated levels of estradiol receptors.

Position 6 gives rise to metabolic stability, whereas replacement of the C-terminal glycinamide residue by an ethylamide group increases strikingly the affinity for the peptides has been attributed to their high binding affinity to LHRH receptors and reduced susceptibility to enzymatic degradation. An amino acid substitution at bonds between amino acids 5 and 6, 6 and 7, and 9 and 10. Analogues of LHRH with different properties have been synthesized by altering amino acids at these positions

The half-life of LHRH is short (2 to 4 minutes) because it is degraded rapidly by peptidases in the hypothalamus and pituitary gland.

Luteinizing HormoneReleasing Hormone Analogues

Luteinizing HormoneReleasing Hormone Agonists

A number of LHRH agonists were generated by substitution of amino acids at the 6 or 10 position. The increased biologic activity of agonistic peptides has been attributed to their high binding affinity to LHRH receptors and reduced susceptibility to enzymatic degradation. An amino acid substitution at position 6 gives rise to metabolic stability, whereas replacement of the C-terminal glycaminide residue by an ethylamide group increases strikingly the affinity for the receptors.

LHRH agonists are administered subcutaneously, intranasally, or intramuscularly. Thus far, many agonistic and antagonistic LHRH analogues with various biologic effects have been produced. These peptidases cleave the bonds between amino acids 5 and 6, 6 and 7, and 9 and 10. Analogues of LHRH with different properties have been synthesized by altering amino acids at these positions

The variations in LHRH pulse frequency are achieved, at least in part, by gonadal steroid feedback. Estradiol increases LHRH pulse frequency, whereas elevated progesterone levels decrease LHRH pulsatility. Therefore, it is conceivable that increased progesterone levels may cause a decrease in LHRH pulse frequency and thereby lead to the preferential biosynthesis and secretion of FSH that is observed in the late luteal phase.

LHRH pulsatility is also modulated by the action of locally released neurotransmitters. Norepinephrine stimulates LHRH release, whereas dopamine exerts an inhibitory effect. Endorphin and other opioids may also serve to suppress the hypothalamic release of LHRH.

The variations in LHRH pulse frequency are achieved, at least in part, by gonadal steroid feedback. Estradiol increases LHRH pulse frequency, whereas elevated progesterone levels decrease LHRH pulsatility. Therefore, it is conceivable that increased progesterone levels may cause a decrease in LHRH pulse frequency and thereby lead to the preferential biosynthesis and secretion of FSH that is observed in the late luteal phase. LHRH pulsatility is also modulated by the action of locally released neurotransmitters. Norepinephrine stimulates LHRH release, whereas dopamine exerts an inhibitory effect. Endorphin and other opioids may also serve to suppress the hypothalamic release of LHRH. The gonadal steroids modify endogenous opioid activity, and the negative feedback of steroids on gonadotropins appears to be mediated, at least in part, by endogenous opioids. Thus, it was proposed that sex steroids enhance the activity of endogenous opioids that in turn exert an inhibitory effect on LHRH secretion. The negative effect of opioids on LHRH secretion is also clinically explicable because the reduced LHRH secretion is associated with hypothalamic amenorrhea may be mediated by an increase in endogenous opioid inhibitory tone.

The half-life of LHRH is short (2 to 4 minutes) because it is degraded rapidly by peptides in the hypothalamus and pituitary gland. These peptidases cleave the bonds between amino acids 5 and 6, 6 and 7, and 9 and 10. Analogues of LHRH with different properties have been synthesized by altering amino acids at these positions (Table 16-1). Thus far, many agonistic and antagonistic LHRH analogues with various biologic effects have been produced.

A number of LHRH agonists were generated by substitution of amino acids at the 6 or 10 position (see Table 16-1). The increased biologic activity of agonistic peptides has been attributed to their high binding affinity to LHRH receptors and reduced susceptibility to enzymatic degradation. An amino acid substitution at position 6 gives rise to metabolic stability, whereas replacement of the C-terminal glycaminide residue by an ethylamide group increases strikingly the affinity for the receptors. LHRH agonists are administered subcutaneously, intranasally, or intramuscularly. An initial agonistic action (i.e., the flare effect) is associated with an increase in the circulating levels of LH and FSH.

The most prominent agonistic response is observed during the early follicular phase, when the combined effects of LHRH agonist and elevated levels of estradiol create a large reserve pool of gonadotropins. Desensitization and down-regulation of the pituitary produce hypogonadotropic hypogonadism. The administration of a long-acting depot formulation of an LHRH agonist gives rise eventually to down-regulation of the gonadotropin-gonadal axis within 1 to 3 weeks. The initial
<table>
<thead>
<tr>
<th>Name</th>
<th>Structure</th>
<th>Route of Administration</th>
</tr>
</thead>
<tbody>
<tr>
<td>LHRH decapptide</td>
<td>pGlu-His-Trp-Ser-Tyr-Gly-Leu-Arg-Pro-Gly-NH₂</td>
<td>Intravenous pulsatile pump</td>
</tr>
<tr>
<td>LHRH agonists</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leuprolide (Lupron)</td>
<td>pGlu-His-Trp-Ser-Tyr-Leu-Leu-Arg-Pro-NHEt</td>
<td>Subcutaneous (daily or depot)</td>
</tr>
<tr>
<td>Nafarelin (Synarel)</td>
<td>pGlu-His-Trp-Ser-Tyr-2Nal-Leu-Arg-Pro-GlyNH₂</td>
<td>Nasal (daily)</td>
</tr>
<tr>
<td>Goserelin (Zoladex)</td>
<td>pGlu-His-Trp-Ser-Tyr-3Ser(O₂Bu)-Leu-Arg-Pro-AzaglyNH₂</td>
<td>Subcutaneous (depot)</td>
</tr>
<tr>
<td>LHRH antagonists</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cetrorelix (Cetrotide)</td>
<td>NAc₂2Nal-c4CIPhe-C3Pai-Ser-Tyr-cGly-Leu-Arg-Pro-cAlaNH₂</td>
<td>Subcutaneous (daily)</td>
</tr>
<tr>
<td>Gantirelix (Antagon)</td>
<td>NAc₂2Nal-c4CIPhe-charg(El₁)-Leu-Harg(El₂)-Pro-calaNH₂</td>
<td>Subcutaneous (daily)</td>
</tr>
</tbody>
</table>

LHRH, luteinizing hormone-releasing hormone.

down-regulation effect is due to desensitization, whereas the sustained response is due to loss of receptors and the uncoupling of the receptor from its effector system. In addition, LHRH agonists may cause ovarian quiescence by the secretion of biologically inactive gonadotropins.

Thus far, the Food and Drug Administration has approved the use of these agonists for the treatment of LHRH-dependent precocious puberty, endometriosis, and prostate cancer. Another indication is preoperative hematologic improvement of patients with anemia caused by uterine leiomyomas. Off-label indications of LHRH agonists include the down-regulation of the pituitary during ovulation induction, induction of endometrial atrophy before endometrial ablation surgery, and the prevention of menstrual bleeding in patients with coagulation defects. LHRH agonists have also been used to suppress ovarian steroidogenesis in hirsute patients. \[^{29}\]

The most prominent side effects of long-term use of depot LHRH agonist formulations are caused by estrogen deficiency. Depot LHRH induces a menopause-like state characterized by hot flashes, vaginal dryness, bone resorption, and osteopenia. Osteopenia is reversible in young women if treatment is maintained for no more than 6 months. \[^{30}\] \[^{31}\] Therefore, the risk-benefit ratio must be considered carefully before LHRH agonist treatment is extended for longer durations. Add-back regimens employing low-dose estrogens or progestins, or both, administered along with LHRH agonists have provided a means to overcome these side effects and permit extending the length of agonist therapy. \[^{32}\]

Luteinizing Hormone-Releasing Hormone Antagonists

Multiple amino acid substitutions also permit the synthesis of LHRH antagonists (see Table 16-1). These include modifications of the pyroglutamic and glycine termini at positions 1 and 10 or deletion and substitution of hydrophobic amino acids at positions 2 and 3. \[^{27}\] LHRH antagonists bind to the LHRH receptor and provide competitive inhibition for the naturally synthesized LHRH. LHRH antagonists have the advantage of inducing an immediate decrease in circulating gonadotropin levels with rapid reversal. \[^{33}\] The early antagonists either lacked potency or were associated with undesirable side effects related to histamine release. Newer LHRH antagonists are at various stages of development. \[^{34}\] \[^{35}\] \[^{36}\] \[^{37}\] \[^{38}\] \[^{39}\] \[^{40}\] These new antagonists are slightly more potent in the down-regulation of ovaries and do not cause histamine release (see Table 16-1).
Reproductive Functions of the Anterior Pituitary

Gonadotroph

Gonadotrophs are specialized cell types of the anterior pituitary that synthesize and secrete LH and FSH. These cells constitute 7% to 15% of the total number of anterior pituitary cells and are detected in this location from early fetal life. The majority of the gonadotrophs are capable of synthesizing both LH and FSH. LH and FSH are each composed of two distinct, noncovalently associated protein subunits called and . In the gonadotroph, the subunit genes are transcribed into messenger ribonucleic acids (mRNAs), which are in turn translated into the subunit precursors. Gonadotrophs contain cell-surface LH-RH receptors that mediate the action of LH-RH. These receptors belong to the seven-transmembrane-domain and G protein-coupled receptor family.

Luteinizing Hormone/Releasing Hormone Receptor

LH-RH is transported from the hypothalamus to the pituitary through the portal circulation and binds to the LH-RH receptor on the cell surface of the gonadotroph with high affinity. The number of LH-RH receptors on the cell surface of the gonadotroph is regulated by varying LH-RH pulse frequencies and other hormonal factors. Changes in LH-RH receptor concentration on the surface of the gonadotroph are often correlated with alterations in gonadotroph response to LH-RH. The concentration of LH-RH receptors in the pituitary gland is highest with an LH-RH pulse frequency of 30 minutes. Markedly low concentrations are observed with an LH-RH pulse frequency of 2 hours. Such changes in the levels of LH-RH receptors are correlated with cyclic synthesis and release of LH and FSH during an ovulatory cycle. The difference in the quantity of LH-RH receptors observed between high-frequency and low-frequency LH-RH pulses is twofold to threefold.

The complementary deoxyribonucleic acid (DNA) for human LH-RH receptor has been cloned and characterized. The single-copy human LH-RH receptor gene is approximately 25 kb in size, contains three exons, and is found on the long arm of chromosome 4. LH-RH binds to a specific seven-transmembrane-domain receptor that is coupled to Gq and sequentially activates phospholipase C. Products of phospholipase C activity then mobilize intracellular pools of Ca and provide lipid-derived messenger molecules.

Luteinizing Hormone and Follicle-Stimulating Hormone

Each gonadotropin is a heterodimer. Both LH and FSH are made of two peptide subunits termed and . The subunits and are associated with noncovalent bonds. The

<table>
<thead>
<tr>
<th>Gonadotropins</th>
<th>Location of Subunit Gene</th>
<th>Size of Subunit (amino acids)</th>
<th>Half-Life in Serum</th>
</tr>
</thead>
<tbody>
<tr>
<td>FSH</td>
<td>Chromosome 11p13</td>
<td>117</td>
<td>34 hr</td>
</tr>
<tr>
<td>LH</td>
<td>Chromosome 19q13.3</td>
<td>121</td>
<td>20 min</td>
</tr>
<tr>
<td>hCG</td>
<td>Chromosome 19q13.3</td>
<td>145</td>
<td>24 hr</td>
</tr>
</tbody>
</table>

FSH, Follicle-stimulating hormone; hCG, human chorionic gonadotropin; LH, luteinizing hormone.

subunits of human LH, FSH, thyroid-stimulating hormone (TSH), and human chorionic gonadotropin (hCG) have an identical polypeptide structure. In contrast, the subunit of each hormone has a unique amino acid sequence and confers the specific activity of heterodimer. Each subunit is cysteine-rich and contains multiple disulfide linkages. Each subunit also contains multiple carbohydrate moieties that play important roles in the biologic activity and metabolism of these hormones. The identical subunit contains 92 amino acids. The subunits of human FSH, LH, and hCG contain 117, 121, and 145 amino acids, respectively.

Upon binding of LH-RH to its receptor, the biosynthesis of the gonadotropins proceeds by transcription of the subunit genes, translation of the subunit mRNAs, post-translational modifications of the precursor subunits and subunit folding and combination, mature hormone packaging, and hormone secretion.

The human subunit gene is located on the short arm of chromosome 6. The encoded precursor polypeptide contains a 24-amino-acid leader sequence that is cleaved post-translationally to produce the mature 92-amino-acid subunit.

The human LH and hCG subunit genes are located in chromosome 19q13.3, which contains a cluster of seven subunit-like genes. Five of these sequences are noncoding pseudogenes arranged in groups of tandem and inverted pairs. Only LH and hCG subunit genes give rise to two distinct and functional mRNA species. The LH subunit mRNA encodes a 145-amino-acid precursor protein that is later cleaved to produce a 24-amino-acid leader peptide and a 121-amino-acid biologically active mature peptide. The hCG subunit mRNA also encodes a 145-amino-acid protein. This protein, however, is not processed post-translationally and functions as the biologically active hCG subunit. The amino acid sequences of the human LH and hCG subunits are 82% homologous. These two subunits confer identical biologic activities when associated with the subunit.

A single gene located on the short arm of chromosome 11 encodes FSH subunit, which is 117 amino acids long. Complementary DNA encoding human FSH-, LH-, or hCG-in combination with the complementary DNA of subunit was expressed in mammalian cells in culture. These cells can synthesize these proteins, modify them after translation, glycosylate and combine the subunits, and secret them as intact FSH, LH, or hCG. These recombinant gonadotropins are currently used clinically to stimulate gonadal function.

Regulation of Circulating Levels of Follicle-Stimulating Hormone and Luteinizing Hormone

The molecular mechanisms responsible for formation and combination of the and subunits of FSH and LH are not completely understood. Production rates of both and subunits are regulated at least in part by negative feedback of estrogen, which is mediated by the pulsatile release of LH-RH from the hypothalamus. The pituitary always contains more subunit than subunit mRNA, and readily detectable levels of free subunit are present in serum. The free subunit, on the other hand, is present at relatively low levels in pituitary and is rarely found in serum or urine. Thus, the specific subunit may be the rate-limiting factor in the synthesis of these glycoprotein hormones.

Serum levels of gonadotropins are proportional to their secretion rates and serum half-lives, which are regulated by the amount of carbohydrate residues. The higher the content of carbohydrate residues, especially the sialic acid residues, the lower the rate of metabolism and the higher the serum half-life. The sialic acid content of gonadotropic hormones and other glycoproteins has a marked effect on their rate of clearance and also influences their apparent molecular size. The higher content of sialic acid in FSH compared with LH is responsible for slower clearance of FSH from the circulation; LH has the most rapid clearance rate. The hCG is highly sialylated and has the longest half-life.
Ovary

The ovary is essential for periodic release of oocytes and the production of the steroid hormones estradiol and progesterone. These activities are integrated into the cyclic repetitive process of follicular maturation, ovulation, and formation and regression of the corpus luteum. Thus, the ovary fulfills two major objectives: (1) the generation of a fertilizable ovum and (2) the preparation of the endometrium for implantation through the sequential secretion of estrogen and progesterone. The ovarian follicle comprising the egg and surrounding granulosa and theca cells constitutes the fundamental functional unit of the ovary.

Adult human ovaries are oval bodies with a length of 2 to 5 cm, a width of 1.5 to 3 cm, and a thickness of 0.5 to 1.5 cm. The combined weight of normal ovaries during the reproductive years is 10 to 20 g (average 14 g). The ovaries lie in approximation to the posterior and lateral pelvic wall and are attached to the posterior surface of the broad ligament by the peritoneal fold, termed the mesovarium. Blood vessels, nerves, and lymphatics traverse the mesovarium and enter the ovary at the hilum.

The ovary consists of three structurally distinct regions: (1) an outer cortex containing the surface germinal epithelium and the follicles, (2) a central medulla consisting of stroma, and (3) a hilum around the area of attachment of the ovary to the mesovarium. The functional anatomy of the adult ovary is illustrated in Figure 16-5 (Figure Not Available). The hilum is the point of attachment of the ovary to the mesovarium. It contains nerves, blood vessels, and hilus cells, which have the potential to become active in steroidogenesis or to form androgen-secreting tumors. These cells are similar to the testis producing Leydig cells of the testes. The outermost portion of the cortex, called the tunica albuginea, is covered by a single layer of surface cuboidal epithelium termed the germinal epithelium. The oocytes, enclosed in complexes called follicles, are in the inner part of the cortex, embedded in stromal tissue (see Fig. 16-5) (Figure Not Available). One

Ovarian differentiation and folliculogenesis depend on coordinate expression and interaction of a multitude of genes. Targeted gene disruption or insertion in mice has made it possible to inquire about the function of specific genes in ovarian differentiation and folliculogenesis. Figure 16-6 (Figure Not Available) summarizes the biologic roles of some of these genes. Transgenic mice represent a first step in attempting to understand in vivo the various gene interactions that result in a functional ovary. Indeed, ovarian pathologic conditions in transgenic mice closely resemble disorders observed in mutant human homologues, as exemplified in cases involving the FSH- subunit and FSH receptor. Many mouse models of ovarian pathologic conditions are available. In general, these can be divided into mice with prenatal ovarian failure with disordered gonad formation and diminished number of germ cells or absent germ cells and mice with postnatal ovarian failure as a result of defects at various stages of folliculogenesis (see Fig. 16-6) (Figure Not Available).

These models should lead to the identification of genetic and molecular mechanisms responsible for the development and function of the human ovary.

Ontogeny of the Ovary

The Oocyte

The primordial germ cells are known to originate outside the embryo proper, from the endoderm of the yolk sac. At this site, they can be identified as early as the end of the third week of gestation by alkaline phosphatase staining. Germ cells migrate to cross a remarkably long distance from the yolk sac to the genital ridge by ameboid movements with the aid of pseudopodia. This long route of migration along the dorsal mesentery of the hindgut is interrupted only by the required lateral curving of the coelomic angle at the level of the genital ridge (Fig. 16-7). Some chemotaxis is operational, but the precise cellular mechanisms underlying the guidance of germ cells to the genital ridge remain uncertain. Germ cells appear unable to persist outside the genital ridge, which may thus be viewed as the only region competent to sustain gonadal development. By the same token, germ cells play an indispensable role in the induction of gonadal development. In fact, no functional gonad is to be expected in the absence of germ cells.

On arrival at the genital ridge by the fifth week of gestation, the premeiotic germ cells are referred to as oogonia. During the subsequent 2 weeks of intrauterine life (weeks 5 to 7 of gestation or the "indifferent" stage), the primordial gonadal structure constitutes no more than a bulge on the medial aspect of the urogenital ridge (see Fig. 16-7). This protuberance is created by proliferation of surface (coelomic) germinal epithelium, by growth of the underlying mesenchyme, and by oogonial multiplication. The oogonia total 10,000 by about 6 to 7 weeks of intrauterine life. Because meiosis and oogonial atresia are not operational, the actual number of germ cells is dictated by mitotic division at this time.

It is during this indifferent phase that the gonadal cortex and medulla are first delineated. However, short of cytogenetic evidence, the precise sexual identity of the gonadal ridge cannot be ascertained at this point. Nevertheless, the absence of testicular development beyond 7 weeks of gestation is generally considered presumptive evidence of ovarian formation. Additional clues to the sexual identity of the gonad can be derived from the detection of oogonial meiosis at about 8 weeks of gestation because no comparable process is observed in the testis until puberty. The sexual identity of the gonadal ridge is histologically clear by 16 weeks of
gestation, when the first primordial follicles can be visualized.

By about 8 weeks of intrauterine life, persistent mitosis increases the total number of oogonia to 600,000 (Fig. 16-8). From this point on, the oogonal endowment is subject to three simultaneous ongoing processes: mitosis, meiosis, and oogonal atresia. Stated differently, the onset of oogonal meiosis and oogonal atresia is now superimposed on oogonal mitosis. As a result of the combined impact of these processes, that is, mitosis counterbalanced by meiosis and oogonal atresia, the number of germ cells peaks at 6 to 7 × 10^6 by 20 weeks of gestation (see Fig. 16-8). At this time, two thirds of the total germ cells are intramictic primary oocytes; the remaining third can still be viewed as oogonal. The midgestational peak and the postpeak decline are accounted for, if only in part, by the progressively decreasing rate of oogonal mitosis, a process destined to end entirely by about 7 months of intrauterine life. Equally relevant is the increasing rate of oogonal atresia, which peaks at about month 5 of gestation (see Fig. 16-8). During this period, regulation of the ovarian developmental process is complex and probably involves a diverse group of genes (see Fig. 16-7).

From midgestation onward, relentless and irreversible attrition progressively diminishes the germ cell endowment of the gonad. Ultimately, some 50 years later, this is finally exhausted. For the most part, this is accomplished through follicular atresia rather than oogonal atresia, begins around month 6 of gestation, and continues throughout life (see Fig. 16-8). In contrast, oogonal atresia is destined to end at 7 months of intrauterine life as follicular atresia sets in. Follicular atresia has a profound effect on germ cell endowment, given that only 1 to 2 × 10^5 germ cells are present at birth (see Fig. 16-8). Remarkably, this dramatic depletion of the germ cell mass occurs during a period as short as 20 weeks. No similar rate of depletion occurs earlier or subsequently. Consequently, newborn females enter life still far from realizing reproductive potential, having lost as much as 80% of their germ cell endowment. This decreases further to approximately 300,000 by the onset of puberty. Of these follicles, only 400 to 500 (i.e., less than 1% of the total) ovulate in the course of a reproductive life span.

Between weeks 8 and 13 of fetal life, some of the oogonia depart from the mitotic cycle to enter the prophase of the first meiotic division. This change marks the conversion of these cells to primary oocytes well before actual follicle formation. Meiosis (beginning at about 8 weeks of gestation) provides temporary protection from oogonial atresia, thereby allowing the germ cells to invest themselves with granulosa cells and to form primordial follicles. Accordingly, oogonia that persist beyond the seventh month of gestation and have not entered meiosis are subject to oogonal atresia. Consequently, no oogonia are usually present at birth.

Once formed, the primary oocyte persists in prophase of the first meiotic division until the time of ovulation, when meiosis is resumed and the first polar body is formed and extruded.
Unlike the testicular seminiferous tubule, the ovary does not constitute an immunologically privileged site. Thus, resident ovarian mononuclear phagocytes constitute the outermost layer of the membrana granulosa and abut the basement layer. High intracellular levels of steroidogenic enzymes and LH receptors in follicular cells suggest that these cells account for the majority of steroidogenesis in the follicle. (Fig. 16-12) (Figure Not Available) . Mural granulosa cells constitute the outermost layer of the membrana granulosa and abut the basement layer. High intracellular levels of steroidogenic enzymes and LH receptors in fetal granulosa cells do not have a direct blood supply. The first sign of follicular recruitment is cuboidal differentiation in the spindle-shaped cells inside the basal lamina, which thereafter undergo successive mitotic divisions to form a multilayered granulosa cell zone. The oocyte enlarges and secretes a glycoprotein-containing mucoid substance called the zona pellucida, which surrounds the oocyte and separates the granulosa cells from the oocyte. This structure is a primary follicle. The secondary follicle is formed by further proliferation of granulosa cells and by the final phase of oocyte growth, in which the oocyte reaches 120 μm in diameter, coincident with proliferation of layers of cells immediately outside the basal lamina to constitute the theca. The portion of the theca adjacent to the basal lamina is termed the theca interna. The theca externa is the outermost layer of the follicle. The theca externa acquires an independent blood supply consisting of one or more arterioles that terminate in a capillary bed at the basal lamina. Capillaries do not penetrate the basement membrane, and the granulosa and oocyte remain avascular. The tertiary follicle is characterized by further hypertrophy of the theca and the appearance of a fluid-filled space among the granulosa cells, named the antrum. The fluid in the antrum consists of a plasma transudate and secretory products of granulosa cells, some of which (estradiol) are found there in strikingly higher concentrations than in peripheral blood. The follicle rapidly increases in size under the influence of gonadotropins to form the mature or graafian follicle. In the graafian follicle, the granulosa and oocyte remain encased by the basal lamina and are devoid of direct vascularization. The antral fluid increases in volume, and the ovum, surrounded by an accumulation of granulosa cells (the cumulus oophorus), occupies a polar, eccentric position within the follicle. The mature graafian follicle is ready to release the ovum by the process of ovulation. (Adapted from Erickson GF, Magoffin DA, Dyer CA. The ovarian androgen producing cells: a review of structure-function relations. Endocr Rev 1985; 6:371-379. Copyright © 1985 by The Endocrine Society.)
are present within the ovarian stroma near perifollicular capillaries. Lymphocytes and polymorphonuclear leukocytes, on the other hand, are observed in the follicle and corpus luteum in varying quantities during the follicular development, corpus luteum formation, and follicular atresia.  

The significance of the preceding observations may be that resident ovarian representatives of the white blood cell series constitute potential in situ modulators of ovarian function, acting through direct local secretion of regulatory cytokines.  

Because the flow of information is probably multidirectional, the same cells are probably targeted for steroid and peptidergic input. Moreover, immune cells are endowed with steroidogenic capabilities that could, in their own right, affect steroid economy.  

Follicles

The follicle represents the most important functional unit in the ovary with respect to germ cell development and steroid production. The follicles are embedded in loose connective tissue of the ovarian cortex and can be subdivided into two functional types: nonmaturing (or primordial) and growing. The majority of follicles (90% to 95%) are nonmaturing throughout reproductive life. Recruitment of a primordial follicle initiates dramatic changes in growth, structure, and function. The growing follicles are divided into four stages: primary, secondary, tertiary, and graafian (see Fig. 16-12) (Figure Not Available). The first three stages of growth can occur in the absence of the pituitary and therefore appear to be controlled by intraovarian mechanisms (Fig. 16-14) . The follicle destined to ovulate is recruited in the first few days of the current cycle.  

The early growth of follicles occurs over the span of several preceding menstrual cycles, but the ovulatory follicle is one of a cohort recruited at the time of transition from the previous cycle's luteal phase and the current cycle's follicular phase (see Fig. 16-14). The total time to achieve preovulatory status is approximately 85 days (see Fig. 16-14). The majority of this period of development is FSH-independent. Eventually, this cohort of follicles reaches a stage at which, unless recruited by FSH, the next step is atresia. Thus, a cohort of follicles measuring 2 to 5 mm is continuously available for a response to FSH (Fig. 16-15) . The late luteal increase in FSH is the critical feature in rescuing this cohort of follicles from atresia, eventually allowing a dominant follicle to emerge and pursue a path to ovulation (see Fig. 16-15) . In addition, maintenance of this increase in FSH for a critical duration of time is essential.  

Recruited primordial follicles either develop into dominant, mature graafian follicles destined to ovulate or degenerate as a result of atresia (see Fig. 16-16). The average time for development of a selected follicle to the point of ovulation is 10 to 14 days (see Fig. 16-15). If a follicle is not recruited, it goes through a process called atresia during which the oocyte and granulosa cells within the basal lamina die and are replaced by fibrous tissue. In contrast, the thecal cells outside the basal lamina do not die but dedifferentiate and return to the pool of cells consisting of ovarian interstitial or stromal cells.  

The process of atresia is generally thought to result from lack of the hormones or growth factors that are formed by the mature dominant follicle through intrinsic intraovarian mechanisms. There is general agreement that atresia of follicles is due to apoptosis.  

Apoptosis is an active and regulated process triggered by a cascade of caspase proteases that lead to characteristic fragmentation of DNA and blebbing of membranes.  

Ovulation

There is a dramatic rise in circulating estradiol level as midcycle approaches (see later). This increase in estradiol is followed by a striking LH and to a lesser extent an FSH surge. This triggers the dominant follicle to ovulate. During each menstrual cycle, usually one follicle ovulates and gives rise to a corpus luteum. In the human, either LH or its surrogate hCG is essential to stimulate the rupture of the mature follicle. It was proposed that increased local prostaglandin biosynthesis in the follicle might mediate the ovulatory effect of LH.  

Ovulation consists of rapid follicular enlargement followed by protrusion of the follicle from the surface of the ovarian cortex. This is followed by the rupture of the follicle and extrusion of an egg-cumulus complex into the peritoneal cavity (Fig. 16-16) (Figure Not Available) . Follicular rupture or ovulation occurs predictably 34 to 36 hours from the start of the LH surge. Elevation of a conical "stigma" on the surface of the protruding follicle precedes rupture (see Fig. 16-16) (Figure Not Available) . Rupture of this stigma is accompanied by a gentle rather than explosive expulsion of the ovum and antral fluid. The gonadotropin-dependent production of proteases acting locally on protein substrates in the basal lamina may play an important role in stigma formation and follicular rupture.  

In particular, plasminogen activator levels increase in the follicle before rupture.  

Thus, plasminogen activatormediated conversion of plasminogen to plasin may contribute to the proteolytic digestion of the follicular wall, which is a prerequisite for follicular rupture.  

Corpus Luteum

After ovulation, the dominant follicle reorganizes to become the corpus luteum (Fig. 16-17) (Figure Not Available). After rupture of the follicle, capillaries and fibroblasts from the surrounding stroma proliferate and penetrate the basal lamina. This rapid vascularization of the corpus luteum may be guided by angiogenic factors, some of which are detected in the follicular fluid.  

Vascular endothelial growth factor has been isolated from corpora lutea and has been postulated, along with basic fibroblast growth factor, to be a potential angiogenic agent in corpora lutea.  

Concurrently, the granulosa and theca cells undergo morphologic changes collectively referred to as luteinization. The granulosa cells become granulosa-lutein cells (large cells), and the theca cells are transformed into theca-lutein cells (small cells; see Fig. 16-17 B) (Figure Not Available) . The so-called K cells, scattered throughout the corpus luteum, are believed to be macrophages.  

The corpus luteum is the endocrine gland that serves as the major source of sex steroid hormones secreted by the ovary during the postovulatory phase of the cycle.  

The human corpus luteum secretes as much as 40 mg of progesterone per day during the midluteal phase of the ovarian cycle.  

In view of the small size of the corpus luteum, it is the most active steroidogenic tissue in humans. An important aspect of corpus luteum formation is the penetration of the follicle basement membrane by blood vessels, which provides the granulosalutein cells with low-density lipoprotein (LDL).  

As stated earlier, LDL cholesterol serves as the substrate for corpus luteum progesterone production.  

A key regulator of steroidogenesis in the corpus luteum is LH. In humans, the LH receptor is maintained throughout the functional life span of corpora lutea and not down-regulated during the maternal recognition of pregnancy.  

The ratelimiting step in LH-mediated progesterone formation in luteinized granulosa cells is the entry of cholesterol into the mitochondria, which is regulated by steroidogenic acute regulatory protein (STAR; see later).  

Thus, the availability of LDL cholesterol and the STAR-mediated mitochondrial entry of cholesterol seem to be the two critical factors that account for the production of large amounts of progesterone in the corpus luteum.
The functional life span of the corpus luteum is normally 14 to 2 days. Thereafter, the corpus luteum spontaneously regresses. It is replaced, unless pregnancy occurs, by an avascular scar referred to as the corpus albicans.

Factors that may regulate luteal life span include hormones such as hCG, maintenance of luteal vascularization, and immune cells. There is little doubt about the central role of LH in the maintenance of corpus luteum function. Withdrawal of LH support in a variety of experimental circumstances has almost invariably resulted in luteal regression. In pregnancy, however, the LH surrogate hCG, secreted by the gestational trophoblast, maintains the ability of the corpus luteum to elaborate progesterone; this stimulus helps to maintain the early gestation until the luteoepithelial shift. Accordingly, the corpus luteum doubles in size (compared with the pregnancy size) during the first 6 weeks of gestation (see Fig. 16-17) (Figure Not Available). This increase is due to proliferation of connective tissues and blood vessels, along with hypertrophy of the luteinized granulosa and theca cells. This early hypertrophy is later followed by regression. The corpus luteum at term is only half the size of that during the menstrual cycle.

Hormones such as estrogens and prostaglandins have been suggested as important factors in the promotion of luteal demise. Immune factors may influence luteal life span because corpus luteal regression is associated with a progressive infiltration of lymphocytes and macrophages.

Apoptosis may be the end-point mechanism by which human corpora lutea are deleted. Corpora lutea during the early luteal phase of the menstrual cycle and corpora lutea of early pregnancy show no evidence of apoptotic DNA fragmentation. DNA fragmentation, on the other hand, is detected in midluteal and late luteal corpora. Thus, it is hypothesized that apoptosis is a major mechanism for the demise of the corpus luteum. LH or hCG inhibits apoptosis in the corpus luteum. In the absence of these trophic factors, apoptosis ensues. The remaining corpus luteum is composed of dense connective tissue and is termed the corpus albicans.

Ovarian Follicle-Stimulating Hormone and Luteinizing Hormone Receptors

The FSH receptor is expressed exclusively by granulosa cells. The LH-hCG receptor is expressed primarily by the theca-interstitial cells of all follicles and by granulosa cells of large preovulatory follicles.

Granulosa cells in primary or secondary follicles that are in the early developmental stages before antrum formation (i.e., preantral follicles) primarily bind FSH but not LH. In these preantral follicles, the binding of LH and hCG is confined to theca-interstitial cells. Granulosa cells in more mature tertiary follicles with an antrum appear capable of binding both LH and FSH. Thus, FSH receptors are found in granulosa cells from follicles of all sizes, but LH receptors are found only in granulosa cells of large preovulatory follicles.

The receptors for the glycoprotein hormones have related structures (Fig. 16-18) (Figure Not Available). The receptors belong to the large family of G protein-coupled receptors, whose members all have a transmembrane domain that consists of seven membrane-traversing -helices connected by three extracellular and three intracellular loops (see Fig. 16-18) (Figure Not Available). The glycoprotein hormone receptors form a separate subgroup within this large family by virtue of their large extracellular hormone-binding domain at the N-terminus. FSH binds to the FSH receptor, and LH and hCG both bind to the same LH receptor. Both LH and FSH receptor genes are located on chromosome 2 at the p21 region. The relationship of the glycoprotein hormone receptors to the other G protein-coupled receptors is indicated by their sequence homology in the C-terminal half of the receptor. This domain, encoded by a single, last exon, contains the seven transmembrane segments and the G-proteincoupling domain. The unusually large extracellular domain of the glycoprotein hormone receptors is encoded by the first 9 or 10 exons (Fig. 16-19).

Role of Follicle-Stimulating Hormone in Ovarian Function

As indicated by its name, FSH is the main promoter of follicular maturation. Given that FSH receptors have been exclusively localized to granulosa cells, it is generally presumed that FSH action in the ovary involves the granulosa cells. The ability of FSH to orchestrate follicular growth and differentiation depends on its ability to exert multiple actions concurrently.

Phenotypes of women with mutations that disrupt the function of the FSH+ subunit gene are in good agreement and demonstrate that FSH is necessary for normal follicular development, ovulation, and fertility. Likewise, pubertal development is hampered in the absence of sufficient numbers of later stage follicles with the granulosa cells needed for adequate estrogen production. Treatment of at least one of these patients with exogenous FSH resulted in follicular maturation, ovulation, and normal pregnancy. The presenting phenotype of FSH- subunit deficiency is practically identical to that caused by inactivating mutations of the FSH receptor. Women with FSH receptor mutations are clinically similar to patients with gonadal dysgenesis, with absent or poorly developed secondary sexual characteristics and high serum levels of FSH and LH. The notable difference was the presence of follicular cells in cases with FSH receptor mutation, consistent with the FSH independence of primordial follicle recruitment and early follicular growth and development. In contrast, total absence of all follicles, including those in the primordial stage, was observed in the cases in which the FSH receptor mutation could not be detected. Thus, the ovarian phenotype of FSH receptor deficiency is distinct from the common form of gonadal dysgenesis as found in Turner's syndrome with streak gonads and absence of growing follicles.

In vivo rodent studies suggest that FSH is capable of increasing the number of its own receptors in the granulosa cell. Whereas estradiol by itself may be without effect on the distribution, number, or affinity of granulosa cell FSH receptors, estradiol has been shown to synergize with FSH to enhance the overall number of granulosa cell FSH receptors. Consequently, changes in the production of estradiol by preantral follicles could increase their response to FSH through the regulation of granulosa cell-surface FSH receptors. This interaction between FSH and estradiol in follicular development has been well established in rodents. It appears that estradiol enhances the activity of ER and may mediate the estrogenic effect on ovarian follicular development and growth. On the other hand, it is not clear at this time whether a similar relationship exists in the human ovary. ER is not detected in the human ovary in significant quantities. The demonstration of ER in the human ovary, however, suggests an interaction between FSH and estrogen in the regulation of normal follicle development and ovulation in women.

One of the major actions of FSH is the induction of granulosa cell aromatase activity. Thus, little or no estrogen can be produced by FSH-unprimed granulosa cells even if they are supplied with aromatizable androgen precursors. On the other hand, treatment with FSH enhances the aromatization capability of granulosa cells, an effect related to enhancement of the granulosa cell aromatase content.

Treatment with FSH has also been shown to induce LH receptors in granulosa cells. The ability of FSH to induce LH receptors is augmented by the concomitant presence of estradiol. Furthermore, progestins, androgens, and LH itself may also induce LH receptors. Once induced, the granulosa cell LH receptor requires the continued presence of FSH for its maintenance.

Circumstantial evidence, as deduced from studies of women with disrupting mutations of the genes that encode FSH and LH receptors and aromatase P450 (P450 arom), indicates that FSH action, but not estrogen or LH action, is essential for follicular growth in humans. In women with deficient LH action or estrogen...
biosynthesis, follicular growth and development up to the antral stage were observed, although these individuals were anovulatory. On the other hand, women with mutations of the FSH- subunit or FSH receptor had only primordial follicles in their ovaries. These data indicate that estrogen or LH is not critical for follicular development at least until the tertiary stage (see Fig. 16-12 (Figure Not Available) and Fig. 16-14). It should be kept in mind, however, that FSH by itself is not sufficient to achieve normal follicular development and ovulation.

Role of Luteinizing Hormone in Ovarian Function

LH is essential for ovulation (follicular rupture) and the sustenance of corpus luteum function. In addition, LH plays other important roles in follicular function. First, it is likely that LH plays a major role in the promotion of theca-interstitial cell androgen production. Moreover, LH may well synergize with FSH in the more advanced phases of follicular development. Last, small and sustained increments in the circulating levels of LH are both necessary and sufficient to cause small antral follicles to grow and develop to the preovulatory stage.

It is presumed that LH acts on theca-interstitial cells of small follicles, where it promotes the biosynthesis of C₁₈ steroids. The consequent increase in estrogen production is presumed to contribute to the growth and development of the follicles. Treatment with small doses of LH also presumably results in an increase in LH receptor content as well as in induction of the key steroidogenic proteins such as STAR, side-chain cleavage P450 (P450sc), 3-hydroxysteroid dehydrogenase type II (3-HSD-II), and 17 hydroxylase (P450ør).

The role of LH action in human ovarian physiology was exemplified by the phenotype of a woman with a disrupting mutation of the LH receptor gene. She presented with amenorrhea with normally developed secondary sexual characteristics, increased circulating FSH and LH levels, and low levels of estradiol and progesterone that were unresponsive to HCG treatment. The ovary contained follicles that developed up to antral stage with a well-developed theca layer but no preovulatory follicles or corpora lutea. These observations collectively support the view that LH is essential for ovulation and sufficient estrogen production, whereas follicular development is initially autonomous and at later stages dependent on intact FSH action.

Ovarian Steroidogenesis

The preovulatory follicle secretes estradiol during the first half of the menstrual cycle, whereas the corpus luteum secretes both estradiol and progesterone during the second half of the cycle (Fig. 16-20). The production of these two biologically active steroids is orchestrated in the follicle and corpus luteum in a cell-specific manner under the control of LH and FSH.

The steroid hormone contents of the ovarian vein effluents and peripheral venous blood were compared to distinguish steroids secreted by the ovary from those secreted by the adrenal and from those produced by peripheral conversion of precursors. These studies revealed that the ovaries secrete pregnenolone, progesterone, 17-hydroxyprogesterone, dehydroepiandrosterone (DHEA), androstenedione, testosterone, estrone, and estradiol. Although such measurements provide insights into the steroidogenic pathways under study, they do not identify the specific ovarian cells involved.

Studies using microdissected preovulatory follicles identified estrone and estradiol as the major steroid products (see Fig. 16-20). Progesterone and 17-hydroxyprogesterone, on the other hand, proved to be the major products of the corpus luteum (see Fig. 16-20).

The biologically active ovarian steroids are estradiol and progesterone (Fig. 16-21). The major C₁₈ -steroid product of the ovary, androstenedione, is not biologically active. Androstenedione, however, acts as a dual precursor and contributes to circulating levels of estrone and testosterone through conversion in extraglandular tissues such as adipose tissue and skin (see later). It is likely that aromatization of weak estrone is further converted to the potent estrogen estradiol, and testosterone is converted to the much more potent androgen dihydrotestosterone (DHT) locally in target tissues such as brain, breast, prostate, and genital skin in order to exert potent biologic effects. The presence of multiple proteins with overlapping enzymatic activities (i.e., 17-HSD and 5-reductase) that catalyze these conversions in a large number of human tissues is supportive of this idea.

The general steroidogenic pathway for the production of estrogens and androgens is depicted in Figure 16-21. There are three major categories of ovarian steroids, as follows.

C₁₈ Steroids

The naturally occurring estrogens are C₁₈ -steroids characterized by the presence of an aromatic A ring, a phenolic hydroxy group at C-3, and either a hydroxy group (estradiol) or a ketone group (estrone) at C-17. Aromatase is the key enzyme for estrogen production in the ovary (see Fig. 16-21). The protein P450ør confers the specific activity of the aromatase enzyme complex. P450ør expression in the ovarian granulosa cell is regulated primarily by FSH. The principal and most potent estrogen secreted by the ovary is estradiol. Although estrone is also secreted by the ovary, another important source of estrone is extraglandular conversion of androstenedione in peripheral tissues. Estradiol (16-hydroxyestradiol) is the most abundant estrogen in urine and is produced by the metabolism of estrone and estradiol in extravascular tissues. All C₁₈ -steroids including estrone, estradiol, and estradiol are commonly referred to as estrogens. It should be pointed out, however, that estrone and estradiol are only weakly estrogenic and must be converted
into this cell type through primarily low-density lipoprotein receptors and for secretion of large amounts of progesterone into the circulation. The entry of cholesterol into mitochondria (by StAR) is likely to be the most critical steroidogenic step for progesterone formation in granulosa lutein cells. Androstenedione produced in theca-lutein cells serves as a substrate for estradiol produced in granulosa-lutein cells. Granulosa cells and the transcription factor SF-1 play key roles in important steroidogenic steps in both cell types. ATP, adenosine triphosphate; cAMP, cyclic adenosine monophosphate; FSH-R, follicle-stimulating hormone receptor; HSD, 3-hydroxysteroid dehydrogenase; LH-R, luteinizing hormone receptor.

to estradiol to show full estrogenic action. There are at least seven enzymes in the 17-HSD family with overlapping activities, which are capable of converting estrone to estradiol in the ovary and extravascular tissues. ['3]

Catechol estrogens are formed by hydroxylation of estrogens at the C-2 or C-4 position. The physiologic role of catechol estrogen, if any, is unclear. Low body weight and hypometabolism are associated with increased formation of catechol estrogens. ['3] Estrone sulfate, formed by peripheral conversion of estradiol and estrone, is the most abundant estrogen in blood but is not physiologically active. ['5] Estradiol sulfate is presumed to serve as a reservoir for estrone formation in a number of tissues, including those that are targets of estrogen. ['6] Estradiol regulates gonadotropin secretion and promotes development of the secondary sexual characteristics of women, uterine growth, thickening of the vaginal mucosa, thinning of the cervical mucus, and linear growth of the dactal region of the breast.

![Figure 16-21](https://example.com/figure1621.png)

**Figure 16-21** Steroidogenic pathway in the ovary. Biologically active steroids progesterone and estradiol are produced primarily in the ovary of a woman of reproductive age. Estradiol production requires the activity of six steroidogenic proteins including StAR and six enzymatic steps. P450c17, product of the CYP17 gene, catalyzes two enzymic reactions. The four rings of the cholesterol molecule and its derivative steroids are identified by the first four letters in the alphabet, and the carbons are numbered in the sequence shown in the insert. 3-HSD-II, 3-hydroxysteroid dehydrogenase [9] 5 isomerase type II; 17-HSD-1, 17-hydroxysteroid dehydrogenase type 1; P450 arom, aromatase; P450c17, 17-hydroxylase/17,20-lyase; StAR, steroidogenic acute regulatory protein.

C_{17}-Steroids

The principal progestogens are C_{19}-steroids and include pregnenolone, progesterone, and 17-hydroxyprogesterone (see Fig. 16-21). Pregnenolone is of primary importance in the ovary because of its key position as precursor of all steroid hormones. Progesterone is the principal secretory product of the corpus luteum and is responsible for the progesterational effects (i.e., cell differentiation and induction of secretory activity in the endometrium of the estrogen-preened uterus). Progesterone is required for implantation of the fertilized ovum and maintenance of pregnancy. It also induces decidualization of the endometrium, inhibits uterine contractions, increases the viscosity of cervical mucus, promotes labor (anterior) development of the breast glands, and increases basal body temperature. On the other hand, 17-hydroxyprogesterone, also secreted by the corpus luteum, has little, if any, biologic activity. ['5]

C_{19}-Steroids

The ovary secretes a variety of C_{19}-steroids, including DHEA, androstenedione, and testosterone (see Fig. 16-21). They are produced by the thecal cells and to a lesser degree by the ovarian stroma. The major C_{19}-sterol is androstenedione, part of which is secreted directly into plasma, with the remainder converted to estrone by the granulosa cells. Androstenedione can be converted to estradiol or testosterone in the ovary and in extravascular tissues. Only testosterone and DHT but not androstenedione are true androgens with the capacity of interacting with the androgen receptor (see later).

Steroids formed by the ovary, as well as other steroid-producing organs, are derived from cholesterol (see Fig. 16-21). There are several sources of cholesterol that can provide the ovary with substrate for steroidogenesis. These include (1) plasma lipoprotein cholesterol, (2) cholesterol synthesized de novo within the ovary, and (3) cholesterol from intracellular stores of cholesterol esters within lipid droplets. In the human ovary, LDL cholesterol is an important source of cholesterol utilized for steroidogenesis. ['1] LH stimulates the activity of adenylyl cyclase, increasing production of cyclic adenosine monophosphate (cAMP), which serves as a second messenger to increase LDL receptor mRNA and the binding and uptake of LDL cholesterol as well as the formation of cholesterol esters. ['4] LDL-derived cholesterol is particularly essential for normal levels of progesterone production in the granulosa lutein cells of the corpus luteum (see Fig. 16-20b). ['5]

The first and rate-limiting step in the synthesis of all ovarian steroid hormones is the movement of cholesterol into the mitochondrion, which is regulated by StAR (see Fig. 16-21). This movement is followed by conversion of cholesterol to pregnenolone, catalyzed by the mitochondrial enzyme complex consisting of P450_{c11}, adrenodoxin, and flavoprotein. ['5] LH induces steroidogenesis by increasing intracellular cAMP, which increases the conversion of cholesterol to pregnenolone in two distinct ways: (1) acute regulation, over minutes, occurs through the phosphorylation of preexisting StAR and rapid synthesis of new StAR protein, and (2) chronic stimulation, within hours to days, occurs through the induction of P450_{c17} expression and consequent increased steroidogenesis (see Fig. 16-20a). Thus, StAR increases the flow of cholesterol to mitochondria, thus regulating substrate availability to whatever amount of P450_{c17} is available on the inner mitochondrial membrane. ['5] In the absence of StAR, only 14% of the maximal StAR-induced level of steroidogenesis persists as StAR-independent steroidogenesis. ['5]

StAR expression in the preovulatory graafian follicle is limited primarily to the thecal cells (see Fig. 16-20a) ['5] The most important product of the thecal cell during the follicular phase is the estrogen precursor androstenedione, whose production is controlled primarily by StAR (see Fig. 16-20b). The biologically active steroid product of the ovary during the follicular phase is estradiol that arises from the granulosa cells located adjacent to theca cells (see Fig. 16-20a). The rate-limiting step for granulosa cell estradiol production is regulated by the FSH-dependent activity of the aromatase enzyme in a cyclic fashion (see Fig. 16-20). During the luteal phase, cells of the corpus luteum, including granulosa lutein cells, also show intense StAR immunoreactivity with a patchy distribution (see Fig. 16-20a). The delivery of cholesterol to the mitochondrial side-chain cleavage enzyme system in the corpus luteum is the rate-limiting step for progesterone biosynthesis and is regulated by StAR (see Fig. 16-20a). Thus, estradiol production seems to be regulated primarily by StAR and P450_{c17} whereas progesterone biosynthesis is under the control of StAR.

The ovarian granulosa, theca, and corpus luteum cells possess StAR plus five distinct proteins with specific enzyme activities for steroid hormone formation. These steroidogenic enzymes are P450_{c17}, 3-HSD-II, P450_{arom}, P450_{c17}, and 17-HSD-1. ['3] ['5] These enzymes are responsible for the conversion of cholesterol to the two major biologically active products estradiol and progesterone. ['1] ['5] Steroidogenesis dependent on LH and FSH in both theca and granulosa cells is mediated by common signaling molecules including cAMP and the transcription factor steroidogenic factor-1 (SF-1) (see Fig. 16-20). ['3] ['5] SF-1 regulates the expression of genes that encode StAR, P450_{c17}, 3-HSD-II, P450_{arom}, and P450_{c17} (see Fig. 16-20). Thus, SF-1 can be regarded as a downstream master switch that orchestrates ovarian steroidogenesis. ['5]

Summary of Updated Two-Cell Theory for Ovarian Steroidogenesis

The classical two-cell theory is supported by molecular findings in the following fashion. Ovarian steroidogenesis in the preovulatory follicle takes place through LH receptors on theca and FSH (possibly plus LH) receptors on granulosa cells (see Fig. 16-20). Cyclic AMP production and increased SF-1 binding to multiple steroidogenic promoters mediate LH action in theca cells. In particular, SF-1 is the primary regulator of production of androstenedione that subsequently diffuses into granulosa cells to serve as the estrogen precursor. In the preovulatory follicle, cholesterol in theca cells arises from circulating lipoproteins and de novo biosynthesis. FSH is responsible for follicular growth and also estrogen formation. FSH induces cAMP formation, increased SF-1 binding activity, and P450_{c17} expression in preovulatory granulosa cells to give rise to estradiol formation primarily through aromatization of androstenedione (see Fig. 16-20).
In the corpus luteum, large deposits of cholesterol (i.e., the yellow color) arise primarily from circulating lipoproteins to support production of extremely high quantities of progesterone. Other key anatomic events in formation of the corpus luteum are the disruption of the basement membrane between the granulosa and theca and strikingly increased vascularization of granulosa lutein cells (see Fig. 16-17) (Figure Not Available). Theca lutein cells possess LH receptors and produce androstenedione. Cyclic AMP, SF-1, and SFAR induced by LH remain as the key regulators for biosynthesis of thecal androstenedione, which serves as the estrogen precursor in neighboring granulosa lutein cells.

The granulosa lutein cell of the corpus luteum is both anatomically and functionally different from its counterpart in the preovulatory follicle. First, these cells are larger and heavily vascularized and contain large quantities of cholesterol. Second, granulosa lutein cells contain high levels of LH receptors in addition to FSH receptors. Third, they produce large quantities of progesterone that is regulated primarily by LH and SFAR. Granulosa lutein cells also aromatize androstenedione of theca origin and eventually give rise to estradiol formation through FSH action and P450 arom. The common known mediators of LH and FSH in granulosa lutein cells are cAMP and increased SF-1 binding activity. Specific functions of these two gonadotropins (i.e., differentiation, growth, and progesterone formation versus estradiol formation) are probably determined by as yet unidentified modifying factors.

---

**Peptide Hormones Produced by Ovary**

The ovary produces a large number of peptides that can act in an intracrine, autocrine, paracrine, or endocrine fashion. These include numerous growth factors (e.g., insulin-like growth factors [IGFs]) and cytokines (e.g., interleukin-1).

The group of peptides including inhibin, activin, and follistatin are produced in ovarian granulosa cells under the control of FSH (see Fig. 16-1). The production of these factors is not limited to the ovary. A number of other tissues, including the adrenal, pituitary, and placenta, synthesize these peptides. It has been well established that inhibin plays a major regulatory role in FSH production in the pituitary. Inhibin is a 32-kd glycoprotein composed of two subunits, (18 kd) and (12 kd), linked by disulfide bonds. Structurally, inhibin is a heterodimer composed of a common subunit but different subunits, denoted α and β. The forms of inhibin, α and β, are termed A and B, respectively. Although inhibin is produced by a number of tissues in the body, the majority is derived from the gonads. In the ovary, the source of inhibin is granulosa cells. The main role of inhibin, for which it was discovered and named, is to suppress FSH production in the pituitary (see Fig. 16-1).

Although both isoforms of inhibin seem to have similar biologic properties, their synthesis is regulated differently during the follicular and luteal phases. Inhibin B is secreted mainly during the early follicular phase, with levels decreasing in midfollicular phase and becoming undetectable after the LH surge. Inhibin A levels are low during the first half of the follicular phase but increase gradually during midfollicular phase with a peak during the luteal phase. All three subunits are detected in small antral follicles by immunohistochemistry and in situ hybridization. The α and β subunits are found in the dominant follicle and in the corpus luteum. All three subunits are expressed in response to gonadotropins or factors that increase intracellular cAMP. The mechanisms underlying the regulation of the different parts of the menstrual cycle by inhibin isoforms need further study for a fuller understanding.

Activin is structurally related to inhibin but may exert opposite actions. Activin contains two subunits that are identical to the subunits of inhibins A and B. Thus, the three activin isoforms are activin A(αA, βA), activin B(αB, βB), and activin AB(αA, βB). In the pituitary, activin stimulates the release of FSH. In the ovarian follicle, activin enhances FSH action.

**Overview of the Hormonal Changes during the Ovarian Cycle**

FSH secretion is suppressed by negative feedback of the ovarian hormones estrogen, inhibin, and progesterone during the early and midluteal phase. Upon the regression of the corpus luteum during the late luteal phase, however, the sharp decline in estrogen, inhibin, and progesterone abolishes this negative feedback of FSH just before and during menses. This initial increase in FSH is essential for follicle recruitment and growth and steroidogenesis (see Fig. 16-22). With continued growth of the follicle, follicular paracrine and autocrine factors produced within the follicle maintain follicular sensitivity to FSH. Continuing and combined action of FSH and activin leads to the appearance of LH receptors on the granulosa cells, a prerequisite for ovulation and luteinization.

Ovulation is triggered by the rapid rise in circulating levels of estradiol. A positive feedback response at the level of the anterior pituitary and possibly at the hypothalamus results in the midcycle surge of LH necessary for expulsion of the egg and formation of the corpus luteum (see Fig. 16-22). A rise in progesterone follows ovulation along with a second rise in estradiol, producing the 14-day-long luteal phase characterized by low FSH and LH levels. The demise of the corpus luteum concomitant with a fall in hormone (progesterone, estradiol, and inhibin A) levels allows FSH to increase again toward the end of the luteal phase, thus initiating a new cycle (see Fig. 16-22). If pregnancy is established by the implantation of a blastocyst, however, the structural integrity and function (progesterone and estradiol production) of the corpus luteum are maintained by HCG that is secreted from the trophoblast. The HCG acts as a surrogate for LH on the corpus luteum.

In addition to FSH and LH, local factors (e.g., activin and inhibin) regulate follicular development and steroidogenesis. In the early follicular phase, activin produced by granulosa in immature follicles enhances the action of FSH on aromatase activity and FSH and LH receptor formation while simultaneously suppressing C19 -steroidogenesis in theca cells. In the late follicular phase, increased production of inhibin by the granulosa and decreased activin promote the synthesis of C19- steroids in the theca in response to LH and local growth factors and cytokines to provide larger amounts of the precursor androstenedione for the production of estrone and ultimately estradiol in the granulosa cells (see Fig. 16-20). Both LH-mediated androstenedione production in the theca cells and FSH-mediated estradiol production in granulosa cells are potentiated by IGFs. The major endogenous IGF produced in the human ovarian follicle is IGF-II (versus IGF-I) in both granulosa and theca cells. The actions of both IGF-I and IGF-II are mediated by IGF receptor type 1 in both cells. IGF receptor type 1 is structurally similar to the insulin receptor. Thus, it appears that gonadotropin-related IGF action in the ovary is regulated primarily by IGF-II and IGF receptor type 1.

In summary, ovulation is under the control of substances functioning as classical hormones (FSH, LH, estradiol, and inhibin) transmitting messages between the ovary and the hypothalamic-pituitary axis and paracrine and autocrine factors such as IGF-II, inhibin, and activin, which coordinate sequential activities within the follicle destined to ovulate. The negative feedback relationship between corpus luteum products (estradiol, progesterone, and inhibin) and FSH results in the critical initial rise in FSH immediately before and during menses, and the positive feedback relationship between estradiol and LH is responsible for the ovulatory stimulus (see Fig. 16-22). Within the ovary, IGF-I, inhibin, and activin modify follicular responses necessary for growth and function. These endocrine, paracrine, and autocrine factors undoubtedly represent only a portion of a complete picture. The causes of anovulation are diverse and may be related to defects in cell-surface receptors, intracellular elements of signal transduction, or cell-cell interactions.

**Extravarian Steroidogenesis**

 Estradiol formation takes place in a number of tissues in the woman of reproductive age. These tissues may be placed in three categories: (1) the ovary, (2) peripheral tissues such as subcutaneous fat and skin, and (3) physiologic and pathologic target sites such as the hypothalamus, breast cancer cells, and the cells of endometriosis (Fig. 16-22). The latter two sources of estrogen are particularly critical in anovulatory premenopausal and postmenopausal women. Although small
quantities of estrogen are produced by an individual adipose or skin fibroblast in a continuous fashion, these cell types contribute to circulating estradiol levels because of their relative abundance.

Figure 16-23  Estrogen biosynthesis in women. The biologically active estrogen estradiol \( (E_2) \) is produced in at least three major sites: (1) by direct secretion from the ovary in reproductive-age women; (2) by conversion of circulating androstenedione \( (A) \) of adrenal or ovarian origins, or both, to estrone \( (E_1) \) in peripheral tissues; and (3) by conversion of \( A \) to \( E_1 \) in estrogen target tissues. In the latter two instances, estrogenically weak \( E_1 \) is further converted to \( E_2 \) within the same tissue. The presence of the enzyme aromatase and 17-hydroxysteroid dehydrogenase \( (17\text{-HSD}) \) is critical for \( E_2 \) formation at these sites. \( E_2 \) formation by peripheral and local conversion is particularly important in post-menopausal women and in estrogen-dependent diseases such as breast cancer, endometriosis, and endometrial cancer. (see Fig. 16-23).

This effect is more pronounced in obese women because of increased mass of the adipose tissue and skin.

P450arom in adipose and skin fibroblasts is responsible for peripheral aromatization of androstenedione that arises from both the ovary and adrenal in premenopausal women and primarily from the adrenal in postmenopausal women (see Fig. 16-23). The product of this reaction, estrone, is only weakly estrogenic, however. Estrone is further converted to estrone sulfate, which serves as a reservoir for estrone in blood and other tissues. Estrone (arising from androstenedione and estrone sulfate) is further converted to the biologically active estradiol in target tissues such as the endometrium and breast by a number of enzymatic proteins with overlapping reductive 17-HSD activity (see Fig. 16-23).

It is likely that local P450arom expression in hypothalamus is critical for the regulation of gonadotropin secretion. Finally, estrogen-dependent pathologic tissues such as those in breast cancer and endometriosis contain extremely high levels of P450arom that enhances tissue growth by increasing local estradiol concentrations. Circulating androstenedione is the major substrate for aromatase activity in these physiologic and pathologic target tissues.

Significant quantities of circulating androstenedione can also be converted to testosterone in peripheral tissues (see later). This is probably accomplished by the presence of multiple 17-HSDs with overlapping reductive activities in peripheral tissues. Androgenic action of testosterone is strikingly amplified by its conversion to DHT in peripheral and target tissues (e.g., skin and prostate). At least two distinct proteins encoded by two separate genes, 5-reductase type 1 and type 2, catalyze the conversion of testosterone to DHT in the liver, prostate, and skin. Local production of DHT in genital skin fibroblasts is critical for normal masculinization of external genitalia of male fetuses in utero. DHT formation in the skin is also important in the etiology of hirsutism (see later).
Endometrium

The endometrium is the mucosal lining of the uterine cavity. The decidua is the highly modified and specialized endometrium of pregnancy. From the evolutionary perspective, the human endometrium is highly developed to accommodate the hemochorial endothelial type of placentation, which requires the presence of spiral arteries (Fig. 16-24). Trophoblasts of the blastocyst invade spiral arteries during implantation and placentation in the establishment of utero-placental vessels.

Spiral arteries of the human endometrium also confer another unique process termed menstruation. Menstruation is shedding of endometrial tissue with hemorrhage that is dependent upon sex steroid hormone-directed changes in blood flow in the spiral arteries. The presence of spiral arteries is essential for menstruation because only the human and a few other primates that have endometrial spiral arteries experience menstruation. With nonfertile but ovulatory ovarian cycles, menstruation affects desquamation of the endometrium. New endometrial growth and development must be initiated with each ovarian cycle, so that endometrial maturation corresponds rather precisely with the next opportunity for pregnancy. There seems to be a narrow window of endometrial receptivity to blastocyst implantation that corresponds to the period between days 20 and 24 during a 28-day menstrual cycle.

Functional Anatomy of the Endometrium

The endometrium can be divided morphologically into an upper two-thirds functionalis layer and a lower one-third basalis layer (see Fig. 16-24). The purpose of the functionalis layer is to prepare for the implantation of the blastocyst; therefore, it is the site of proliferation, secretion, and degeneration. The purpose of the basalis layer is to provide the regenerative endometrium following menstrual loss of the functionalis. Major histologic components of the endometrium include (1) stromal cells that constitute the skeleton of the tissue, (2) a single layer of epithelial cells that line the lumens of the endometrial cavity and invaginations of the stroma, (3) blood vessels, and (4) resident immune cells. The epithelial cells that line the rather deep invaginations of the stroma are also referred to as glandular cells. It should be noted, however, that these deep crypts represent simply extensions of the intracavitary lumen and are not true glands. These invaginations lined by epithelial cells extend from the surface of the functionalis layer (i.e., luminal epithelium) deep into the basalis level (the so-called glandular epithelium). Thus, after the functionalis layer is shed at the time of menstruation, the basalis that contains both epithelial and stromal cells can give rise to a new functionalis layer for the upcoming cycle (Fig. 16-25).

The cellular components of the functionalis layer undergo a striking progression during the menstrual cycle, whereas the basalis shows only modest alterations. The sequence of endometrial changes associated with an ovulatory cycle has been carefully studied by Noyes and colleagues in the human and Markee and Bartelmez in the subhuman primate. The histologic changes that occur in the endometrium during the nonfertile but ovulatory menstrual cycle are summarized in Figure 16-26 as described originally by Noyes and co-workers. As described originally by Noyes and colleagues.

Hormone-Induced Morphologic Changes of the Endometrium

The cyclic changes in endometrial histology are faithfully reproduced during each ovulatory ovarian cycle. These sex steroid hormone-induced modifications can be summarized as follows. (1) During the preovulatory, or follicular, phase of the cycle, estradiol is secreted (principally by a single dominant follicle of one ovary) in increasing quantities until just before ovulation. (2) During the postovulatory, or luteal, phase of the cycle, progesterone is secreted by the corpus luteum in increasing amounts (up to 40 to 50 mg/day) until the midluteal phase. (3) Beginning about 7 to 8 days after ovulation, the rates of progesterone (and estrogen) secretion by the corpus luteum begin to decline and then diminish progressively before menstruation (see Fig. 16-22).

In response to these cyclic changes in the rates of ovarian sex steroid hormone secretion, there are five main stages of the corresponding endometrial cycle: (1) menstrual-postmenstrual reapithelialization, (2) endometrial proliferation in response to stimulation by estradiol; (3) abundant epithelial secretion, in response to the combined action of estradiol and progesterone; (4) premenstrual ischemia, the result of endometrial tissue volume involution, which causes stasis of blood in the spiral arteries; and (5) menstruation, which is preceded and accompanied by severe vasconstriction of the spiral arteries and collapse and desquamation of all but the deepest layer of the endometrium. In the final analysis, menstruation is the consequence of the withdrawal of factors that maintain endometrial growth and differentiation. Commonly, the initiation of menstruation is attributed to progesterone withdrawal. This concept was developed because the administration of estrogen to
postmenopausal women and thence treatment or withdrawal with a progestin affect menstruation, even with continued estrogen treatment. Moreover, progesterone facilitates and permits decidualization of the endometrium and the maintenance of pregnancy, whereas progesterone withdrawal favors the initiation of menstruation, lactation, and parturition. There are probably multiple additional coordinated and interactive processes (other than progesterone withdrawal) that are operative and essential for the success of each of these events.

Both the preovulatory (follicular or proliferative) phase and the postovulatory (luteal or secretory) phase of the ovarian-endometrial cycles are customarily divided into early and late stages (see Fig. 16-22). The normal secretory phase of the endometrial (menstrual) cycle can be subdivided rather finely (almost day by day), by histologic criteria, from shortly after ovulation until the onset of menstruation. In fact, Noyes and other investigators have provided an extremely detailed description of the histologic features of the secretory phase endometrium, which permit accurate dating during the luteal phase (see Fig. 16-26). Gynecologists use the histologic dating of the endometrial biopsies obtained during the luteal phase to evaluate ovulation, progesterone production, or the degree of biologic response of the endometrium to progesterone. Normal endometrial development is assumed when the histologic and chronologic endometrial dating agree within 2 days. When they differ by more than 2 days, the endometrium is considered to be out of phase. Out-of-phase endometrial tissue was proposed to be a cause of implantation failure giving rise to infertility. Understanding the limitations of this test is important, if infertility treatments are based on biopsy results, because the sensitivity and the specificity of the dating of endometrial biopsy for the evaluation of infertility are unknown. For

**Effects of Ovarian Steroids on Endometrium**

Estradiol or synthetic estrogens cause a striking thickening of endometrial tissue. Both stromal and epithelial cells of the endometrium proliferate rapidly under the influence of estradiol. Estrogen increases mitotic activity and DNA synthesis in both cell types strikingly (Fig. 16-27). While promoting growth, estrogen also renders endometrial tissue responsive to progesterone by inducing the expression of progesterone receptors (PRs) in this tissue because progesterone action is dependent on previous or concurrent estrogen exposure of the endometrium.

In contrast to the proliferative effects of estrogen, progesterone action primarily gives rise to the differentiation of the endometrium. For example, progesterone can inhibit and even reverse the proliferative action of estrogen on the functionalis layer (Fig. 16-28). Moreover, progesterone action prepares the endometrium for implantation of the embryo through differentiation of both epithelial and stromal cells. Progesterone induces the production and secretion of a glycoprotein-rich substance from the epithelial cells. Progesterone also causes an increase in the stromal cell cytostasis, a process called pseudodecidualization.

**Estrogen Action**

Estrogens, the biologically potent, naturally occurring estrogens, which is secreted by the granulosa cells of the dominant ovarian follicle, acts to promote responses of the endometrium in a manner that is classical for steroid hormone action. Estradiol enters cells from blood by simple diffusion, but in estrogen-responsive cells, binding to the ER sequesters estradiol. ERs are proteins with high affinity for estradiol and other biologically active estrogens, that is, synthetic estrogens. The ER subtypes and are discussed in various chapters of this textbook. Although both ER and PR are present in the endometrium, ER seems to be the primary mediator of the estrogenic action in the endometrium. The estradiol-receptor complex, after transformational changes, is a transcriptional factor that becomes associated with the estrogen response elements of specific genes. This interaction brings about ER-specific initiation of gene transcription, which promotes the synthesis of specific mRNAs and thereafter the synthesis of specific proteins. Among the many proteins synthesized in most estrogen-responsive cells are additional ERs, as well as PRs. Thus, estradiol acts in the endometrium and in other estrogen-responsive tissues to promote the perpetuation of estrogen action and to promote the responsiveness of that tissue to progesterone.

The endometrial epithelial cells are estrogen-responsive but probably do not replicate as a result of the direct action of estradiol on the epithelial cells. Replication of human endometrial epithelial cells in culture is not increased appreciably, if at all, when estrogen is added to the medium. Further, estrogen acts on mouse uterine stromal cells to promote the synthesis of epithelial cell growth factor (see Fig. 16-27). These growth factors operate in a paracrine manner to cause increased DNA synthesis and replication in the adjacent epithelial cells. This type of paracrine arrangement may be the primary mechanism that mediates estrogen action in hormone-responsive tissues.

**Progesterone Action**

Progesterone action is the differentiation of the endometrium. The two PR isoforms, PR-A and PR-B, are both present in the human endometrium. Because PR-B but not PR-A levels in the endometrium are tightly regulated during the human menstrual cycle, PR-B is presumed to play a more important biologic role. Commonly, the cellular content of PRs is dependent on previous estrogen action. The progesterone-PR complex also promotes gene transcription, but the response to progesterone is strikingly different from that evoked by the estradiol-ER complex.

Progesterone action includes a decrease in the synthesis of ER molecules. This is one means by which progesterone (and synthetic progestins) attenuates estrogen action. Progesterone also acts to increase the rate of enzymatic inactivation of estradiol to estrone through an increase in the activity of an oxidative type 17-HSD enzyme. Progesterone-dependent transcription of a specific gene, namely 17-HSD type 2, is responsible for this enzyme activity. Progesterone also acts to increase sulfation of estrogens (estrogen sulfotransferase), another means of estrogen inactivation. Therefore, progesterone acts as an antiestrogen in at least three ways: (1) by reducing the rate of synthesis of ERs, (2) by bringing about a decrease in the tissue levels of estradiol (through conversion to estrone), and (3) by enhancing estrogen inactivation through sulfation. As in the case of estrogen action in the uterus, tissue recombination experiments using uteri of PR knockout and normal mice demonstrated that many effects of progesterone on epithelial cells are also mediated in a paracrine fashion by PRs in stromal cells but not by those in epithelial cells (see Fig. 16-28).

The most striking consequence of progesterone action is the differentiation of the endometrium. The histologic correlates of differentiation, stromal decidualization and epithelial secretion, are correlated with the presence of nuclear PRs and increased levels of circulating progesterone during the luteal phase. Molecular correlates of progesterone action with respect to differentiation include increased production of lactoferrin and glycoprotein in epithelial cells and prolactin and IGF...
Unless the ovum is fertilized within 24 hours of ovulation, it does not survive. Fertilization takes place in the ampullary (one-third distal) portion of the oviduct. Over the next 2 days, the fertilized ovum remains unattached within the tubal lumen, utilizing tubal fluids and residual attached cumulus granulosa cells to sustain nutrition and energy for early cellular cleavage. After this stage, the solid ball of cells (morula), which is the embryo, leaves the oviduct and enters the uterine cavity. Fortunately, by this time endometrial secretions under the influence of luteal progesterone have filled the cavity and bathe the embryo in nutrients. This is the first of many neatly synchronized events that mark the conceptus-uterine relationship. By 6 days after ovulation, the embryo (now a blastocyst) is ready to attach and implant. At this time, it finds an endometrial lining of sufficient depth, vascularity, and nutritional richness to sustain the important events of early placentation to follow. Just below the epithelial lining, a rich capillaryplexus has been formed and is available for creation of the trophoblasticmaternal blood interface. Later, the surrounding superficial portion of the functionalis zone, now occupying more and more of the endometrial cavity, provides a sturdy splint to retain endometrial architecture despite the invasive inroads of the burgeoning trophoblast.

Progesterone is essential for the maintenance of pregnancy. The blastocyst is dependent on progesterone produced by the corpus luteum at this time. The hCG that is secreted by the trophoblast prevents the regression of the corpus luteum by acting as a surrogate LH. This serves to maintain a continued supply of progesterone for the maintenance of pregnancy until the placental tissue itself starts to produce sufficient quantities of progesterone by 6 to 7 weeks after fertilization.

Studies in experimental and domestic animals have demonstrated that there must be synchronous development of the embryo and endometrium for normal implantation and development to occur. In laboratory animals, there is a discrete window for implantation, which in some species lasts only a matter of hours. The receptive phase of the endometrium is the temporal window of endometrial maturation during which the trophoectoderm of the blastocyst can attach to the endometrial epithelial cells and subsequently proceed to invade the endometrial stroma. In the study of human endometrial receptivity, a key question is the determination of the temporal window of implantation. Only factors expressed during this temporal window can be considered either markers or functional mediators of the receptive state.

The window of uterine receptivity can be inferred from what has been learned from transfer of embryos to uteri of women primed with exogenous estrogen and progesterone preparations. There is a distinct window for embryo transfer leading to implantation, which spans endometrial cycle days 16 to 20. Presumably, the actual window of implantation follows this window of transfer because embryos need to develop further from the four-cell to eight-cell stage to the blastocyst stage before initiation of attachment and frank invasion.

The window of implantation in the humans was estimated to be between days 20 and 24 of the cycle using serial measurements of serum hCG as a marker of initial embryonic-maternal interaction. Thus, it appears that the window of implantation in the human is relatively wide (approximately 4 days). These observations agree with the earlier morphologic data from Hertig and colleagues.

Mechanism of Menstruation

In the absence of pregnancy, failure of the appearance of hCG, despite otherwise appropriate tissue reactions, leads to the vasomotor changes associated with estrogen-progesterone withdrawal and menstrual desquamation. A program of endometrial remodeling is initiated; alterations in the extracellular matrix and infiltration of leukocytes lead to hypoxia-reperfusion injury and sloughing of the functionalis, followed by activation of hemostatic and regenerative processes. The main histologic features of the premenstrual phase are degradation of the stromal reticular network, stromal infiltration by polymorphonuclear and mononuclear leukocytes, and secretory exhaustion of the endometrial glands, whose epithelial cells now have basal nuclei. The endometrium shrinks preceding menstruation, in part as a result of diminished secretory activity and the catabolism of extracellular matrix.

The most prominent and final effect of progesterone and estrogen withdrawal is menstruation. The window of uterine receptivity can be inferred from what has been learned from transfer of embryos to uteri of women primed with exogenous estrogen and progesterone preparations. There is a distinct window for embryo transfer leading to implantation, which spans endometrial cycle days 16 to 20. Presumably, the actual window of implantation follows this window of transfer because embryos need to develop further from the four-cell to eight-cell stage to the blastocyst stage before initiation of attachment and frank invasion.

Control of Endometrial Function Employing Synthetic Hormones

The fertility potential of a woman is primarily determined by the biologic quality of her oocytes, reflected in part by the capacity of the fertilized ovum to divide at an optimal rate and contain a normal chromosomal complement. This biologic quality of the oocyte declines sharply after the age of 35. The biologic potential of the endometrium for successful implantation, however, remains intact even at advanced ages. Oocyte donation from a fertile woman and in vitro fertilization of these donor eggs with the recipient's male partner's sperm, followed by embryo transfer into the uterine cavity of the recipient woman who does not have functioning ovaries (e.g., premature ovarian failure), have been used successfully as a therapeutic strategy to treat infertility. This clinical application has provided unique opportunities to examine the hormonal requirements for endometrial maturation. A number of hormone replacement protocols have been proposed, and many pregnancies have resulted from donor oocytes in women with ovarian failure. The success of these procedures has averaged about 50% (pregnancy rate) per embryo transfer.

The degree of endometrial differentiation in response to exogenous hormones has been evaluated by histologic analysis of endometrial biopsy specimens. The epithelial elements exhibit delayed maturation early during progesterone administration on days 20 to 22, but catch up by day 26. Despite this apparent dysynchrony, the pregnancy rate in these patients with donor oocytes is higher than in conventional in vitro fertilization.

The majority of infertility specialists currently use step-up administration of oral micronized estradiol in 2-, 4-, and 6-mg daily doses followed by 4 to 6 mg of estradiol combined with daily intramuscular progesterone (50 mg) to promote the secretory transformation. Serum estradiol levels in these subjects during the replacement "fulfloch" phase reach preovulatory peak levels of 600 to 1000 pg/mL. Intramuscular injection of 50 mg/day of progesterone in oil generates serum levels of progesterone greater than 10 ng/mL. The length of exposure to progesterone, but not absolute plasma progesterone concentrations achieved after adequate priming...
of the endometrium with estrogen, is a key factor for the development of uterine receptivity. Thus, the exogenous administration of only estradiol and progesterone is sufficient to prepare the endometrium for implantation in the absence of ovarian function. This observation further underscores the essential roles of these steroids in uterine physiology.
APPROACH TO THE WOMAN WITH REPRODUCTIVE DYSFUNCTION

Reproductive dysfunction in an adult woman is most often manifest by disruption of cyclic, predictable menses. Efficient diagnosis of the underlying disorder requires a thorough understanding of female reproductive physiology and pathology and an accurate history and physical examination. Without a critical analysis of clinical findings based on thorough knowledge of normal and abnormal reproductive function, the application of predetermined algorithms of laboratory testing causes unnecessary use of hormone measurements or imaging studies and delays diagnosis.

History

An essential tool for the evaluation of a woman with a reproductive disorder is a carefully recorded history. The history should be obtained from the patient with the aim of assessing the biologic effects of each of the various hormones. Recording the details of pubertal development as a starting point for the rest of the particular symptoms provides critical clues to the etiology of certain reproductive disorders. For example, anovulation manifest by irregular uterine bleeding associated with the polycystic ovary syndrome (PCOS) most often begins during the pubertal years. The onset of gradually progressing hirsutism around puberty is suggestive of nonclassical adrenal hyperplasia or PCOS. In these cases, measurement of serum 17-hydroxyprogesterone may help to differentiate nonclassical adrenal hyperplasia from PCOS (see later). The appearance of hirsutism before puberty or several years after normal pubertal development should alert the clinician to the possibility of ovarian or adrenal neoplasms. The sudden (versus gradual) onset of hirsutism at any age or the presence of virilization should prompt the physician to rule out steroid-secreting ovarian or adrenal tumors. Most women with symptomatic endometriosis suffer from severe episodes of painful menses (dysmenorrhea), which start during pubertal years.

Evaluation of female reproductive function starts with a detailed history of the menses. For example, PCOS is extremely unlikely without a long-standing history of irregular periods since the menarche. By the same token, history of a period of cyclic, predictable menses before the onset of menstrual irregularities should draw attention to hypothalamic or other causes of anovulation. The current frequency, regularity, length, and quantity of uterine bleeding should be carefully recorded for several reasons. First, this information reflects tightly regulated interactions of several tissues, including the hypothalamus, pituitary, ovaries, and endometrium. Second, regular, predictable menses imply ovulation. Third, defining the type of menstrual irregularity may help with the diagnosis of the underlying etiology. For example, prolonged amenorrhea in a thin and estrogen-deficient woman suggests anovulation of hypothalamic origin. Infrequent periods of varying duration and amount of blood loss in a well-estrogenized overweight woman, on the other hand, suggest a primary ovarian dysfunction such as PCOS. It should be kept in mind that anovulation in a thin but well-estrogenized woman may also be due to PCOS. Regular but heavy and prolonged menses with intermittent spotting may be due to uterine anatomic disorders such as adenomyosis or leiomyomas. Finally, neoplastic disorders of the endometrium including endometrial polyps, hyperplasia, or malignancies may be manifest by any pattern of irregular bleeding. The combination of vaginal ultrasonography and endometrial biopsy is extremely sensitive for the diagnosis of endometrial neoplasia.

Disruption of cyclic and predictable menses is a common and alarming symptom that initially brings the patient to the clinician. After a careful evaluation of the menstrual symptoms, the clinician should identify other obvious symptoms of endocrine disorder underlying irregular periods. Pregnancy is the most common cause of amenorrhea (and possibly any other menstrual irregularity) in a woman of reproductive age. In a woman presenting with amenorrhea or any other menstrual irregularity, normal pregnancy, ectopic pregnancy or gestational trophoblastic disease must be excluded at the onset. Careful evaluation of any past reproductive history, as well as of the patient’s sexual activity and contraceptive practices, can provide useful indications of the likelihood of pregnancy. Furthermore, the reproductive history may suggest the possibility of Sheehan’s syndrome of postpartum pituitary necrosis if menses did not resume after a delivery complicated by significant hemorrhage. In such instances, evidence of adrenal and thyroid insufficiency should be sought. A classical symptom of Sheehan’s syndrome is the absence of lactation after delivery related to prolactin deficiency.

Amenorrhea is traditionally categorized as either primary (no history of mensturation) or secondary (cessation of menses after a variable time). The causes of primary amenorrhea are diverse and discussed extensively in Chapter 19 and Chapter 21. Although the distinction between primary and secondary amenorrhea is useful for identifying the mechanism of disease and differential diagnosis, the clinician should be aware that some disorders can initially present with either primary or secondary amenorrhea. For example, most women with gonadal dysgenesis have primary amenorrhea, but some patients have residual follicles and ovulate, and in these women with partial gonadal dysgenesis some menstruation and rare pregnancies may occur before the cessation of ovarian function. Patients with PCOS usually have secondary amenorrhea but occasionally have primary amenorrhea.

Secondary amenorrhea is most often due to chronic anovulation, which can be broadly categorized as (1) hypothalamic dysfunction, (2) galactorrhea-associated, (3) ovarian failure, (4) androgen excess, (5) chronic illness, and (6) primary uterine disease (e.g., intrauterine adhesion formation after a postpartum curettage). Establishing any association of secondary amenorrhea with any life events is extremely useful. Strenuous exercise is often associated with amenorrhea. Weight loss often precedes or accompanies secondary amenorrhea and has been suggested as evidence of hypothalamic dysfunction. An unusual dietary history may be suggestive of bulimia or anorexia nervosa. A history of dilatation and curettage, postpartum endometritis, or disseminated tuberculosis with absent to scant menses should suggest the possibility of intrauterine adhesion.

The presence of any signs or symptoms of estrogen deficiency, including painful intercourse, atrophic vagina, emotional lability, and vasomotor instability, should suggest anovulation of a central nature with low concentrations of circulating gonadotropins (hypogonadotropic hypogonadism) or ovarian failure with elevated gonadotropins (hypergonadotropic hypogonadism).

Galactorrhea in the absence of a recent history of pregnancy is suggestive of a host of diagnostic possibilities and is frequently a manifestation of excessive prolactin secretion, although it may be a result of increased sensitivity of breast tissue to the hormones necessary for milk production. This history frequently reveals drug ingestion as the cause. Various drugs (including several psychotropic agents and antihypertensive agents as well as oral contraceptives) have been implicated. Primary hypothyroidism may be associated with precocious puberty with galactorrhea in the child and with amenorrhea, galactorrhea, or both in the adult woman. A history of excessive nipple manipulation or chest wall disease should be elicited and may well be the cause of galactorrhea. Prolactinomas, the prolactin-secreting adenomas of the pituitary, are a common etiology of galactorrhea related to abnormally high serum levels of prolactin.
Physical Examination

The quantity and distribution of excessive hair growth should be considered in light of the familial history. Hypertrichosisexcessive growth of hair on the extremities, the head, and the backmust be distinguished from true hirsutism, which is the development of facial hair, chest hair, and a male escutcheon with or without signs of virilization in response to increased production of or sensitivity to biologically active androgens. Some degree of hypertrichosis is not uncommon in women of Mediterranean descent, whereas the occurrence of any facial hirsutism in the relatively hairless Asian woman may require thorough investigation. Hirsutism is best documented and quantified with the help of photographs. Virilization is characterized as thickening of voice, severe cystic acne, hair loss, increased muscle mass, and clitoromegaly and implies a more severe degree of androgen excess than that found with hirsutism. The syndrome of complete androgen insensitivity is characterized by sparse to absent pubic and axillary hair because of resistance to androgen.

A careful inspection of the breasts is essential for a thorough physical examination. Classification of the stage of breast development according to the method of Marshall and Tanner is a convenient and valuable adjunct. Whether the breasts appear to have decreased in size recently (e.g., severe androgen excess), whether the areolae are well formed and pigmented (as they are in pregnancy), and whether a discharge (e.g., galactorrhea) can be expressed should be assessed.

A woman with PCOS who has never ovulated or taken a progestin-containing medication may have Tanner stage 4 breast development related to adequate estrogen production, whereas the progression to Tanner stage 5 requires exposure to progesterone through either ovulation or ingesting a progestin (e.g., administration of oral contraceptives). See Chapter 22 for a detailed description and hormonal basis of Tanner staging of breast development.

The vulva, vagina, and cervix also represent sensitive indicators of sex steroid action. Because sensitivity of the genital skin and mucosa to androgen decreases with time from the early stages of fetal development to adulthood, the extent of any virilization can be helpful in suggesting the timing of androgen exposure. The most profound androgenic effects, such as posterior labial fusion with or without formation of a penile urethra, are generally observed in patients exposed to androgens during the first trimester (12 weeks) of pregnancy. Such findings have been described in patients with virilizing congenital adrenal hyperplasia, true hermaphroditism, and drug-induced virilization. Significant postnatal clitoromegaly, on the other hand, requires marked hormonal stimulation and, in the absence of significant exogenous steroids, strongly implicates an androgen-secreting tumor. Measurement of the base of the clitoris versus its length is a more accurate method for the determination of androgen-dependent clitoral growth. A clitoral index, defined as the product of the sagittal and transverse diameters at the base, greater than 35 mm falls outside the 95% confidence interval.

The vagina and uterine cervix are the most sensitive indicators of estrogen action. Under the influence of estrogen, the vaginal mucosa progresses during sexual maturation from a tissue with a shiny, bright red appearance with sparse, thin secretions to a dull, gray-pink rugated surface with copious, thick secretions. Well-estrogenized vaginal mucosa with stretchable cervical mucus (spinnbarkeit) may be indicative of the proliferative phase of the menstrual cycle in an ovulatory woman or extraovarian estrogen formation in an anovulatory woman with PCOS. The biologic activity of estrogen can also be quantified by vaginal cytology.

To summarize, irregular uterine bleeding is a common symptom that brings the woman with reproductive dysfunction to the physician’s office. Various disorders of the hypothalamus, pituitary, ovaries, or uterus or other issues that affect reproductive function may be responsible for this alarming symptom. When pregnancy is ruled out, a detailed history and physical examination should be carefully recorded. In particular, the physician should pay attention to the salient features in the history and biologic indicators of hormone action at target tissues during the physical examination. Analysis of these findings most often leads to a tentative diagnosis. This diagnosis should then be confirmed with laboratory testing.
DISORDERS OF THE FEMALE REPRODUCTIVE SYSTEM

Chronic Anovulation

Chronic anovulation is one of the most common gynecologic problems encountered by the practitioner. These women may present with secondary amenorrhea, infrequent uterine bleeding (oligomenorrhea), or irregular episodes of excessive uterine bleeding. Infertility is an obvious consequence of chronic anovulation.

One group of anovulatory patients are estrogen-deficient. Common findings in this group include hypothalamic anovulation, galactorrhea-hyperprolactinemia (e.g., hypothyroidism, prolactinoma, nonfunctioning pituitary tumor), and premature ovarian failure in a woman of reproductive age. These patients are usually amenorrheic and deficient in estrogen. One serious consequence is bone loss giving rise to osteopenia and osteoporosis. If possible, the underlying cause should be corrected. Hormone replacement should be provided if ovulation cannot be restored.

Women with androgen excess constitute the second major group of anovulatory patients. A serious consequence of anovulation in this group is the greater risk for carcinoma of the endometrium because of unopposed action of estrogen formed continuously in extravarian tissues. The most common disorder of the ovary associated with androgen excess and anovulation is PCOS. There is a new appreciation for the role of insulin resistance in this condition and for the clinical effects of insulin resistance and hyperandrogenism on the risks of developing cardiovascular disease and diabetes mellitus. The clinician must recognize the long-term impact of PCOS and undertake therapeutic management of these anovulatory patients to avoid unwanted consequences. The clinician should also develop a plan with the patient to address long-term complications of unopposed estrogen formation associated with PCOS (e.g., endometrial neoplasia). Oral contraceptives or periodic progesterin supplementation may be provided to prevent endometrial hyperplasia and cancer.

For practical purposes, the following five broad categories include the majority of the etiologic factors giving rise to chronic anovulation in a woman of reproductive age:

1. Hypothalamic anovulation.
2. Hyperprolactinemia.
3. Androgen excess.
4. Premature ovarian failure.
5. Chronic illness (e.g., hepatic or renal failure, acquired immunodeficiency syndrome).

There may be multiple mechanisms responsible for anovulation in chronic illness. Effective treatment of the primary illness may restore normal menses. Alternatively, anovulatory bleeding may be managed by exogenous hormones in these chronically ill patients as outlined further subsequently. The following are detailed descriptions of specific disorders that cause chronic anovulation in a reproductive-age woman.

Hypothalamic Anovulation

Production of LHRH in the neurons of the arcuate nucleus in the hypothalamus and its secretion into the portal vessels in the median eminence in a pulsatile fashion are necessary for the production and secretion of FSH and LH from the pituitary. LHRH neurons depolarize and release LHRH at critical pulse frequencies of 60 to 200 minutes during specific phases of the menstrual cycle to increase or decrease secretion of FSH and LH. Variations in LHRH pulse frequency are achieved, at least partially, by gonadal steroid feedback. Local neuroendulators in the brain, including norepinephrine, dopamine, and -endorphin, mediate the actions of gonadal steroids on the hypothalamus (see Fig. 16-4).

Any disorder of the central nervous system that interferes with this intricate process can thus cause functional hypothalamic anovulation. Some of these disorders may be demonstrated by defined genetic or anatomic evidence such as isolated gonadotropin deficiency (with or without anosmia), infection, suprasellar tumors (pituitary adenomas, craniopharyngioma), and head trauma. These genetic and anatomic disorders affect the function of the hypothalamus, and some of them may be ruled out by history, physical examination, and imaging of the head (Table 16-3).

The most commonly observed form of hypothalamic anovulation, however, is not associated with a demonstrable neuroanatomic finding. This common form is called functional hypothalamic anovulation because it is presumed to involve aberrant but reversible regulation of otherwise normal neuroendocrine pathways. Changes in lifestyle usually result in the return of normal ovulatory cycles. Functional hypothalamic anovulation may be associated with excessive exercise, abrupt weight loss, and emotional distress. It is hypothesized that these stress factors cause anovulation by affecting brain function and the LHRH pulse generator. Other causes of hypothalamic anovulation, however, are relatively rare, and these are covered in detail in other chapters of this textbook (Chapter 8 and Chapter 24).

Functional Hypothalamic Anovulation

Anovulation of hypothalamic origin is characterized by estrogen deficiency and low levels of gonadotropins. No identifiable genetic or anatomic disorders are present in the majority of these patients. The concept of functional hypothalamic anovulation was first postulated 60 years ago as the failure of the hypothalamic-pituitary pathways to release LH from the anterior pituitary. Since then, many clinical studies have confirmed this idea. The data accumulated thus far suggest that the common underlying defect is an alteration in the pulsatile secretion of LHRH. Intriguingly, it has been shown in patients with this disorder that diverse etiologic factors such as malnutrition or caloric restriction, depression or psychogenic stress, excessive energy expenditure related to exercise, or combinations of these have preceded the onset of functional hypothalamic anovulation. Heightened awareness of diet and exercise and unrealistic expectations with respect to the body image of women have most likely contributed to the epidemic of this disorder.

Diagnosis of Functional Hypothalamic Anovulation

Patients with functional hypothalamic anovulation most commonly present with secondary amenorrhea characterized by the absence of menstrual cycles for more than 6 months without evidence of an organic disorder. It should be emphasized once again that the diagnosis of hypothalamic anovulation is one of exclusion. There are many neuroanatomic or genetic disorders that can mimic functional hypothalamic anovulation (see Table 16-3). Thus, a careful and complete diagnostic evaluation is essential to make this diagnosis.

Women with functional hypothalamic anovulation usually present with a history of regular menses for a period of variable length after menarche. Thereafter, this period of normal ovulatory function (by history) is interrupted by anovulation usually manifest by secondary amenorrhea. It should be emphasized that women with functional hypothalamic anovulation may occasionally present with primary amenorrhea.

Women with functional hypothalamic anovulation are typically normal to thin in body weight, driven, and involved in high-stress occupations. The occupation of the patient (e.g., a ballerina or competitive athlete) may be an extremely important clue. A detailed interview may reveal a variety of emotional crises or stressful events (e.g., divorce, death of a friend) preceding the onset of amenorrhea. During the interview, additional environmental and interpersonal factors may become evident, including academic pressure, social maladjustment, and psychosexual problems. When evaluating the patient, one should take note of the current diet regimen, the use of any sedatives or hypnotics, and the rigorousness of the patient's exercise habits. Despite a careful interview, a history of stress, excessive physical exercise, or an eating disorder may not be readily revealed in some women with functional hypothalamic anovulation. These women usually do not complain of hot flashes, which, on the other hand, are commonly observed in ovarian failure.
The physician should exclude a possible hyperprolactinemic etiology (e.g., prolactinoma, hypothyroidism) and evidence of androgen excess (e.g., PCOS) during the physical examination. These women have normal secondary sexual characteristics. The pelvic examination usually shows a thinning vaginal mucosa accompanied by scant to absent cervical mucus with a normal to small uterus, all evidence of estrogen deficiency. Signs

**TABLE 16-3 -- Classification of Anovulation Caused by Disorders of the Hypothalamic-Pituitary Unit**

<table>
<thead>
<tr>
<th>Classification of Anovulation Caused by Disorders of the Hypothalamic-Pituitary Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Functional Hypothalamic Anovulation</strong></td>
</tr>
<tr>
<td>Stress (psychogenic or physical)</td>
</tr>
<tr>
<td>Dieting</td>
</tr>
<tr>
<td>Vigorous exercise</td>
</tr>
<tr>
<td>Chronic illness (e.g., chronic liver or renal failure, AIDS)</td>
</tr>
<tr>
<td><strong>Psychiatric-Medical Emergencies</strong></td>
</tr>
<tr>
<td>Anorexia nervosa</td>
</tr>
<tr>
<td>Medications</td>
</tr>
<tr>
<td>Dopamine antagonists (e.g., haloperidol)</td>
</tr>
<tr>
<td>Opiates</td>
</tr>
<tr>
<td>Antihypertensives (e.g., methyldopa, reserpine)</td>
</tr>
<tr>
<td><strong>Hypothyroidism</strong></td>
</tr>
<tr>
<td>Anatomically or Genetically Defined Pathologies of the Hypothalamic-Pituitary Unit</td>
</tr>
<tr>
<td>Pituitary tumors</td>
</tr>
<tr>
<td>Prolactinoma</td>
</tr>
<tr>
<td>Clinically nonfunctioning adenoma</td>
</tr>
<tr>
<td>GH-secreting adenoma (acromegaly)</td>
</tr>
<tr>
<td>ACTH-secreting adenoma (Cushing's disease)</td>
</tr>
<tr>
<td>Other pituitary tumors (e.g., metastasis, meningioma)</td>
</tr>
<tr>
<td>Pituitary stalk section</td>
</tr>
<tr>
<td>Pituitary apoplexy (including Sheehan's syndrome)</td>
</tr>
<tr>
<td>Pituitary aneurysm</td>
</tr>
<tr>
<td>Infiltrative disease of the pituitary (e.g., lymphocytic hypophysitis, sarcoidosis, histiocytosis X, tuberculosis)</td>
</tr>
<tr>
<td>Empty sella syndrome</td>
</tr>
<tr>
<td>Isolated gonadotropin deficiency (including Kallmann's syndrome)</td>
</tr>
<tr>
<td>Tumors that affect hypothalamic function (e.g., metastasis, craniopharyngioma)</td>
</tr>
<tr>
<td>Infiltrative granulomatous disease of the hypothalamus (e.g., sarcoidosis, histiocytosis X, tuberculosis)</td>
</tr>
<tr>
<td>Head trauma</td>
</tr>
<tr>
<td>Irradiation to the head</td>
</tr>
<tr>
<td>CNS infection</td>
</tr>
<tr>
<td>Other</td>
</tr>
<tr>
<td>ACTH, adrenocorticotropic hormone; AIDS, acquired immunodeficiency syndrome; CNS, central nervous system; GH, growth hormone.</td>
</tr>
</tbody>
</table>

of a well-estrogenized vagina and cervix observed during the physical examination make the diagnosis of hypothalamic anovulation unlikely.

Laboratory tests are obtained to exclude other causes of anovulation and secondary amenorrhea. LH and FSH levels should be obtained. Gonadotropin levels are usually lower than the normal values ordinarily found in the early follicular phase. TSH and prolactin levels are obtained to rule out hypothyroidism and hyperprolactinemia. The progestin challenge test (medroxyprogesterone acetate at 10 mg/day for 10 days) shows either a small spotting episode or absence of withdrawal uterine bleeding in most patients. This confirms that there is a scant or absent estrogenic effect on the endometrium because circulating estradiol levels are typically in the low or early follicular phase range. Measurement of the serum estradiol level is not necessary. Because a suprasellar or large pituitary tumor is in the differential diagnosis, a magnetic resonance imaging (MRI) scan of the head is necessary to rule this out. Imaging of the head is especially important if amenorrhea develops suddenly or is associated with a neurologic sign, both of which make the presence of a tumor more likely.

**Pathophysiology of Functional Hypothalamic Anovulation**

A key observation in functional hypothalamic anovulation is the absence of increased gonadotropin secretion despite the lack of inhibitory factors of ovarian origin, such as estradiol

and inhibit. The secretory pattern of LH is abnormal. The causative factor in women with this type of anovulation is a slowdown in the frequency of pulsatile LHRH secretion. Frequent peripheral blood samples were obtained from these patients to quantify the episodic secretion of LH, which provided an indirect assessment of endogenous LHRH secretion. There is considerable variability in the amplitude and frequency of the pulsatile LH secretion in functional hypothalamic anovulation. When the LH secretory patterns are compared with that of the follicular phase of the menstrual cycle, a characteristic abnormality in the LH pulse frequency and amplitude and on occasion a regression to a pronounced variability similar to what is seen in the prepubertal pattern are present. In severe cases, the frequency and amplitude of LH pulses are markedly reduced. These LH patterns also suggest that LHRH pulsatile secretion is not altered to the same degree in each individual. During the recovery phase of hypothalamic anovulation, a reversal of the LHRH-LH secretory pattern is often present, characterized by a sleep-associated increase in LH amplitude. The response of the pituitary gland to LHRH with respect to production and release of gonadotropins is not impaired in functional hypothalamic anovulation. Intravenous pulsatile LHRH administration can regulate levels of LH and FSH.

Norepinephrine, dopamine, and serotonin produced in the brain have been shown to modulate LHRH or LH release in animal studies. Patients receiving medication that alters these neurotransmitters (e.g., sedatives, antidepressants, stimulants, and antipsychotics) have presented with abnormalities in their menstrual cycles. Thus, it can be noted as circumstantial evidence that disruptions of neural pathways can alter LHRH release in the human. From these observations, it appears that activation of the noradrenergic neurons principally stimulates release of LHRH, whereas dopaminergic and serotoninergic neurons can stimulate
or inhibit LH-RH-LH secretion. A number of neuropharmacologic agents have been used as probes to determine whether LH-RH-LH secretion can be normalized. For example, metoclopramide blocks the action of dopamine. A metoclopramide injection results in a prompt increase in LH secretion in patients with functional hypothalamic anovulation. These studies suggest enhanced dopaminergic activity in functional hypothalamic anovulation. It should be emphasized, however, that chronic administration of dopamine antagonists (e.g., haloperidol, metoclopramide) may also cause anovulation.

Another group of substances that have inhibitory influences on LH-RH secretion are endogenous opioid peptides. Blockage of endogenous opiate receptors by the administration of naloxone, an opiate antagonist, to women with this disorder caused an increase in the frequency and amplitude of pulsatile LH release. Gonadotropin secretion resumes if the activity of the opiate receptor is blocked by long-term naloxone use in these anovulatory patients, and ovulatory function may even be regained in some cases. These studies suggest that there is an overall increase in endogenous opiate activity, which can reduce pulsatile LH secretion in functional hypothalamic anovulation.

Reproductive function may be disrupted by chronic exposure to stress. In fact, activation of the pituitary-adrenocortical system is a common response in patients with chronic stress. In functional hypothalamic anovulation, stressors such as exercise or emotional stress can chronically activate the hypothalamic-pituitary-adrenal axis. Daytime cortisol levels are markedly elevated, and the pituitary response to corticotropin-releasing hormone (CRH) is blunted. The stress response is associated with increased secretion of CRH, adrenocorticotropic hormone (ACTH), cortisol, prolactin, oxytocin, vasopressin, epinephrine, and norepinephrine.

The association between emotional or physical stress and disruption of the reproductive function of the hypothalamus is complex and involves several mechanisms. In the animal model, CRH seems to be an important factor in the inhibition of LH-RH pulsatility. This inhibitory effect can be prevented by coadministration of a CRH antagonist or reversed by the opiate antagonist naloxone, which suggests that the action of CRH is mediated, in part, by activation of the opioidergic system. Moreover, ACTH administration blocks the pituitary response to LH-RH at the pituitary level. In addition, another stress hormone, oxytocin, can inhibit hypothalamic LH-RH secretion. In summary, overproduction of CRH and other stress-related hormones in the brain and activation of the pituitary-adrenocortical system by chronic stress seem to play causative roles in the inhibition of gonadotropin secretion in functional hypothalamic anovulation.

Hypothalamic Anovulation and Exercise

Regular vigorous exercise can lead to menstrual disturbances, a delay in menarche, luteal phase dysfunction, and secondary amenorrhea. Thirty percent of adolescent ballet dancers have problems with the progression of puberty. The mean age of menarche is delayed until 15. In fact, advancement of pubertal stages seems to coincide with times of prolonged rest or following recovery from an injury. The intensity, length, and type of the sport determine the severity of the disease. Activities associated with an increased frequency of reproductive dysfunction are those that favor a lower body weight and include middle-distance and long-distance running, competitive swimming, gymnastics, and ballet dancing.

Competitive athletes show endocrine abnormalities in the central nervous system consistent with those in other forms of functional hypothalamic anovulation. These include elevations on central CRH and endorphin levels.

The management of exercise-related anovulation is dependent on the patient’s choices and expectations. Side effects such as osteoporosis and delay of puberty must be discussed at length with the patient. Decrease in exercise level and behavioral modification may be sufficient for the return of ovulatory function. Hormone replacement should be provided if sufficient results are not achieved. A low-dose oral contraceptive is a suitable option for women of reproductive age.

Hypothalamic Anovulation Associated with Eating Disorders

Two common eating disorders associated with hypothalamic dysfunction are anorexia and bulimia. In anorexia nervosa, there is an extreme loss of weight (weight decrease of greater than 25% of original body weight) and a distorted body image accompanied by a striking fear of obesity. Bulimia is a related disorder characterized by alternating episodes of binge eating followed by periods of food restriction, self-induced vomiting, or excessive use of laxatives or diuretics. About 90% to 95% of these patients are female. Most patients with female eating disorders anorexia and bulimia are from middle-class or upper-middle-class families. The incidence of classical anorexia nervosa is about 1 per 100,000 in the general population. Among high school and college female students, bulimia, however, is fairly common (2%). The incidence of anorexia nervosa peaks twice during the teen years at ages 13 and 17. Bulimia usually begins at a later age, between 17 and 25 years. Anorexia nerva has an extremely high mortality of 9% and is a true medical emergency. Death may be secondary to cardiac arrhythmia, which may be precipitated by diminished heart muscle mass and associated electrolyte abnormalities. These patients are also at increased risk for suicide.

Gonadotropin secretion in anorexic women exhibits a prepubertal pattern that is similar to other forms of hypothalamic anovulation. Transitional patterns of LH secretion are seen when there are moderate degrees of weight recovery and there is a normal or supranormal response to LH-RH. Anovulation can persist in up to 50% of anorexic patients even after achieving normal weight. Both anorexic and bulimic patients exhibit hyperactivation of the hypothalamic-pituitary system. Although the diurnal variation is maintained, there is a persistent hypersecretion of cortisol throughout the day. Cushionoid features, however, are not present, in part because of mild hypercortisolism and also a reduction of peripheral glucocorticoid receptors. Levels of both CRH and endorphin are increased in the central nervous system.

In anorexia nervosa, basal metabolism is decreased because peripheral conversion of thyroxine (T4) to biologically potent triiodothyronine (T3) is decreased. Instead, T3 is converted to reverse T3, an inactive isomer. This alteration is also observed in severely ill patients and during starvation. Anorexics also have partial diabetes insipidus and are unable to concentrate urine appropriately because of the impaired secretion of vasopressin.

Both anorexia nervosa and bulimia are extremely difficult to treat. The most accepted approaches include individual psychotherapy, group therapy, and behavior modification. Patients with eating disorders should have psychiatric consultation and follow-up. This helps with both the diagnosis and treatment. In patients who weigh less than 90% of their ideal body weight, chronic hospitalization is recommended. The treatment of anorexia nervosa includes nutritional therapy; other consequences are estrogen deficiency and generalized effects of malnutrition. Hormone replacement in the form of an oral contraceptive should be provided until ovulatory function is achieved.

Treatment and Management of Functional Hypothalamic Anovulation

Treatment of chronic anovulation resulting from central nervous system hypothalamic disorders should be directed at reversal of the primary cause (e.g., stress management, reduction of exercise, or correction of weight loss). The successful treatment of this disease state is underscored because these women are prone to the development of osteoporosis. For a considerable number of patients, spontaneous recovery of menstrual function takes place after a modification of lifestyle, psychological guidance, or accommodation to environmental stress. Therefore, the initial treatment should be directed to a change in lifestyle and tailored to the individual patient. For individuals who remain amenorrheic, periodic assessment of reproductive status (every 4 to 6 months) is prudent. If anovulation persists for more than 6 months or if reversal of the primary cause is not practical (e.g., professional athletes, ballerinas), a major concern is the long-term effect of hypogonadism, especially on bone metabolism. In addition to estrogen deficiency, IGF-I deficiency, hypercortisolism, or nutritional factors may all contribute to bone loss in this disorder. Unfortunately, epidemiologic data on the risk of fractures and the benefits of hormone replacement are scant.

On the basis of studies of reproductive-age women who have been ovariectomized or who have undergone treatment with LH-RH agonist for endometriosis, bone density would be expected to decrease significantly even within the first 6 months of amenorrhea. Because these patients are often reluctant to take medications, serial bone density studies of the lumbar spine and femur may be necessary to convince them of the necessity to begin estrogen replacement therapy. If the patient is not at risk for thromboembolism and does not smoke cigarettes, a low-dose combination oral contraceptive is a reasonable replacement option. Alternatively, a combination of conjugated estrogens (0.625 mg) and medroxyprogesterone acetate (2.5 mg) daily may be administered to provide estrogenic support.
The progestin (medroxyprogesterone acetate) is added solely to prevent endometrial hyperplasia.

If the patient desires ovulation in order to achieve pregnancy, the most physiologic approach is ovulation induction with pulsatile LHRH. This is currently the best physiologic means of induction because the cause of the anovulatory state is the decrease in endogenous LHRH secretion. Pulsatile intravenous LHRH, 5 µg every 90 minutes, was shown to be effective. Monitoring of serum estradiol levels or follicular development can be minimized because the ovarian follicular response and gonadotropin output mimic the natural menstrual cycle. In these patients, either continuation of pulsatile LHRH or human chorionic gonadotropin, 1500 units intramuscularly every 3 days for a total of four doses, can support the corpus luteum function. The intravenous LHRH treatment results in ovulation rates of approximately 90%, pregnancy rates up to 30%, and hyperstimulation rates of less than 1% per treatment cycle. Because the intravenous LHRH pump is not a practical choice for many women, an alternative strategy is the use of subcutaneous recombinant FSH for the development of one to three follicles and the induction of ovulation with intramuscular hCG followed by luteal support using either intramuscular hCG or progesterone in oil.

**Hyperprolactinemia**

**Prolactin and Reproductive Function**

**Structure and Function of Prolactin**

Prolactin is secreted from the anterior pituitary and is essential for lactation. It is a 198-amino-acid polypeptide with a molecular mass of 22 kd. Its structure is remarkably similar to that of growth hormone. The secondary structure is folded into a globular shape, and three disulfide bonds connect the folds. The amino acid sequences among prolactin, growth hormone, and human placental lactogen show an impressive degree of homology. Prolactin is produced in the lactotroph of the anterior pituitary. In the human pituitary, lactotrophs constitute 10% to 30% of the total pituitary cell mass and are located primarily in the posterior and lateral aspects of the adenohypophysis.

**Regulation of Prolactin Secretion**

A process that includes activation of a number of signaling pathways and gene transcription regulates prolactin synthesis and release. Prolactin is synthesized and packaged into secretory granules and is stored in the cytoplasm pending its release. Exposure to secretagogues, lactotrophs release prolactin from a readily releasable pool, and newly synthesized prolactin replenishes the releasable pool as well as a storage pool. Dopamine, thyrotropin-releasing hormone (TRH), and estradiol were shown to regulate transcription of the prolactin gene.

Dopamine is a well-characterized inhibitor of prolactin secretion. Dopamine release is achieved through the portal vessels of the tuberoinfundibular dopaminergic system. Cell bodies of these neurons are located in the arcuate nucleus, and the axons extend to the median eminence. The biosynthesis and release of dopamine occur within the axonal terminals, which are adjacent to the portal capillaries. Thus, dopamine reaches the lactotroph by way of the portal circulation. Dopamine binds to its receptors on the lactotroph with resultant inhibition of prolactin secretion. Dopamine concentrations are higher in the central portal vessels than in the lateral vessels. The availability of lower amounts of circulating dopamine in the lateral aspects of the pituitary may account for the common presence of prolactinomas in the peripheral portions of the anterior pituitary.

In contrast to dopamine, TRH stimulates the release of pituitary prolactin. TRH stimulates prolactin gene transcription within minutes because specific TRH receptors are present on the lactotroph. The result is an increase in mRNA accumulation in the cytoplasm and acute release of translated prolactin protein. Although the prolactin- and TSH-releasing actions of TRH are distinct, circulating levels of TSH and TSH influence prolactin release in response to TRH stimulation. Subnormal serum levels of TSH and T3, as in the case of primary hypothyroidism, increase TRH-induced prolactin release, whereas higher than normal serum levels of T4 and T3 inhibit prolactin mRNA accumulation and release of protein. Unlike TSH, estradiol amplifies the stimulatory effects of TRH on prolactin release. Thus, a number of factors, including thyroid hormones, estradiol, and antithyroid medications, can modify the effects of TRH on prolactin release.

Estrogen stimulates the secretion of prolactin from the anterior pituitary in a dose- and time-dependent fashion. Administration of synthetic or natural estrogen increases blood prolactin levels in both premenopausal and postmenopausal women. Estrogen increases prolactin synthesis and release by direct stimulation of prolactin gene transcription, lactotroph proliferation, increased TRH receptors, and a decrease in dopaminergic activity. The stimulatory effect of estrogen on prolactin secretion in vivo is usually subtle, as illustrated by mild increases in prolactin levels in women taking oral contraceptives.

**Prolactin Action**

The prolactin receptor belongs to the cytokine receptor superfamily that also includes the growth hormone receptor. Human prolactin, on the other hand, does not bind to the human growth hormone receptor. There are long and short isoforms of the prolactin receptor. Both isoforms can bind prolactin with high affinity and can stimulate prolactin-responsive cells to grow. The functions of prolactin and growth hormone receptors are mediated, at least in part, by two families of signaling molecules: Janus kinases (JAKs) and signal transducers and activators of transcription (STATs).

The prolactin receptor is widely distributed; in some instances in the same tissue in which the ligand prolactin is also expressed. The receptor has been found in the hypothalamus, pituitary gland (both normal and neoplastic), gastrointestinal tract, prostate, bone, decidua, fetal membranes, and Leydig cells as well as in normal and neoplastic breast. The expression of prolactin and its receptors in diverse tissues (pituitary gland, gastrointestinal tract, prostate, decidua, and the breast) is suggestive of the presence of autocrine and paracrine interactions in these tissues. Although a classical endocrine negative feedback cycle between prolactin and pituitary gonadotropin and its target tissues has not been described, these potential local mechanisms within target tissues may play important roles in the regulation of prolactin biosynthesis and action.

**Hyperprolactinemia**

Galactorrhea associated with hyperprolactinemia is one of the most common entities associated with chronic anovulation or secondary amenorrhea, or both. The impaired inhibition of prolactin secretion because of decreased dopamine production.

<table>
<thead>
<tr>
<th>TABLE 16-4 Causes of Hyperprolactinemia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothalamic</td>
</tr>
<tr>
<td>Tumors that affect hypothalamic function</td>
</tr>
<tr>
<td>Metastasis</td>
</tr>
<tr>
<td>Cranopharyngioma</td>
</tr>
<tr>
<td>Glioma</td>
</tr>
<tr>
<td>Infiltrating granulomatous lesions</td>
</tr>
<tr>
<td>Sarcoidosis</td>
</tr>
<tr>
<td>Histiocytosis X</td>
</tr>
<tr>
<td>Tuberculosis</td>
</tr>
<tr>
<td>CNS infection</td>
</tr>
<tr>
<td>Irradiation</td>
</tr>
</tbody>
</table>
Head trauma

Pituitary

Pituitary tumors

- Clinically nonfunctioning adenoma
- GH-secreting adenoma (acromegaly)
- ACTH-secreting adenoma (Cushing's disease)
- Other pituitary tumors (e.g., metastasis, meningioma)

Pituitary stalk section

Pituitary aneurysm

Infiltrative disease of the pituitary

- Lymphocytic hypophysitis
- Sarcoidosis
- Histiocytosis X
- Tuberculosis

Empty sella syndrome

Hypothyroidism

- Chronic illness
- Renal failure
- Liver failure

Ectopic secretion of prolactin

- Bronchogenic carcinoma
- Renal carcinoma

Neurogenic

- Breast manipulation
- Chest trauma
- Chest or upper abdominal surgery
- Herpes zoster infection

Medications causing inappropriate secretion of prolactin

- Estrogen-containing medications (e.g., combination oral contraceptives)
- Dopamine antagonists (e.g., phenothiazines, haloperidol)
- Antihypertensives (reserpine, methyldopa)
- Amphetamines, hallucinogens
- Opiates
- Cimetidine
- Other

Idiopathic

- ACTH, adrenocorticotropic hormone; CNS, central nervous system; GH, growth hormone.

Hyperprolactinemia results in functional and secondary hyperprolactinemia, such as that caused by psychotropic medications. Another common cause of hyperprolactinemia is hypothyroidism associated with overproduction of TRH, which is a known inducer of prolactin secretion. The most common neuroanatomically demonstrable cause of hyperprolactinemia is a prolactin-secreting adenoma or prolactinoma. The processes that lead to the elevation of prolactin are numerous and can be found in Table 16-4.

Hyperprolactinemia Induced by Medications

A large number of medications impair dopaminergic inhibition of prolactin and give rise to hyperprolactinemia. The drugs that block synthesis, metabolism, reuptake, or receptor binding of dopamine reduce dopamine availability and result in the hypersecretion of prolactin (see Table 16-4). Conversely, drugs that enhance dopamine biosynthesis or dopamine agonists suppress the release of prolactin. Galactorrhea, therefore, is a relatively common complication in patients treated with phenothiazines, metoclopramide, reserpine, methyldopa, or similar agents.

Menstrual irregularities or amenorrhea often develops in these patients, reflecting the importance of neurotransmitters and hyperprolactinemia in the regulation of LHRH secretion and impairment of ovarian function.

Hyperprolactinemia and Primary Hypothyroidism

Hyperprolactinemia and anovulation may be associated with primary hypothyroidism. Enlargement of the pituitary gland is frequently seen in long-standing primary hypothyroidism. A number of mechanisms may be involved. First, the clearance of prolactin tends to be decreased in hypothyroidism. Second, patients with severe hypothyroidism may have elevated total and free estradiol levels, giving rise to increased prolactin production stimulated by excess free estrogen. The third, and possibly the most significant, mechanism involves the inhibitory effects of T3 on TRH production and on TRH receptor expression. A decrease in T3 feedback in hypothyroidism may induce an increase in hypothalamic TRH production and in the number of TRH receptors in the lactotroph. Increased TRH action on the lactotroph, in turn, may stimulate prolactin secretion.

Pituitary Prolactinoma

Prolactinomas are the most common hormone-producing pituitary tumors in women. Signs and symptoms include galactorrhea, anovulatory bleeding, amenorrhea, headache, and bitemporal hemianopsia. Prolactinomas are commonly classified as microadenomas for tumors less than 1 cm in diameter and macroadenomas for tumors larger than 1 cm. Macroprolactinomas are more likely to cause headaches and visual symptoms than microprolactinomas. Macroprolactinomas usually give rise to higher serum prolactin levels than microprolactinomas, although circulating prolactin levels associated with microprolactinomas and macroprolactinomas overlap extensively. Prolactin-secreting adenomas are almost always benign. Prolactin measurements and radiologic imaging lead to the diagnosis of these tumors. MRI is a suitable
choice for imaging because it does not involve ionizing radiation and sensitivity is high. The lateral aspects of the pituitary represent the most common locations for prolactinomas. Currently, the most common form of treatment is a dopamine agonist. Among these agents, bromocriptine has been the most commonly used drug. Tumor shrinkage is achieved in most cases. Prolactin levels return to normal; this usually results in the disappearance of headaches and galactorrhea, and the resumption of menses follows. Treatment must usually be continued indefinitely to maintain the euneprolactinemic state. Treatment should be administered to amenorrhoic women even if pregnancy is not the goal because osteopenia often occurs within 6 months of the amenorrhea because of the associated hypoestrogenemia. The increase in parathyroid hormone-related protein (PTHrP) levels in hyperprolactinemic women may also contribute to bone loss.

Side effects of bromocriptine are not uncommon and include syncope, nausea, and vomiting. A long-acting, parenteral form of bromocriptine decreases the incidence of side effects but is not yet approved for use in the United States. Selective D2-type dopamine receptor agonists have become available. These include cabergolide and quinagolide. Both of these drugs have been shown useful in the treatment of bromocriptine-resistant tumors, in which there is decreased expression of the two D2 dopamine receptor isoforms. Cabergoline has a long half-life, requiring only twice-weekly dosing, and is approved for use in the United States; quinagolide requires daily administration and is not yet approved for use in the United States. The long half-life and decreased side effects of cabergoline have led to its increased use.

Treatment in the past was commonly transphenoidal resection of these prolactin-secreting adenomas. There is, however, frequently a recurrence of hyperprolactinemia within 5 to 7 years after the surgery. Thus, dopamine antagonists are the first line of therapy for both microadenomas and macroadenomas.

Continuation of bromocriptine therapy during pregnancy appears to be safe. If bromocriptine is stopped during pregnancy, however, clinically significant enlargement of prolactinomas is relatively uncommon. Approximately 6% of microadenomas and 30% of macroadenomas continue to grow during pregnancy. Eventually, the tumor may reach a size sufficient to cause headache and visual symptoms related to chiasmal compression. When visual symptoms become evident during pregnancy, resuming a dopamine agonist usually reduces the tumor mass and thus the visual symptoms. Transphenoidal resection during pregnancy is rarely required when medical treatment is not sufficient to control visual disturbances and tumor growth. Placental apoplexy is an uncommon but serious complication of prolactinomas during pregnancy.

During pregnancy, serum prolactin levels cannot be used reliably to monitor the size of pituitary adenomas because the decidua (differentiated endometrial tissue under the unique hormonal influence of pregnancy) and normal prolactin-producing pituicytes are additional potential sources of serum prolactin. Therefore, central nervous system symptoms, such as subjective visual symptoms, and objective visual field examinations are used monthly to observe pregnant patients.

**Hyperprolactinemia and Androgen Excess**

Androgen excess and hirsutism were found in a significant number of patients with classical galactorrhea-anovulation syndromes in original reports. This finding was verified by subsequent publications, which demonstrated that almost 40% of patients with pituitary adenomas and hyperprolactinemia had abnormal secretion and metabolism of androgen. Hyperprolactinemia and ultrasound evidence of polycystic ovaries frequently overlap. Hirsutism was observed in 59% of hyperprolactinemic patients with polycystic-appearing ovaries and in 41% of hyperprolactinemic patients with normal-appearing ovaries on ultrasoundography.

Levels of testosterone and androgen precursors, namely dehydroepiandrosterone sulfate (DHEAS) and DHEA, are elevated, whereas testosterone-binding globulin (TeBG, see later) is reduced. Reversal of these changes occurs after lowering the prolactin levels using oral bromocriptine. Furthermore, the levels of elevated androgen precursors are suppressed by dexamethasone, suggesting that hyperprolactinemia may exert a stimulatory action on adrenal C17-steroid secretion.

**Hyperprolactinemia and Bone Loss**

Hyperprolactinemia is commonly associated with reduced bone mineral density. Although hypoestrogenism can cause a decrease in bone mineral content, some eustrogenic patients have also been found to have reduced bone density. A direct action of prolactin on calcium mobilization may be a possible underlying mechanism. Prolactin has been shown to stimulate calcium mobilization from the bone independent of vitamin D and parathyroid hormone in the rat. In addition, prolactin receptors are present in bone cells, thus, prolactin may directly act on the bone. Alternatively, prolactin-dependent increases in PTHrP levels in women with hyperprolactinemia may contribute to the osteoporotic effects of excessive prolactin. Regardless of the mechanism, hyperprolactinemic patients are at risk for development of osteoporosis, and assessment of estrogen and bone mineral status is indicated. Prompt correction of hyperprolactinemia is the approach of choice. If hyperprolactinemia cannot be corrected, hormone replacement in the form of oral contraceptives is indicated.

**Hyperprolactinemia and Hypothalamic Anovulation**

Hyperprolactinemia is usually associated with anovulation, as exemplified by postpartum lactational amenorrhea and the galactorrhea-amenorrhea syndrome. Increased levels of prolactin inhibit the hypothalamic-pituitary-ovarian axis. Both opioid peptides and hypothalamic dopamine regulate the pulsatile secretion of LHRH. Hyperprolactinemia inhibits LHRH activity by interacting with the hypothalamic dopaminergic and opioidergic systems through a short-loop feedback mechanism or by a direct effect on LHRH neurons, in which prolactin receptors are expressed. Both possibilities are consistent with the observation that suppression of prolactin by the dopamine receptor antagonist bromocriptine restores ovulatory function.

**Chronic Anovulation and Androgen Excess**

**Approach to the Patient with Androgen Excess**

Two natural androgens are testosterone, which is transported to target tissue by the circulation, and DHT, which is produced primarily by target tissues. Increased levels of these androgens can lead to hirsutism, which is excessive androgenic hair growth, or to virilization, a more severe form of androgen excess. Hirsutism is defined as the presence of terminal (coarse) hair in locations at which hair is not commonly found in women. It includes facial hair on the cheek, above the upper lip, and on the chin (Figs. 16-30 and B). The presence of midline chest hair is also significant (Fig. 16-30 C). In addition, a male escutcheon, hair on the inner aspects of the thighs, and midline lower back hair entering the intergluteal area are hair growth patterns compatible with androgen excess. A moderate amount of hair on the forearms and lower legs by itself may not be abnormal, although it may be viewed by the patient as undesirable and may be mistaken for hirsutism. Numerous scoring systems are available for quantifying hirsutism. One of the most detailed scales was proposed by Ferriman, Galloway, and Lorenz. A practical and clinically useful means of quantifying hirsutism is recording the hair growth in detail using simple drawings and photographs. In particular, photographs are invaluable for documenting hirsutism accurately.

In contrast to hirsutism, virilization is a more severe form of androgen excess and implies significantly higher rates of testosterone production. Its manifestations include temporal balding, deepening of voice, decreased breast size, increased muscle mass, loss of female body contours, and clitoral enlargement (Fig. 16-31). Even if testosterone levels are moderately increased (>1.5 ng/mL), temporal balding and clitoromegaly may be observed over a long period of time (>1 year) in the presence of persistent androgen excess. A marked increase in androgen secretion, as may occur from production by neoplasms, however, leads to a more full-blown picture of virilization over a short duration of time (less than a few months).

Measurements of an enlarged clitoris may be used for the quantification of virilization. A clitoral length more than 10 mm is considered abnormal. A clitoral length is quite variable, however. An increase in clitoral diameter is a much more sensitive indicator of androgen action. Normal values for clitoral diameter are less than 7 mm at the base of the glans. The most accurate definition of clitoromegaly involves the use of the clitoral index (the product of the width and length of the glans clitoris). A clitoral index greater than 35 mm² is abnormal and correlates statistically with androgen excess.
For practical purposes, measuring the levels of all of these C₁₇-steroids is not clinically necessary in the majority of patients presenting with androgen excess. The purpose of measuring serum testosterone is to establish circulating androgen action (see Fig. 16-32). Although the achievement of this diagnostic yield of this measurement is clearly superior to that of total serum testosterone, the correlation between total and nonTeBG-bound testosterone is excellent and can frequently be predicted. TeBG is one of the primary regulators that determines the amounts of circulating bound and bioavailable testosterone available to act on target tissues. Conditions that decrease TeBG binding (e.g., androgen excess, obesity, acromegaly, hypothyroidism, and liver disease) also increase bioavailable testosterone, thus augmenting the effect of testosterone. TeBG also regulates the circulating amounts of bioavailable estradiol by binding a significant fraction of circulating estradiol. Hence, conditions that decrease TeBG binding (e.g., hirsutism, virilization) and estrogen-dependent disorders (e.g., malignancies of breast and endometrium) also give rise to increased bioavailable (nonTeBG-bound) estradiol.

TeBG is the portion not bound to TeBG but rather loosely associated with albumin, and the fraction not bound by either TeBG or albumin, that is, free or dialyzable testosterone. Biologically active testosterone includes both the free and albumin-bound fractions. Thus, the blood concentration of testosterone available to diffuse into target tissues is referred to as bioavailable or nonTeBG-bound testosterone. The remainder is tightly bound to the protein TeBG.

TeBG synthesis in women. There are two biologically active androgens, testosterone (T) and dihydrotestosterone (DHT). Depending on the menstrual cycle phase or postmenopausal status, 20% to 30% of T is secreted by the ovary. The rest of T production (blood) is accounted for by the conversion of circulating androstenedione (A) to T in various peripheral tissues. Both the adrenal and ovary contribute to circulating A directly or indirectly depending on the cycle phase, reproductive-age versus postmenopausal status, and chronicologic age. Moreover, T may also be formed locally in androgen target tissues. Finally, T is converted to the potent androgen DHT within the target tissues and cells. For example, local conversion of T to DHT in sex skin fibroblasts and hair follicles amplifies androgenic action for clitoral enlargement and hirsutism. DHEA, dehydroepiandrosterone; HDSD, hydroxysteroid dehydrogenase.

Two natural C₁₇-steroids are capable of acting as androgens on target organs: testosterone and DHT. In this chapter, the use of the term androgen refers to either of these steroids. Testosterone in reproductive-age women is produced by two major mechanisms: (1) direct secretion by the ovary, accounting for roughly one third of testosterone production, and (2) conversion of the precursor androstenedione to testosterone in the peripheral (extragonadal) tissues, accounting for two thirds of testosterone production (Fig. 16-32). The peripheral tissues include the skin and adipose tissue. Androstenedione, the direct precursor of testosterone, is produced in both the ovary and the adrenal. The C₁₇-steroids DHEAS and DHEA of adrenal origin and DHEA of ovarian origin indirectly contribute to testosterone formation by first being converted to androstenedione that is subsequently converted to testosterone (see Fig. 16-32).

Whereas testosterone is an androgen, DHEAS and DHEA is a biologically inert steroid. Up to 20 mg of DHEAS is produced daily versus only 3 mg of androstenedione and 8 mg of DHEA per day. These C₁₇-steroids of adrenal origin (DHEAS, DHEA) exert their effects after conversion to the potent androgen testosterone (see Fig. 16-32). Only androstenedione can be converted directly to testosterone. The conversion rate of circulating androstenedione to testosterone in extragonadal tissues is about 5% in both men and women.

TeBG measurement is not clinically necessary in the majority of patients presenting with androgen excess. The purpose of measuring serum testosterone is to establish circulating androgen action (see Fig. 16-32). Although the achievement of this diagnostic yield of this measurement is clearly superior to that of total serum testosterone, the correlation between total and nonTeBG-bound testosterone is excellent and can frequently be predicted. The purpose of measuring serum testosterone is to establish circulating androgen excess, to estimate the source of androgen production, and to detect extremely high values that might originate from an androgen-secreting neoplasm.

The normal serum levels of androgens and their precursors in women vary from laboratory to laboratory. Some approximate levels are as follows: testosterone, less than 0.6 ng/mL; free testosterone, less than 8 pg/mL; nonTeBG-bound testosterone (free + albumin-bound), less than 0.1 ng/mL; androstenedione, less than 3 ng/mL; DHEA, less than 9 ng/mL; DHEAS, less than 2.5 µg/mL; and DHT, less than 0.3 ng/mL. Androstenedione, DHEA, testosterone (total and free), and DHT fluctuate during the ovulatory cycle, and highest levels are found during the luteal phase because of increased secretion of these steroids or their precursors from the corpus luteum. Thus, steroid measurements should be obtained routinely during the early follicular phase, preferably at cycle day 3 (±1 day), so that these levels can be compared and interpreted more easily.

For practical purposes, measuring the levels of all of these C₁₇-steroids is not clinically necessary in the majority of patients presenting with androgen excess. The...
most useful initial test is a serum total testosterone level. An abnormal level in the presence of hirsutism or virilization may be associated with PCOS, hyperthecosis, nonclassical adrenal hyperplasia, or an androgen-secreting neoplasm. The majority of androgen-secreting tumors are of ovarian origin. The likelihood of a neoplasm correlates roughly with increasing testosterone levels. The following tests may be added on the basis of the clinical presentation: serum 17-hydroxyprogesterone (nonclassical adrenal hyperplasia), serum prolactin and TSH (mild androgen excess associated with hyperprolactinemia), serum FSH and LH (elevated LH:FSH ratio in PCOS), serum DHEAS (adrenal tumors), and imaging of ovaries and adrenals (PCOS, tumors).

### Causes of Androgen Excess

A variety of disorders give rise to androgen excess. These include unusual causes such as iatrogenic or drug-induced androgen excess, congenital genital ambiguity (e.g., excessive in utero androgen formation in female pseudohermaphroditism), and conditions unique to pregnancy (luteoma of pregnancy and hyperreactio luteinalis). These uncommon causes and relatively more prevalent disorders associated with androgen excess are listed in Table 16-5. The term extra-ovarian steroid formation is used synonymously for extraglandular, extragonadal, or peripheral steroid formation in this text.

In most hyperandrogenic disorders, androgen originates from more than one source (see Fig. 16-32). For example, testosterone secretion is somewhat increased from the ovary in PCOS, but the bulk of testosterone comes from extraovarian conversion of significantly elevated circulating androstenedione of ovarian origin to testosterone. To add a further twist, patients with PCOS also show increased adrenal output of DHEAS, which (after peripheral conversion to DHEA that is further converted to androstenedione) contributes indirectly to extraovarian testosterone formation (see Fig. 16-32).

When androgen excess is associated with primary amenorrhea, abnormal in utero sexual differentiation should be strongly suspected. These disorders are covered in detail in Chapter 22. Furthermore, before embarking on a major workup for hirsutism or virilization, the physician is well advised to rule out exogenous androgen use. It is best to ask the patient to list all prescriptions and over-the-counter medications that she takes on her own, including injections. This is usually more rewarding than simply asking the patient whether she takes any androgens. Medications that can cause hirsutism or virilization are related to testosterone. These include anabolic steroids and similar compounds.

The most common identifiable cause of androgen excess is PCOS. PCOS is discussed under a separate heading in this chapter. In this section, we first define some of the other disorders associated with hirsutism or virilization. This is followed by a simplified treatment strategy, which may be applied to the majority of hirsute patients within the categories PCOS, nonclassical adrenal hyperplasia, and idiopathic hirsutism.

#### Idiopathic Hirsutism

Excessive hair growth in the absence of demonstrable androgen excess in ovulatory women is also referred to as idiopathic or constitutional and occurs more frequently in certain ethnic populations, particularly in women of Mediterranean ancestry. It is defined as hirsutism in conjunction with regular menstrual cycles and normal levels of serum testosterone. Idiopathic hirsutism is not associated with any sign of virilization. Its cause is not understood completely. It has been proposed that women with idiopathic hirsutism have significantly increased cutaneous 5-reductase activity. At this time, the presence or absence of such an association is not clear. Likewise, it is still unclear which of the 5-reductase isoenzymes (type 1 or 2), if any, is predominant in the development of idiopathic hirsutism.

Idiopathic hirsutism is diagnosed in women who have (1) hirsutism, (2) normal ovulatory function, and (3) normal total or free testosterone levels. Overall, more than 80% of women with cyclic predictable menses are ovulatory. Ovulatory function may be verified by a luteal phase day 7 progesterone level, which should be at least 5 ng/mL. Luteal phase day 7 corresponds to cycle day 17 for 24-day intervals, cycle day 21 for 28-day intervals, and cycle day 28 for 35-day intervals. The presence of oligo-ovulation or anovulation in hirsute women after the exclusion of related disorders (e.g., hyperprolactinemia, hyperprolactinemia, or nonclassical adrenal hyperplasia) is consistent with the diagnosis of PCOS. Thyroid dysfunction and hyperprolactinemia should be excluded by the measurements of TSH and prolactin. The follicular phase basal 17-hydroxyprogesterone level should be measured to exclude 21-hydroxylasedeficient, nonclassical adrenal hyperplasia. The use of exogenous androgens should also be excluded. In summary, the diagnosis of idiopathic hirsutism is one of exclusion, in which ovulatory dysfunction, elevated circulating testosterone, and other causes of androgen excess are ruled out.

#### PCOS, Polycystic Ovary Syndrome

The majority of androgen-secreting tumors arise from the ovary. These ovarian tumors secrete large quantities of testosterone or its precursor androstenedione. They include Sertoli-Leydig cell tumors, hilus cell tumors, and, infrequently, granulosa-theca tumors. Steroidogenically inert ovarian neoplasms such as epithelial cystadenomas or cystadenocarcinomas may produce factors that stimulate steroidogenesis in adjacent non-neoplastic ovarian stroma and induce production of sufficient amounts of androgen precursors such as androstenedione. Approximately 5% of androstenedione is converted to testosterone in extraglandular tissues to give rise ultimately to androgen excess (see Fig. 16-32). Sertoli-Leydig cell tumors, which account for less than 1% of all solid ovarian tumors, tend to occur during the second to fourth decades of life, whereas hilus cell tumors occur more frequently in postmenopausal women. By the time the signs and symptoms of androgen excess cause the patient to seek medical assistance, Sertoli-Leydig cell tumors are usually so large that they are readily palpable on pelvic examination, whereas hilus cell tumors are still small. In women with either type of tumor, serum testosterone is markedly elevated. Granulosa-theca tumors primarily produce estradiol but may occasionally produce testosterone.

| PCOS | Hyperthecosis (a severe PCOS variant) |
| Ovarian tumor (e.g., Sertoli-Leydig cell tumor) |
| Adrenal | Nonclassical adrenal hyperplasia |
| Cushing's syndrome |
| Glucocorticoid resistance |
| Adrenal tumor (e.g., adenoma, carcinoma) |
| Specific conditions of pregnancy |
| Luteoma of pregnancy |
| Hyperreactio luteinalis |
| Aromatase deficiency in fetus |
| Other |
| Hyperprolactinemia, hypothyroidism |
| Medications (danazol, testosterone, anabolizing agents) |
| Idiopathic hirsutism (normal serum testosterone in an ovulatory woman) |
| PCOS, polycystic ovary syndrome |

624
Laboratory Testing to Aid the Differential Diagnosis of Androgen Excess

Non-Neoplastic Adrenal Disorders and Androgen Excess

A number of adrenal disorders, such as classical congenital adrenal hyperplasia, Cushing's syndrome, and glucocorticoid resistance, give rise to androgen excess related to overproduction of testosterone precursors from the adrenal. These disorders are discussed in other chapters. Here, we discuss nonclassical adrenal hyperplasia.

The debate regarding the diagnosis and prevalence of nonclassical adrenal hyperplasia still continues, although the disorder clearly exists. Other terms that have been used to describe this syndrome include late-onset, adult-onset, attenuated, incomplete, and cryptic adrenal hyperplasia. This form of adrenal hyperplasia is caused by a partial deficiency in 21-hydroxylase activity. Although deficiencies in 11-hydroxylase and 3-HSD may result in the disorder, defects in 21-hydroxylase account for more than 90% of cases.

The clinical presentation is almost identical to that of patients with PCOS. The prevalence of this disorder varies according to ethnic background, and the prevalence reported by different investigators has varied widely. The characteristic presentation consists of anovulatory uterine bleeding and progressive hirsutism of the proliferative phase in affected or disease-free ovulatory women.

The prevalence of this disorder varies according to ethnic background, and the prevalence reported by different investigators has varied widely. The characteristic presentation consists of anovulatory uterine bleeding and progressive hirsutism of the proliferative phase in affected or disease-free ovulatory women.

The diagnosis of nonclassical adrenal hyperplasia can be made if the basal 17-hydroxyprogesterone level is higher than 8 ng/mL. No further testing is required in these cases. Values between 2 and 8 ng/mL are considered increased but not diagnostic of nonclassical adrenal hyperplasia. For example, disease-free women or patients with PCOS may also have basal 17-hydroxyprogesterone levels higher than 8 ng/mL. Therefore, use of an LHRH agonist cannot be relied upon to distinguish a neoplasm from another functional state.

Under these circumstances, the scan can be deferred until further investigation has been carried out. Levels of a variety of adrenal steroids including corticosteroids may be elevated in various combinations in the presence of an adrenal tumor. Thus, it is not possible to describe a particular pattern of hormones that defines an adrenal tumor. In general, high levels of serum DHEAS (>8 µg/mL) are suggestive of an adrenal tumor. Testosterone-secreting adrenal tumors are extremely rare. Virilizing ovarian tumors, on the other hand, are encountered much more frequently than those of an adrenal origin. If the presentation is compatible with an androgen-secreting tumor and the ovaries are normal by transvaginal ultrasonography, the adrenals should be evaluated next by imaging.

Testosterone levels three times the upper normal range (or > 2 ng/mL) and DHEAS levels higher than 8 µg/mL have been used as guidelines to investigate further whether neoplasms of the ovary or adrenal are the sources of androgen excess. It should be emphasized that these numbers are provided only as guidelines and not as rules. The following exceptions to these guidelines must be pointed out. First, because tumors secrete androgens episodically, more than one value may be required to detect a significantly elevated level. Second, other precursor steroids are often elevated as well (particularly androstenedione), and their measurement should be considered. Finally, the tumors may give rise to milder elevations of DHEAS and testosterone levels. In particular, even mild elevations in a postmenopausal woman are highly suspicious of an androgen-secreting tumor. By the same token, severely elevated serum testosterone levels (three times the upper normal range or > 2 ng/mL) may be observed in women with severe ovarian hyperthecosis (a severe variant of PCOS) in the absence of a tumor.

Virilization of recent onset and short duration should warrant further investigation, even if testosterone and DHEAS are mildly elevated. With improvements in scanning techniques and ultrasonography for the ovary; abdominal ultrasonography; computed tomography, and MRI for the adrenal the diagnosis of a small (ovarian or adrenal) tumor may be made. However, if no neoplasm can be localized, imaging of the ovary or adrenal after intravenous administration of radiolabeled analogs of hCG (e.g., hCG-B, hCG-A, hCG-D), which detects active steroid-producing tumors, has proved useful. These diagnostic studies should be pursued aggressively before the surgical exploration of a suspected tumor.

Screening may first be carried out by obtaining an 8:00 AM serum 17-hydroxyprogesterone level in an anovulatory patient on any day. Although the majority of women with nonclassical adrenal hyperplasia are anovulatory, some women with this disorder present with regular periods and hirsutism of pubertal onset or with only unexplained infertility. If nonclassical adrenal hyperplasia is suspected in an ovulatory patient on the basis of clinical presentation, an 8:00 AM serum 17-hydroxyprogesterone level should be obtained during the follicular phase because 17-hydroxyprogesterone levels are higher in the luteal phase versus the proliferative phase in affected- or disease-free-ovulatory women. A level less than 2 ng/mL effectively rules out this diagnosis.

The diagnosis of nonclassical adrenal hyperplasia can be made if the basal 17-hydroxyprogesterone level is higher than 8 ng/mL. No further testing is required in these cases. Values between 2 and 8 ng/mL are considered increased but not diagnostic of nonclassical adrenal hyperplasia. For example, disease-free women or patients with PCOS may also have basal 17-hydroxyprogesterone levels in this indeterminate range. The only way to distinguish nonclassical adrenal hyperplasia from PCOS under these circumstances is with an ACTH stimulation test. A rise of 17-hydroxyprogesterone to at least 10 ng/mL 60 minutes after intravenous injection of ACTH has been considered diagnostic of nonclassical adrenal hyperplasia. It should be noted, however, that a higher basal 17-hydroxyprogesterone level within the 2 to 8 ng/mL range is associated with a higher likelihood of nonclassical adrenal hyperplasia. For example, an 8:00 AM 17-hydroxyprogesterone level higher than 4 ng/mL had a sensitivity of 90% for the diagnosis of nonclassical adrenal hyperplasia.

In a patient with androgen excess who belongs to an ethnic group in which there is high prevalence, a baseline 17-hydroxyprogesterone level should be measured at 8:00 AM. In addition, the following patients should have a screening baseline 17-hydroxyprogesterone level obtained: patients with premature pubarche, those with androgen excess of early pubertal onset, women with progressive hirsutism or virilization, and patients with strong family histories of severe androgen excess.
importance to guide laboratory testing. The most important features include the onset and severity of the signs and the rapidly with which they progress. Rapidly progressing severe androgen excess implies an androgen-secreting tumor until proved otherwise. The possibility of a tumor is further underscored in a postmenopausal woman or in a reproductive-age woman with a recent history of cyclic, predictable periods. Ovarian hyperthecosis, a severe variant of PCOS, also gives rise to severe androgen excess that may progress rapidly, especially at the time of expected puberty. Androgen excess emerging at the time of puberty may be indicative of PCOS or nonclassical adrenal hyperplasia.

The most useful initial test to evaluate androgen excess is serum total testosterone (Table 16-6). Testosterone levels in most normal ovulatory women are below 0.6 ng/mL, although the value may vary from laboratory to laboratory. Women with idiopathic hirsutism have cyclic menses and normal testosterone levels. No further testing for androgen excess is required in this group.

If the testosterone level is elevated in an anovulatory woman, serum TSH and prolactin should be obtained next to rule out anovulation associated with hyperprolactinemia. Ultrasonography of the ovaries is also helpful at this time to assess the presence or absence of an ovarian tumor or polycystic ovaries. If the ethnic background of the patient (Ashkenazi Jews, Hispanics, and those of central European ancestry), onset of hirsutism (puberty), or family history is suggestive of nonclassical adrenal hyperplasia, a baseline serum 17-hydroxyprogesterone level should be obtained at 8:00 AM. Rare etiologies of androgen excess include an adrenal tumor, Cushning's syndrome, and glucocorticoid resistance. A serum DHEAS level and adrenal imaging are required to assess the presence or absence of an adrenal tumor. A computed tomographic scan, MRI scan, or abdominal ultrasonography may be used to assess the adrenals, depending on the expertise of the local radiology laboratory. A screening test for Cushings's syndrome and glucocorticoid resistance may be performed to explore rare adrenal causes of androgen excess. See Chapter 14.

Most women with chronic anovulation and mild to moderate hirsutism of pubertal onset fall into the category of PCOS. These women have high normal or elevated testosterone levels and no other laboratory abnormalities. When other diagnoses are ruled out either by laboratory testing or on clinical grounds, a diagnosis of PCOS can be made.

Treatment of Hirsutism

Therapy for androgen excess should be directed toward its specific cause and at suppression of abnormal androgen secretion. Specific treatments for hirsutism and virilization would be indicated for the following conditions: ovarian and adrenal tumors, hyperthecosis, Cushings's syndrome, and adrenal hyperplasia. Neoplasms warrant surgical intervention and are not discussed in greater detail. Suppression with an LHRH analogue may be tried initially for ovarian hyperthecosis. Unfortunately, bilateral oophorectomy is inevitable to control androgen excess arising from hyperthecosis in the majority of patients (see later). Patients with adrenal disease are treated specifically. For Cushings's syndrome, treatment is according to the source of hypercortisolism. For nonclassical adrenal hyperplasia, glucocorticoid replacement should be implemented as for adrenal insufficiency. When treating androgen excess associated with nonclassical adrenal hyperplasia, an androgen antagonist (e.g., spironolactone) in combination with an oral contraceptive or a glucocorticoid may be used. The doses of glucocorticoids needed to suppress the adrenal, however, can often cause symptoms and signs of glucocorticoid excess during long-term treatment. Thus, a combination oral contraceptive plus spironolactone should be favored to treat androgen excess if the patient responds to this treatment with decreased hirsutism. Greater details of glucocorticoid therapy may be found in Chapter 14 and in Chapter 22.

The general treatment of androgen excess is directed toward the prevention of abnormal hair growth and virilization. For practical purposes, the same approach is used for androgen excess associated with idiopathic hirsutism, PCOS, and nonclassical adrenal hyperplasia. The existing hair follicles and manifestations of virilization (e.g., thickening of voice, clitoromegaly, temporal balding) remain even after the elimination of excessive androgen production. Therefore, terminal hair should be removed by mechanical methods (e.g., electrolysis) at least 3 months after androgen suppression is achieved. Patients with clitoromegaly may be referred to a urologist for clitoral reduction surgery after the source of virilization is effectively eliminated. The following medications are available for the treatment of androgen excess and hirsutism.

Oral contraceptives reduce circulating testosterone and androgen precursors by suppression of LH and stimulation of TeBG levels and, thereby, reduce hirsutism in hyperandrogenic patients. Oral contraceptives decrease circulating androgen in patients with PCOS and synergize with the effects of antiandrogens. It is possible that oral contraceptives may further improve the results of antiandrogen therapy in idiopathic hirsutism. It is advisable to use an oral contraceptive containing either 30 or 35 µg of ethinyl estradiol to achieve effective suppression of LH.

Spironolactone

The most common androgen blocker used for the treatment of hirsutism in the United States is spironolactone, an aldosterone antagonist structurally related to progestins. Spironolactone is effective for abnormal hair growth associated with PCOS or idiopathic hirsutism.

Because spironolactone acts through mechanisms different from that of oral contraceptives, the overall effectiveness is improved by combining these two medications, even in patients with idiopathic hirsutism. Apart from the inhibition of steroidogenesis and acting as an androgen antagonist, spironolactone has a significant effect in inhibiting 5-reductase activity. Basic and several clinical studies clearly point to the efficacy of spironolactone for hyperandrogenism and suggest that the

TABLE 16-6 – Laboratory Tests for the Differential Diagnosis of Androgen Excess

<table>
<thead>
<tr>
<th>Test</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Testing</td>
<td>Total testosterone, Prolactin</td>
</tr>
<tr>
<td>Further Testing Based on Clinical Presentation</td>
<td>17-Hydroxyprogesterone (8:00 AM), 17-Hydroxyprogesterone 60 min after intravenous ACTH, Cortisol (8:00 AM) after 1 mg dexamethasone at midnight, DHEAS, Androstenedione, Imaging of ovaries (transvaginal ultrasonography), Imaging of adrenals (abdominal ultrasonography, CT scan, MRI)</td>
</tr>
<tr>
<td>Nuclear imaging after intravenous administration of radiolabeled cholesterol</td>
<td>ACTH, adrenocorticotropic hormone, CT, computed tomography; DHEAS, dehydroepiandrosterone sulfate; MRI, magnetic resonance imaging; TSH, thyroid-stimulating hormone.</td>
</tr>
</tbody>
</table>

*See text.*
principal effect is related to its peripheral blocking ability. [338]

Doses of spironolactone have varied in clinical studies from 50 to 400 mg daily. Although doses of 100 mg/day are generally effective for the treatment of hirsutism, higher doses (200 to 300 mg/day) may be preferable in extremely hirsute or markedly obese women. [337] [358] Thus, it is recommended to start with 100 mg/day and gradually increase the dose by 25 mg/day increments every 3 months up to 200 mg/day on the basis of the response. This approach may be helpful to minimize side effects such as gastritis, dry skin, and anovulation.

In patients with normal renal function, hyperkalemia is almost never seen. Hypotension is rare except in older women. Monitoring, however, is imperative for electrolytes and blood pressure within the first 2 weeks at each dose level. Adjustments in dose should be made only after 3 to 6 months, as with other androgens, to account for the slow changes in the hair cycle. Patients usually note an initial transient diuretic effect. Some women with normal cycles complain of menstrual irregularity with spironolactone. The latter complaint is remedied by either a downward dose adjustment or the addition of an oral contraceptive. The mechanism for abnormal bleeding is unclear. In women with oligomenorrhea, such as those with PCOS, resumption of normal menses may occur. In part, this may be due to an alteration in levels of circulating androgens, although LH levels have only occasionally been noted to decrease. [338] Another important consideration is the potential in utero feminizing effect of this antiandrogen on the genitalia of a 46,XY fetus. Thus, effective contraception should always be provided in women taking spironolactone.

Cyproterone Acetate

Cyproterone acetate is a 17-hydroxyprogesterone acetate derivative with strong progestagenic properties. Cyproterone acetate acts as an antiandrogen by competing with DHT and testosterone for binding to the androgen receptor. There is also some evidence that cyproterone acetate and ethinyl estradiol in combination can inhibit 5-reductase activity in skin. [361] Cyproterone acetate is currently not available in the United States but has been used in other countries. The drug is mostly administered in doses of 50 to 100 mg from days 5 through 15 of the treatment cycle. Because of its slow metabolism, it is administered early in the treatment cycle, whereas ethinyl estradiol, when added, is usually used at 50-µg doses between days 5 and 26. This regimen is needed for menstrual control and is usually referred to as the reverse sequential regimen. Cyproterone acetate in doses of 50 to 100 mg/day, combined with ethinyl estradiol at 30 to 35 µg/day, is as effective as the combination of spironolactone, 100 mg/day, and an oral contraceptive in the treatment of hirsutism. [364] In smaller doses (2 mg), cyproterone acetate has been administered as an oral contraceptive in daily combination with 50 or 35 µg of ethinyl estradiol. This regimen is primarily suited for individuals with a milder form of hyperandrogenism. [365]

Finasteride

Finasteride inhibits 5-reductase activity and has been used primarily for the treatment of prostatic hyperplasia. [366] It can also be used in the treatment of hirsutism. [367] [368] At a dose of 5 mg/day, a significant improvement of hirsutism is observed after 6 months of therapy, without significant side effects. In hirsute women, the decline in circulating DHT levels is small and cannot be used to monitor therapy. Although this treatment regimen increases testosterone levels, testosterone dihydrogenone (TeBG) levels remain unaffected. [359]

Finasteride primarily inhibits 5-reductase type 2. As hirsutism results from a combination of effects of type 1 and type 2, this agent is only partially effective. Although prolonged experience with finasteride is lacking, one of the potential advantages of this agent appears to be its benign side-effect profile. One study showed efficacy with 1 year of hirsutism treatment. [369] It was also reported that finasteride is less effective than spironolactone with respect to the reduction of hirsutism. [370] Nevertheless, finasteride represents a useful option for treating women with hirsutism at a dose of 5 mg/day for prolonged periods because of its benign side-effect profile and good tolerance by patients.

Flutamide

Flutamide is a potent antiandrogen used in the treatment of prostate cancer. [357] It has been shown to be effective in the treatment of hirsutism. [358] [359] [360] [361] Nevertheless, occasional severe hepatotoxicity makes this drug unsuitable for the indication of hirsutism. [362] [363]

Summary

The preceding medications may be effective when administered as individual treatments. Patients with the most common form of hirsutism (i.e., PCOS) are often initially treated with a combination of two agents, one that suppresses the ovary (e.g., oral contraceptive) and another agent that suppresses the extraglandular (peripheral) action of androgens (e.g., spironolactone). Thus, an oral contraceptive containing 30 to 35 µg of ethinyl estradiol combined with spironolactone, 100 mg/day, is the initial treatment of choice. Even in women with idiopathic hirsutism, the addition of an oral contraceptive to the antiandrogen spironolactone can improve efficacy and prevent abnormal bleeding. For women with only minor complaints of hirsutism, the use of an oral contraceptive alone may be an appropriate first approach.

Because the growth phase of body hairs lasts 3 to 6 months, one should not expect a response before 6 months from the onset of the treatment. Objective means should be used to assess changes in hair growth. Scoring systems and evaluation of anagen hair shafts are difficult; taking photographs is the simplest and most objective tool. Patients are often unaware that change is indeed taking place unless there is some objective measurement. Pictures of face and selected midline body areas before and during therapy are especially useful for the encouragement of the patient and compliance with the treatment. Suppression of androgen production and action only inhibits new hair growth. Thus, existing coarse hair should be removed mechanically. Plucking, waxing, and shaving are ineffective for hair removal and cause irritation, folliculitis, and ingrown hairs. Electrolysis is still the method of choice. Laser epilation is relatively new and needs further evaluation. [364]

The majority of patients with PCOS and idiopathic hirsutism respond to this treatment within 1 year. Patients should be encouraged to continue treatment for at least 2 years. After this, depending on the wishes and clinical responses of patients, therapy can be stopped and the patient reevaluated. Many patients require continuous treatment for the suppression of hirsutism.

The Polycystic Ovary Syndrome

PCOS is the most common form of chronic anovulation associated with androgen excess, perhaps occurring in 5% to 10% of reproductive-age women. [363] The diagnosis of PCOS is made by excluding other hyperandrogenic disorders (e.g., non-classical adrenal hyperplasia, androgen-secreting tumors, and hyperprolactinemia) in women with chronic anovulation and androgen excess. [365]

During the reproductive years, PCOS is associated with important reproductive morbidity including infertility, irregular uterine bleeding, and increased pregnancy loss. [364] The endometrium of the patient with PCOS must be evaluated by biopsy because long-term unopposed estrogen stimulation leaves these patients at increased risk for endometrial cancer. [366] PCOS is also associated with increased metabolic and cardiovascular risk factors. [367] These risks are linked to insulin resistance and compounded by the common occurrence of obesity, although insulin resistance is also present in nonobese women with PCOS. [368]

PCOS is now viewed as a heterogeneous disorder of multifactorial etiology. PCOS risk is significantly increased with a positive family history of chronic anovulation and androgen excess, and this complex disorder may be inherited in a polygenic fashion. [369] [370] [371] [372] [373] [374] [375]

Historical Perspective

In their pioneering studies, Stein and Leventhal described an association between the presence of bilateral polycystic ovaries and signs of amenorrhea,
Prolactin and TSH should be obtained routinely to rule out mild androgen excess and anovulation that may be associated with hyperprolactinemia. If basal LH levels are elevated, it is important to consider the possibility of androgen excess. As emphasized earlier, elevated total testosterone is the most direct evidence for androgen excess. Varying levels of testosterone are present in women with PCOS.

Glucocorticoid resistance is characterized by preserved diurnal rhythm despite significantly elevated cortisol, ACTH, and adrenal Cₐₙ-s-steroid levels. This morphologic change is called stromal hyperthecosis and appears to be directly correlated with circulating insulin levels. Glucocorticoid resistance and the polycystic ovary syndrome: mechanism and implications for pathogenesis. Endocr Rev 1997; 18:774800. Copyright © 1997 by The Endocrine Society.

One of the most significant discoveries regarding the pathophysiology of PCOS was the demonstration of a unique form of insulin resistance and associated hyperinsulinemia. For the first time, Burghen and co-workers reported this finding in 1980. The presence of insulin resistance in PCOS has since been confirmed by a number of groups worldwide. Diagnosis of Polycystic Ovary Syndrome and Laboratory Testing

One of the most prominent features of PCOS is the history of ovulatory dysfunction (amenorrhea, oligomenorrhea, or other forms of irregular uterine bleeding) of pubertal onset. Thus, a clear history of cyclic predictable menses of menarchal onset makes the diagnosis of PCOS unlikely. Acquired insulin resistance associated with significant weight gain or an unknown cause, however, may occasionally induce the clinical picture of PCOS in a woman with a history of previously normal ovulatory function. Hirsutism may develop prepubertally or during adolescence, or it may be absent until the third decade of life. Seborrhea, acne, and alopecia are other common clinical signs of androgen excess. In extreme cases of ovarian hyperthecosis (a severe variant of PCOS), clitoromegaly may be observed. Nonetheless, rapid progression of androgenic symptoms and virilization are rare in ordinary PCOS. Some women may never have signs of androgen excess because of genetic differences in target tissue sensitivity to androgens. Infertility related to the anovulation may be the only presenting symptom.

During the physical examination, it is essential to search for and document signs of androgen excess (hirsutism or virilization or both), insulin resistance (acanthosis nigricans), and the presence of unopposed estrogen action (well-rugated vagina and stretchable clear cervical mucus) to support the diagnosis of PCOS. It should be noted that none of these signs are specific for PCOS and may be associated with any of the conditions listed under the differential diagnosis of PCOS.

The currently recommended diagnostic criteria for PCOS are androgen excess and ovulatory dysfunction with the exclusion of other causes of androgen excess such as nonclassical adrenal hyperplasia (21-hydroxylase deficiency), hyperprolactinemia (with or without hypothyroidism), or androgen-secreting neoplasms (Table 16-8). The exclusion of hyperprolactinemia, hyperthyroidism, nonclassical adrenal hyperplasia, and tumors requires a careful history and physical examination as well as laboratory testing as detailed previously (see Table 16-6). Cushing's syndrome and glucocorticoid resistance may give rise to androgen excess and anovulation after a period of normal ovulatory function in teens. An 8:00 AM cortisol level after dexamethasone (1 mg) administration at midnight is a useful screening test for both conditions. Cushing's syndrome may be recognized by its typical signs, whereas 8:00 AM and 4:00 PM cortisol levels are essential to suspect the diagnosis of glucocorticoid resistance. Glucocorticoid resistance is characterized by preserved diurnal rhythm despite significantly elevated cortisol, ACTH, and adrenal C₁₇α-steroid levels and absence of cushingoid symptoms and signs.

As emphasized earlier, elevated total testosterone is the most direct evidence for androgen excess. Varying levels of testosterone are present in women with PCOS. Rarely, serum testosterone levels higher than 2 ng/mL may be encountered in association with the most severe form of PCOS, ovarian hyperthecosis. Overall, it is much more common to observe high normal levels or borderline elevations of testosterone in women with PCOS.

Prolactin and TSH should be obtained routinely to rule out mild androgen excess and anovulation that may be associated with hyperprolactinemia. If basal LH levels

---

**TABLE 16-7 -- Differential Diagnosis of Polycystic Ovary Syndrome**

<table>
<thead>
<tr>
<th>Idiopathic hirsutism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyperprolactinemia, hypothyroidism</td>
</tr>
<tr>
<td>Nonclassical adrenal hyperplasia</td>
</tr>
<tr>
<td>Ovarian tumors</td>
</tr>
<tr>
<td>Adrenal tumors</td>
</tr>
<tr>
<td>Cushing's syndrome</td>
</tr>
<tr>
<td>Glucocorticoid resistance</td>
</tr>
<tr>
<td>Other rare causes of androgen excess</td>
</tr>
</tbody>
</table>

**TABLE 16-8 -- Criteria for Clinical Diagnosis of Polycystic Ovary Syndrome**

<table>
<thead>
<tr>
<th>Androgen excess with or without skin manifestations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Irregular uterine bleeding (anovulation or oligo-ovulation)</td>
</tr>
<tr>
<td>Absence of other causes of androgen excess (e.g., nonclassical adrenal hyperplasia, Cushing's syndrome, glucocorticoid resistance or ovarian or adrenal tumor)</td>
</tr>
</tbody>
</table>

(Demonstration of polycystic ovaries on ultrasonography is a common feature that is not essential for diagnosis)
are used as a marker for PCOS, a significant number of patients slip through the cracks because they do not all manifest elevated LH levels or increased LH:FSH ratios. This issue prompted a National Institute of Child Health and Human Development-sponsored consensus conference on diagnostic criteria for PCOS in 1990 and the recommendation that LH and the LH:FSH ratio are not required for the diagnosis of PCOS. The heterogeneity of LH values in PCOS may be caused by the pulsatile nature of LH secretion and negative effects of obesity on LH levels. Thus, an elevated LH:FSH ratio is supportive of the diagnosis of PCOS and may be useful in differentiating mild cases of nonobese PCOS without prominent androgen excess from hypothalamic anovulation. Failure to exhibit an elevated LH level, however, is of no diagnostic value.

By definition, nonclassical adrenal hyperplasia is not manifest as genital virilization of external genitalia. Hyperandrogenic symptoms most commonly appear peripubertally or postpubertally. A 17-hydroxyprogesterone level at 8:00 AM is essential to rule out 21-hydroxylase-deficient nonclassical adrenal hyperplasia. The majority of symptomatic patients with nonclassical adrenal hyperplasia are anovulatory and can be tested on any day at 8:00 AM. In an occasional ovulatory woman with androgen excess and suspected nonclassical adrenal hyperplasia, the 17-hydroxyprogesterone level should be obtained during the follicular phase to maximize specificity. A basal 17-hydroxyprogesterone level less than 2 ng/mL effectively rules out nonclassical adrenal hyperplasia. Patients with (proliferative phase 8:00 AM) 17-hydroxyprogesterone levels higher than 2 ng/mL should undergo an ACTH stimulation test. A 17-hydroxyprogesterone level higher than 10 ng/mL at 60 minutes after intravenous injection of ACTH is diagnostic of nonclassical adrenal hyperplasia. Please refer to Chapter 11 and Chapter 19 for details of the ACTH stimulation test. A screening test for Cushing's syndrome or glucocorticoid resistance should be performed as clinically indicated (see Chapter 14 and Chapter 22).

Serum DHEAS levels may be increased (up to 8 µg/mL) in about 50% of anovulatory women with PCOS. DHEAS originates almost exclusively from the adrenal. The etiology of adrenal hyperactivity in PCOS is not known. Obtaining a DHEAS level routinely in a patient with PCOS is not recommended because it does not change the diagnosis or management. On the other hand, if an adrenal tumor is suspected, a DHEAS level should be obtained. DHEAS levels above 8 µg/mL may be associated with steroidogenically active adrenal tumors, and imaging is then indicated.

The use of ultrasonography in the diagnosis of PCOS must be tempered by an awareness of the broad spectrum of women with ultrasonographic findings characteristic of polycystic ovaries. The typical polycystic-appearing ovary emerges in a nonspecific fashion when a state of anovulation persists for any length of time (Fig. 16-35)(Figure Not Available). Whether diagnosis is by ultrasonography or by the traditional clinical and biochemical criteria, a cross-section of all anovulatory women at any point in time reveals that approximately 75% have polycystic-appearing ovaries as determined by ultrasonography. As there are numerous causes of anovulation, there are also numerous reasons for polycystic ovaries. A similar clinical picture and ovarian condition can reflect any of the dysfunctional states discussed previously. In other words, the polycystic-appearing ovary is the result of a functional derangement but not a specific central or local defect.

The application of rigid endocrine or clinical criteria for the diagnosis of PCOS results in a focused portion of the broad clinical spectrum of PCOS. This particularly applies to diagnosing PCOS with the use of ultrasonography. Criteria include an increase in the number of follicles and their frequent necklace-like arrangement, accompanied by an increase in ovarian volume related to stromal increase (see Fig. 16-35)(Figure Not Available). From 8% to 25% of normal women demonstrate ultrasonographic findings typical of polycystic ovaries. Even 14% of women taking oral contraceptives have been found to have this ultrasonographic picture. Thus, ultrasonography is not a tool of choice and its use as a diagnostic aid is not of value. Magnetic resonance studies further confirm the unreliability of the imaging finding (i.e., the polycystic ovary) that was once presumed to be diagnostic of this condition.

Biochemical evidence of insulin resistance or glucose intolerance is also not necessary for the diagnosis of PCOS. Glucose intolerance should nonetheless be investigated. Therefore, plasma glucose levels should be measured after a 75-g glucose load as a screen for glucose intolerance.

Women with PCOS commonly present with irregular uterine bleeding in the form of infrequent periods (oligomenorrhea) or amenorrhea. It is not necessary to document anovulation by ultrasonography, progesterone levels, or otherwise, especially if menstrual cycles are irregular with periods of amenorrhea. To confirm the diagnosis of chronic anovulation and unopposed estrogen exposure, most clinicians perform a progesterin challenge test after a negative urine pregnancy test. Because endometrium is exposed to estradiol chronically in PCOS, these women respond to a challenge with a progestin (e.g., medroxyprogesterone acetate 10 mg/day orally for 10 days) by uterine bleeding within a few days after the last pill of progestin. The reasons for lack of uterine bleeding after a progestin challenge include pregnancy, insufficient progesterone exposure of the endometrium, or anatomic defect. If uterine bleeding does not follow progestin challenge, pregnancy should be ruled out again along with other causes of chronic anovulation as described in this chapter. An anatomic defect such as intrauterine adhesions may be ruled out with a hysterosalpingogram or hysteroscopy.

Finally, during the initial work-up, it is advisable to obtain an endometrial biopsy specimen using a plastic miniscissure cannula (e.g., Pipelle) in the physician's office. If chronic anovulation persists, endometrial biopsies should be repeated periodically. Response to oral contraceptives or periodic progestin treatment with predictable withdrawal bleeding episodes is reassuring, and these patients with predictable bleeding patterns do not need endometrial sampling during these treatments. In untreated patients, the risk of endometrial hyperplasia and malignancy is significantly increased even in young women with PCOS because of unopposed estrogen exposure.

Gonadotropin Production in Polycystic Ovary Syndrome

Women with PCOS have higher mean concentrations of LH but low or low-normal levels of FSH compared with levels found in normal women in the early follicular phase. The elevated LH levels are partly due to increased sensitivity of the pituitary to LHRH stimulation manifest by increases in LH pulse frequency and, in particular, LH pulse amplitude. Intestinally, an increased level of LH bioactivity accompanies high levels of LH in women with PCOS.

The elevated LH levels in PCOS are presumed to be primarily due to accelerated LH:RH-LH pulsatility. Central opioid tone appears to be suppressed because the pattern of LH secretion does not change in response to naloxone. Indeed, the enhanced pulsatile secretion of LH has been attributed to a reduction in hypothalamic opioid inhibition caused by the chronic absence of progesterone. An increase in amplitude and frequency of LH secretion also correlates with the steady-state levels of circulating estrogen.

In obese women with PCOS, LH levels are not increased. The increase in LH pulse frequency is characteristic of the anovulatory state regardless of the body fat content. LH pulse amplitude, however, is comparatively normal in overweight women with PCOS, whereas it is...
increased in nonobese women with PCOS. The overall LH reduction in obese women with PCOS may also be due to factors other than changes in LH pulse amplitude. It should be noted again that a low LH value does not rule out the diagnosis of PCOS, whereas a high LH/FSH ratio is supportive of this diagnosis in an anovulatory woman.

Steroid Production in Polycystic Ovary Syndrome

Ovulatory cycles are characterized by cyclic fluctuating hormone levels that regulate ovulation and menses. Anovulation in women with PCOS, on the other hand, is associated with steady-state levels of gonadotropins and ovarian steroids. In patients with persistent anovulation, the average daily production of estrogen and androgens is both increased and dependent on LH stimulation. This is reflected in higher circulating levels of testosterone, androstenedione, DHEA, DHEAS, 17-hydroxyprogesterone, and estrone. Testosterone, androstenedione, and DHEA are secreted directly by the ovary, whereas DHEAS, elevated in about 50% of anovulatory women with PCOS, is almost exclusively an adrenal contribution. Circulating levels of androstenedione, secreted by polycystic ovaries, are particularly high.

Estrogen arises primarily from peripheral aromatization of androstenedione and, in part, from ovarian secretion. Estrone itself is not a potent estrogen but can be viewed as a precursor that must be converted to estradiol to exert full estrogenic action. The presence of a number of 17-HSD isoenzymes with overlapping activities that catalyze the conversion of estrone to estradiol in peripheral (extraovarian) tissues is, in part, responsible for maintaining estradiol production in women with PCOS. Increased androstenedione leads to a detectable increase in circulating levels of estradiol in women with PCOS compared with estradiol levels measured during the first few days of an ovulatory cycle. This occurs through aromatase and 17-HSD activities in extraovarian tissues such as skin and subcutaneous adipose tissue. Also, local conversion of estrone to estradiol is an important physiologic process for certain estrogen target tissues such as disease-free breast and genital skin. Finally, local conversion can also promote the growth of pathologic estrogen-dependent tissues such as breast cancer and endometriosis.

Overall, androstenedione of ovarian origin is the most strikingly elevated steroid in PCOS. Androstenedione is not biologically active but serves as a dual precursor for both androgens (testosterone that is further converted to the biologically far stronger androsterone DHT) and estrogen (estrogen that is further converted to biologically active estradiol in target tissues). Estradiol is an extremely potent steroid. Biologically effective circulating levels of estradiol are measured using units of pg/mL or pmol/L, whereas biologically effective levels of testosterone are measured in units of ng/mL or nmol/L, and circulate at 10 to 100 times the physiologic levels of estradiol. Thus, even small rates of conversion of androstenedione to estrone may have a significant biologic impact, whereas markedly elevated production of androstenedione is required to produce significant amounts of testosterone and manifestations of androgen excess. Because such elevated production of androstenedione does occur in PCOS, extraovarian production of testosterone is biologically significant in this disease. In contrast, in postmenopausal women, who have much lower levels of androstenedione, extraovarian production of testosterone is less important. On the other hand, relatively small quantities of estrone (and estradiol) produced primarily by peripheral aromatization of androstenedione have a biologic impact in men and postmenopausal women.

Production of Testosterone-Binding Globulin in Polycystic Ovary Syndrome

TeBG binds both testosterone and estradiol and thus decreases the biologic activities of these critical steroids. In PCOS, there is an increase in the net production of androgen and estrogen. Increased estrogenic and androgenic effects in PCOS, however, are also due to a decrease in TeBG concentration giving rise to increased free or biologically active quantities of both estradiol and testosterone. The levels of TeBG are controlled by a balance of hormonal influences on its synthesis in the liver. Testosterone and insulin inhibit, whereas estrogen and FSH stimulate, TeBG formation. In anovulatory women with PCOS, circulating levels of TeBG are reduced approximately 50%, this may be a hepatic response to increased circulating levels of testosterone and insulin. Circulating free estradiol and testosterone levels are increased because of the significant decrease in TeBG in patients with PCOS.

Thus, three mechanisms contribute to the presence of increased quantities of biologically available estradiol in PCOS: (1) increased production of estradiol from estrone in peripheral (extraovarian) tissues giving rise to increased levels of circulating estradiol; (2) increased biologically available circulating estradiol because of decreased TeBG, and (3) local conversion of estrone to estradiol at target tissues. The last local mechanism is likely to be physiologically significant in estrogen targets such as the breast that proliferates in response to estrogen and in the central nervous system, which produces LHRH and gonadotropins under feedback regulation by estrogen.

In addition to giving rise to increased biologically available estradiol, decreased serum TeBG causes elevations in biologically available free testosterone levels. In turn, testosterone decreases serum TeBG levels, giving rise to a vicious feedback circle favoring low TeBG and high bioavailable testosterone levels.

Follicular Fate in Polycystic Ovary Syndrome

Under the influence of relatively low but constant levels of FSH, follicular growth is continuously stimulated, but not to the point of full maturation and ovulation. Despite the fact that full growth potential is not realized, the follicular life span may extend several months in the form of multiple follicular cysts. Most of these follicles in polycystic ovaries are 2 to 10 mm in diameter, whereas some can be as large as 15 mm. Hyperplastic theca cells, often luteinized in response to the high LH levels, surround these follicles. The accumulation of follicles arrested at various stages of development allows increased and relatively constant production of steroids in response to steady-state levels of gonadotropins.

These follicles are also subject to atresia and are replaced by new follicles of similar limited growth potential. A steady state of stromal cell turnover contributes to the stromal compartment of the ovary, and it is sustained by tissue derived from follicular atresia. A degenerating granulosa compartment, leaving the theca cells to contribute to the stromal compartment of the ovary, accompanies atresia. This functioning stromal tissue secretes significant amounts of androstenedione under the influence of increased LH. Androstenedione is not a biologically active steroid but acts as a double precursor converted to both estradiol and testosterone in extraovarian tissues. In turn, elevated testosterone suppresses TeBG synthesis, resulting in elevated free testosterone and free estradiol levels. The elevation in free testosterone further decreases TeBG. From the point of view of steroidogenesis and steroid action, the PCOS is the result of a complex vicious circle that includes a number of positive and negative feedback mechanisms.
These findings suggest an in vivo resistance to FSH action in PCOS, possibly related to the pathologic absence of an interaction between FSH- and IGF-related signaling pathways. Induction of ovulation in PCOS is achieved, therefore, by increasing FSH levels to overcome the block at the granulosa cell level. Two currently popular treatments, oral clomiphene citrate and injectable recombinant FSH, aim to provide increased levels of endogenous or exogenous FSH to overcome this in vivo defect. Granulosa cells develop and grow in response to FSH and produce estradiol; this increase in estradiol production precedes the resumption of ovulation. In fact, the polycystic ovary often overreacts to pharmacologic levels of FSH by the recruitment of a large number of developing follicles at once, occasionally giving rise to the ovarian hyperstimulation syndrome.

Ovarian Hyperthecosis

Ovarian hyperthecosis is a severe variant of PCOS. The term refers to significantly increased stromal tissue with luteinized theca-like cells scattered throughout large sheets of fibroblast-like cells. Both clinical and histologic findings represent an exaggerated version of PCOS. This diagnosis can be made on clinical grounds; an ovarian biopsy is not necessary except to rule out an ovarian tumor. Increased androgen production leads to the clinical picture of more intense androgenization. The higher testosterone levels may also lower LH levels by blocking estrogen action at the hypothalamic-pituitary level.

Hyperthecosis seems to be an exaggerated version of the same process that gives rise to chronic ovulation in PCOS. A correlation exists between the severity of insulin resistance and the degree of insulin resistance. And in turn, because insulin and IGF-I stimulate proliferation of thecal interstitial cells, hyperinsulinemia may be an important pathophysiologic factor in the etiology of hyperthecosis.

It is not uncommon to encounter markedly high levels of testosterone, even above 2 ng/mL, in ovarian hyperthecosis. Virilization is common. These patients usually do not ovulate in response to clomiphene or recombinant FSH. It is usually difficult to suppress testosterone production even using an LHRH agonist. Bilateral oophorectomy should be used as the last resort but, unfortunately, may be necessary to control testosterone production in a significant portion of these patients.

Increased waist-to-hip ratio compounded by significantly increased body mass index is called android obesity because this type of adipose tissue distribution is

Role of Obesity in Insulin Resistance and Anovulation

Increased body mass index is associated with increased estradiol and decreased sex hormone-binding globulin levels. In turn, the circulating estradiol level increases and androgen production is influenced by a genetic predisposition. In fact, women with hyperandrogenism, anovulation, and polycystic ovaries have a higher incidence of females with hyperinsulinemia and male relatives with baldness. Finally, familial aggregation of increased serum testosterone levels in PCOS suggests that androgen excess per se is a genetic trait. Genetic linkage studies are under way to identify individual gene defects that may be responsible for PCOS.

Insulin Resistance and Polycystic Ovary Syndrome

Insulin resistance is a major factor in the pathogenesis of noninsulin-dependent diabetes mellitus (NIDDM). The term insulin resistance can be defined as impaired whole-body insulin-mediated glucose disposal, as determined using techniques such as the hyperinsulinemic clamp technique. Insulin resistance is defined clinically as the inability of a known quantity of exogenous or endogenous insulin to increase glucose uptake and utilization in an individual as much as it does in the normal control. Insulin resistance is frequently observed in hyperandrogenic and obese women with PCOS. More severe degrees of insulin resistance or impaired glucose tolerance, however, are more common in obese women with PCOS.

The association between a disorder of carbohydrate metabolism and androgen excess was first described in 1921 by Archard and Thiers and was called the "diabetes of bearded women." Since then, the association between PCOS and insulin resistance or impaired glucose tolerance has been well recognized. This clinical association of insulin resistance and anovulatory hyperandrogenism is commonly found throughout the world and among different ethnic groups. In addition, the strong link between hyperinsulinemia and hyperandrogenism suggests that the stimulatory effect of insulin on ovarian androgen production is influenced by a genetic predisposition. In fact, women with hyperandrogenism, anovulation, and polycystic ovaries have a higher incidence of females with hyperinsulinemia and male relatives with baldness. Finally, familial aggregation of increased serum testosterone levels in PCOS suggests that androgen excess per se is a genetic trait. Genetic linkage studies are under way to identify individual gene defects that may be responsible for PCOS.

Insulin resistance is characterized by an impaired glucose response to a specific amount of insulin. In many of these patients, normal glucose levels are maintained at the expense of increased circulating insulin to overcome the underlying defect. More severe forms of insulin resistance in PCOS range from impaired glucose tolerance to frank NIDDM. Resistance to insulin-stimulated glucose uptake is a relatively common phenomenon in the general population, sometimes referred to as syndrome X. The fundamental abnormality leading to the manifestations that make up syndrome X is resistance to insulin-mediated glucose uptake in muscle and increased lipolysis giving rise to elevated circulating free fatty acid levels. These individuals also have dyslipidemia, hypertension, and increased risk of developing cardiovascular disease. Not surprisingly, the incidences of dyslipidemia and cardiovascular risk are also increased significantly in women with PCOS.

The incidence of hypertension increases significantly after the menopause in women with a history of PCOS. Thus, there is probably a significant clinical and pathologic overlap between syndrome X and PCOS. The extent of this overlap is not known at this time.

The clinical presentation of patients with insulin resistance depends on the ability of the pancreas to compensate for the target tissue resistance to insulin. During the first stages of the development of this condition, compensation is effective, and the only metabolic abnormality is hyperinsulinemia. In many patients, the beta cells of the pancreas eventually fail to meet the challenge, and declining insulin levels lead to impaired glucose tolerance and eventually frank diabetes mellitus. In fact, beta cell dysfunction is demonstrable in women with PCOS before the onset of glucose intolerance.

Studies of well-characterized causes of hyperinsulinemia and androgen excess have illuminated various mechanisms of insulin resistance. Factors such as a decrease in insulin binding related to autoantibodies to insulin receptors, postreceptor defects, and a decrease in insulin receptor sites in target tissues are all involved in insulin resistance. These rare syndromes, however, are found in an extremely small portion of women with anovulation, androgen excess, and insulin resistance, leaving the majority of PCOS patients without any demonstrable abnormalities in the number or quality of receptors or antibody formation. The exact nature of insulin resistance in the great majority of women with PCOS is not well understood.

In order to understand the molecular defect underlying insulin resistance in PCOS, Dunai and co-workers studied the differences between skin fibroblasts from women with and without PCOS with respect to insulin-dependent signal transduction. The fibroblasts of women with PCOS showed no change in insulin binding or receptor affinity. In half of the women with PCOS, however, a postreceptor defect was observed. This defect is characterized by increased basal insulin receptor serine phosphorylation and a decrease in insulin-dependent tyrosine phosphorylation of the insulin receptor. These abnormal patterns of phosphorylation of specific residues of the insulin receptor might represent a molecular mechanism responsible for the insulin resistance, anovulation, and androgen excess of PCOS. The cause of this abnormal phosphorylation pattern and consequences for insulin action are important topics for future study.

Rule of Obesity in Insulin Resistance and Anovulation

Increased waist-to-hip ratio compounded by significantly increased body mass index is called android obesity because this type of adipose tissue distribution is observed more commonly in men. Overweight women with anovulatory androgen excess commonly have this particular body fat distribution. Android obesity is the result of fat deposited in the abdominal
wall and visceral mesenteric locations. This fat is more sensitive to catecholamines, less sensitive to insulin, and more active metabolically. Android obesity is associated with insulin resistance, glucose intolerance, diabetes mellitus, and an increase in androgen production rate resulting in decreased levels of TeBG and increased levels of free testosterone and estradiol. 414 430 432 433 Not surprisingly, android obesity is associated significantly with cardiovascular risk factors, including hypertension and dyslipidemia. It is also important to emphasize that android obesity has been linked to a notable increase in the risk of breast cancer, with a poor prognosis. 432 433 No direct association, however, has been reported between PCOS and breast cancer risk. 433

Although the combination of insulin resistance and androgen excess is often observed in obese women overall, women with android-type obesity appear to be at a significantly higher risk for insulin resistance and androgen excess. However, insulin resistance and androgen excess are not confined to obese anovulatory women but also occur in nonobese anovulatory women. 414 430 432 433 Although obesity by itself causes insulin resistance, the combination of insulin resistance and androgen excess is a specific feature of PCOS. Not surprisingly, the combination of obesity and PCOS is associated with more severe degrees of insulin resistance than those found in nonobese women with PCOS. 432 433 434 Android-type obesity, in contrast to general obesity, is a much more specific risk factor for PCOS.

Diagnosis of Insulin Resistance

In everyday clinical practice, the biochemical diagnosis of insulin resistance in an individual patient has not been standardized and is extremely complex. First, a quarter of the normal population has fasting and glucose-stimulated insulin levels that overlap those of insulin-resistant individuals 434 435 because of great variability of insulin sensitivity in normal subjects. Second, clinically available measures of insulin action, such as fasting or glucose-stimulated insulin levels, do not correlate well with more detailed measurements of insulin sensitivity in research settings.

In view of these constraints, it is reasonable to consider all women with PCOS at risk for insulin resistance and the associated abnormalities of the insulin resistance syndrome (syndrome X) dyslipidemia, hypertension, and cardiovascular disease. A lipid profile should be obtained in all cases of PCOS. Especially obese women with PCOS should have fasting glucose levels and glucose levels 2 hours after a 75-g glucose load as a screen for glucose intolerance. The clinician should encourage the patient to take every possible measure (e.g., weight reduction and exercise) to reduce insulin resistance.

Use of Antidiabetic Drugs to Treat Anovulation and Androgen Excess

A logical approach to the management of PCOS includes the use of medications that improve insulin sensitivity in target tissues, thus achieving reductions in insulin secretion and stability of glucose tolerance. Oral agents employed in the treatment of diabetes mellitus, such as metformin and troglitazone, have been used to induce ovulation and decrease circulating androgen in anovulatory women with PCOS. The biguanide metformin improves insulin sensitivity, but the primary effect is a significant reduction in glucose tolerance, thus decreasing hepatic glucose production.

Metformin at a dose of 500 mg three times a day reduced hyperinsulinemia, basal and stimulated LH levels, and free testosterone concentrations in overweight women with PCOS. 414 430 436 A significant number of these anovulatory women ovulated and achieved pregnancy. 432 433 The addition of metformin to clomiphene citrate resulted in a remarkable improvement in the ovulation rate in obese women with PCOS. 432 433 The improvements in ovulation and androgen levels in this study might in part result from the weight loss that often accompanies the use of metformin. 432 433 Investigators of subsequent published studies, however, concluded that in both lean and obese anovulatory women, metformin treatment reduced hyperinsulinemia and androgen excess independent of changes in body weight. 434 435 437 Metformin thus has the added benefit of having a promising medication for the reduction of cardiovascular risk and induction of ovulation in women with PCOS. The benign side-effect profile of this medication also makes it an attractive choice.

The thiazolidinediones are pharmacologic ligands for the nuclear receptor peroxisome proliferator-activated receptor (PPAR). A member of this family, troglitazone, at a dose of 400 mg/day, markedly improves insulin action and insulin secretion through improved peripheral glucose utilization and beta cell function without weight changes. Troglitazone decreased insulin, LH, and testosterone levels and increased circulating TeBG. 432 434 435 Resumption of ovulation in obese women has been reported with the use of troglitazone. 434 435 However, troglitazone, the first thiazolidinedione approved by the Food and Drug Administration in the United States, proved to be hepatotoxic and was withdrawn from the market after the report of several dozens of deaths and cases of severe hepatic failure requiring liver transplantation. The safety and therapeutic potential of new thiazolidinediones currently are being tested in PCOS. 432

Management of Long-Term Deleterious Effects of Polycystic Ovary Syndrome

The long-term consequences of PCOS include irregular uterine bleeding, anovulatory infertility, androgen excess (hirsutism or virilization or both), chronically elevated free estrogen associated with an increased risk of endometrial cancer, and insulin resistance associated with an increased risk of cardiovascular disease and diabetes mellitus. Therefore, treatment must encompass the following: (1) in achieving a healthy lifestyle and normal body weight, protection of the endometrium from unopposed estrogen effects, and a reduction in testosterone levels.

If the patient desires pregnancy, she is a candidate for the medical induction of ovulation. When pregnancy is achieved, patients with polycystic ovaries appear to have an increased risk of spontaneous miscarriage. However, patients with PCOS who have not ovulated spontaneously have a normal spontaneous pregnancy rate. This increased risk may be related to elevated levels of LH that may produce an adverse environment for the oocyte and the endometrium. Therefore, LH levels should be suppressed with oral contraceptives before inducing ovulation. This suppression can be achieved in most patients with PCOS by the use of an oral contraceptive for 4 to 6 weeks before ovulation induction with clomiphene citrate or recombinant FSH.

If the patient does not wish to become pregnant, therapy is directed toward the interruption of unopposed effect of estrogen on the endometrium. Nonfluctuating levels of unopposed estradiol in the absence of progesterone cause irregular uterine bleeding, amenorrhea, and infertility and increase the risk of endometrial cancer. Anovulatory women with PCOS may have endometrial cancer even in their early 20s. 432 434 435 Therefore, endometrial biopsy should be performed periodically in untreated women with PCOS regardless of age. The uterine bleeding pattern should not influence the decision to perform an endometrial biopsy. The presence of amenorrhea does not rule out endometrial hyperplasia. The critical factor that determines the risk of endometrial neoplasia is the duration of anovulation and exposure to unopposed estradiol. Long-term treatment with a progestin or oral contraceptive significantly decreases the risk of endometrial cancer.

One of the simplest and most effective ways to administer a progestin in the long term is to use an oral contraceptive. Also, oral contraceptives provide two more benefits: reduction of androgen excess and contraception. Oral contraceptive pills reduce circulating androgen levels through suppression of circulating LH and stimulation of TeBG levels and have been shown to reduce hirsutism in hyperandrogenic patients. 435

A concern regarding possible insulin-desensitizing effects of oral contraceptives has been raised. 435 Older oral contraceptives cause increased insulin resistance because of the high estrogen component. 435 Long-term follow-up studies, however, have failed to detect any increase in the incidence of diabetes mellitus in past or current users of high-dose pills. 435 436 In fact, more recent studies demonstrated that new oral contraceptives induced neither change or a significant decrease in insulin resistance in women with PCOS. The oral contraceptives that did not induce insulin resistance in women with PCOS contained an ethinyl estradiol dose of 30 µg or less and a desogestrel, norgestimate, or gestodene as the progestin component. 435 436 Furthermore, past users of oral contraceptives have no increased cardiovascular risk. 435 Low-dose oral contraceptives have also been administered to women with gestational diabetes or insulin-dependent diabetes mellitus without an adverse impact. 435 436 437 438 Low-dose oral contraceptives have not increased the risk of retinopathy or nephropathy in diabetic patients, nor has there been any deterioration of lipid or biochemical markers. 435 436 437 438 It should be emphasized again that the major contributing factor to hyperinsulinemia and insulin resistance in PCOS is obesity. 435 438 In summary, oral contraceptive treatment for anovulatory and hyperinsulinemic women with androgen excess does not increase cardiovascular risk.

For the patient who does not complain of hirsutism but is anovulatory and has irregular bleeding, treatment with a single progestin may be attempted as an alternative to oral contraceptives. Progestin therapy is directed toward interruption of the chronic exposure of endometrium to unopposed effects of estrogen.
Medroxyprogesterone acetate, 10 mg daily for the first 10 days of every month, can be administered to ensure withdrawal bleeding and prevent endometrial hyperplasia. This treatment does not decrease androgen excess, nor does it provide contraception. Because new oral contraceptives (with an ethinyl estradiol content of 30 µg or less and a new progestin) suppress androgen excess of ovarian origin, provide contraception, protect the endometrium, and do not increase insulin resistance, a new low-dose oral contraceptive is the treatment of choice for nonsmokers with PCOS. An oral contraceptive together with the antiandrogen spironolactone, 100 mg/day, is the recommended starting treatment for a hirsute woman with PCOS. The dose of spironolactone can be increased in increments to suppress hair growth as previously described in this section.

Treatment with an oral contraceptive (plus or minus spironolactone) may not be effective in androgen suppression in severe cases of PCOS. In these patients resistant to oral contraceptives, suppression of the ovary with an LHRH agonist may be required. Because glucocorticoids increase insulin resistance, they should be used with caution in patients with hyper-insulinemia. Spironolactone does not affect insulin sensitivity in anovulatory women and can be used safely without causing adverse effects on carbohydrate or lipid metabolism. Therefore, long-term follow-up studies have shown a significantly increased risk for the development of frank diabetes mellitus in anovulatory patients with PCOS. It is therefore important to monitor glucose tolerance with periodic glucose levels after fasting and after a 75-g glucose load.

Because insulin resistance contributes to the abnormal lipid profile and increased cardiovascular risk in women with PCOS, weight loss is a high priority for patients who are overweight. Both insulin resistance and androgen excess can be reduced with a weight reduction of at least 5%. Significant weight loss also resulted in ovulation and pregnancy in a number of patients with PCOS. Therefore, long-term nutritional counseling and an emphasis on lifestyle changes are essential components of the long-term management of PCOS.

The place of insulin sensitizers, such as metformin and thiazolidinediones, in the treatment of PCOS remains to be determined by data from future clinical trials. When one takes into account the design of how this treatment should work, the potential for its preventive benefits is impressive. Preliminary findings suggest that thiazolidinediones are a potent group of medications that restore ovulation by reducing insulin resistance and androgen excess. Unfortunately, troglitazone, the first compound approved by the Food and Drug Administration in the United States, proved to be hepatotoxic and was withdrawn from the market after the report of a small but significant number of liver failures and deaths. The most important factor determining the extent of use of new thiazolidinediones in PCOS will be their side-effect profiles (e.g., liver toxicity and teratogenesis). By contrast, the majority of publications on metformin show no serious side effects or teratogenicity. Metformin does seem to improve metabolic abnormalities and restore ovulation. Further studies are in progress to determine the roles of these antidiabetics as therapeutic agents in PCOS.

In long-term follow-up of women with PCOS, android obesity and hyperinsulinemia persisted during the postmenopausal years. Thus, postmenopausal women who have previously been anovulatory, hyperandrogenic, and hyperinsulinemic are at risk for cardiovascular disease and diabetes mellitus. Aggressive preventive health care interventions that lower cardiovascular risk and other unfavorable consequences of PCOS are appropriate.

Genetic counseling is of paramount importance for the families of patients with PCOS. A growing body of evidence suggests that up to half of first-degree relatives and sisters may be affected by PCOS or at least by androgen excess in the presence of regular menses. These individuals may be at higher than average risk for cardiovascular disease and may benefit from preventive measures that reduce this risk.

Ovulation Induction by an Aromatase Inhibitor

Clomiphene Citrate.

To induce ovulation in PCOS, FSH levels are increased through the use of, for example, clomiphene citrate or by injection of recombinant FSH. Presumably, pharmacologic levels of FSH overcome the ovarian defect responsible for anovulation in PCOS.

Clomiphene citrate is a nonsteroidal ovulation-inducing ER ligand with mixed agonistic-antagonistic properties. Acting as an antiestrogen, clomiphene citrate is thought to displace endogenous estrogen from hypothalamic ERs, thereby removing the negative feedback effect exerted by endogenous estrogens. The resultant increase in pulsatile LHRH release is thought to normalize the release of pituitary FSH and LH, followed by follicular recruitment and selection, assertion of dominance, and, ultimately, ovulation.

Clomiphene citrate treatment can be started at any time in an amenorrheic and anovulatory patient provided that a pregnancy test is performed beforehand. Alternatively, urine

bleeding may be induced after a 21-day treatment with an oral contraceptive or 10-day treatment with medroxyprogesterone acetate (10 mg/day). Clomiphene citrate at 50 mg/day is started orally on day 3, 4, or 5 of the cycle and continued for 5 days. Over the past 20 some years, we have used a practical approach termed triple-7 in order to monitor indicators of ovulation after the administration of clomiphene citrate. The protocol is depicted in Figure 16-38. This approach is timed favorably to detect the preovulatory surge of serum estradiol 7 days after the last clomiphene citrate dose (arrow pointing up) and a serum progesterone level 14 days after the last clomiphene citrate dose (arrow pointing down) in order to document ovulation. A repeated office visit is scheduled 7 days after the progesterone determination, that is, 21 days after the last clomiphene citrate dose (see Fig. 16-38). The patient should be encouraged to have intercourse every other day during the 10-day period following the last clomiphene citrate dose. Alternatively, measurement of urinary LH to detect an LH surge can be used to time intercourse.

If ovulation does not occur after the first course of therapy with clomiphene citrate at 50 mg/day, a second course of 100 mg daily for 5 days may be started. Lack of response at doses of 150 to 200 mg daily for 5 days should be an indication for a change of treatment. Most patients desired to conceive do so with the starting dose of clomiphene citrate (50 mg/day for 5 days). Most clomiphene citrate initiated conceptions are likely to occur within the first six ovulatory cycles. The incidence rate for multiple gestation in clomiphene citrate induced pregnancies is 7.9%, of which 6.9% are twins.

Letrozole.

The aromatase inhibitor letrozole has been used as an experimental medication to induce ovulation. The mechanism of action appears to be similar to that of clomiphene citrate. Oral administration of letrozole (2.5 mg/day on days 3 to 7 after uterine bleeding) is effective for ovulation induction in anovulatory infertility.

Ovulation induction by an aromatase inhibitor is presumed to be mediated by estrogen deficiency induced at the level of the hypothalamus. Letrozole appears to avoid the unfavorable effects on the endometrium frequently seen with the use of antiestrogens (e.g., clomiphene citrate) for ovulation induction. Clinical data regarding this experimental treatment are extremely scarce at the moment.

For women who do not ovulate in response to clomiphene citrate, recombinant FSH is administered subcutaneously at a starting dose of two ampules (equivalent to 150 IU of FSH), starting on day 3 of spontaneous or progesterin-induced uterine bleeding and increasing by 75 IU at 3- to 7-day intervals until serum estradiol concentrations begin to increase. The dose is then maintained until follicular rupture, which is induced by intramuscular administration of hCG (10,000 IU). Follicular growth is monitored by transvaginal ultrasonography and blood estradiol levels, which serve as biochemical markers for the granulosa cell mass in the growing

Figure 16-38 Hormonal monitoring in clomiphene citrate initiated ovulation: use of the triple-7 regimen. Please see text. (From Adashi EY. Clomiphene citrate-initiated ovulation: a clinical update. Semin Reprod Endocrinol 1996; 4:255-271.)
Three important complications of gonadotropin therapy in PCOS are significantly increased rates of multiple pregnancies, severe ovarian hyperstimulation syndrome, and spontaneous miscarriage. Conventional-dose gonadotropin therapy causes two important complications: (1) an alarming number of multiple pregnancies (range, 14% to 50% of treatment cycles) and (2) a significantly increased risk of severe ovarian hyperstimulation syndrome (range, 1.3% to 9.4% of treatment cycles).

Conventional-dose FSH regimens for induction of ovulation for women with PCOS have succeeded in reducing the rate of multiple pregnancies to as low as 6% in some series. The low-dose regimen also practically eliminated the complication of severe ovarian hyperstimulation syndrome. This has been achieved by reaching, but not exceeding, the threshold level of FSH, starting with a daily dose of 75 IU for 14 days and using small incremental dose increases when necessary. This regimen induces the development of a single follicle in 70% of cycles. Conception rates are comparable to those achieved with conventional therapy. The miscarriage rate remains somewhat higher than that after spontaneous conception (20% to 25%). The treatment time for the low-dose regimen is significantly longer than that required for conventional gonadotropin treatment. New data obtained using recombinant FSH for a low-dose regimen, rather than urinary gonadotropins, suggest that treatment time may be shortened.

Premature Ovarian Failure

Premature ovarian failure, which is defined as early depletion of ovarian follicles before the age of 40, is a state of hypergonadotropic hypogonadism. These patients present with amenorrhea or oligomenorrhea. They go through a normal puberty and a variable period of cyclic menses followed by oligomenorrhea and amenorrhea. Therefore, premature ovarian failure should always be included in the differential diagnosis of chronic anovulation. History and physical examination may reveal menstrual irregularity or secondary amenorrhea accompanied by symptoms and signs of estrogen deficiency, such as hot flashes and urogenital atrophy. Elevated FSH levels (above the 95% confidence limits of the midcycle gonadotropin peak of the normal menstrual cycle, i.e., > 40 IU/L) on at least two occasions confirm the diagnosis.

On average, the menopause occurs at the age of 50 years, with 1% of women continuing to menstruate beyond the age of 60 years and another 1% whose menopause occurs before 40 years. Thus, premature menopause or ovarian failure has been arbitrarily defined as the cessation of menses before 40 years of age. In most cases, the etiology of premature ovarian failure is not clear. The patient may be counseled that the disorder is probably a genetic one causing ovarian follicles to disappear at a rate faster than normal. Specific sex chromosome anomalies may be identified in a subset of patients presenting with premature ovarian failure. Among these, 45,X and 47,XXY are the most common, followed by mosaicism involving various combinations.

The underlying ovarian defect may be manifest at varying ages, in the number of follicles present in the ovaries. The different symptoms may be regarded as phases in the process of perinatalgonadal change regardless of the actual age of the patient. If loss of follicles occurs rapidly before puberty, primary amenorrhea and lack of secondary sexual development ensue. The degree to which the adult phenotype develops and when the secondary amenorrhea actually occurs depend on whether follicle loss took place during or after puberty. In cases of primary amenorrhea associated with sexual infantilism, the ovarian remnants exist as streaks, and transvaginal ultrasonography usually cannot detect any ovaries.

Premature ovarian failure can be due to an autoimmune process, as this condition is frequently detected in association with autoimmune polyendocrine syndromes. Other causes of premature ovarian failure can be related to the sudden destruction of the follicles through factors such as chemotherapy, radiation, or infections such as mumps or gonorrhea. The effect of radiation is dependent upon age and the x-ray dose. Steroid levels begin to fall and gonadotropins rise within 2 weeks after radiation of the ovaries. Young women exposed to radiation are less likely to have permanent ovarian failure because of the higher number of oocytes present at younger ages. When the radiation field excludes the pelvis or the ovaries are transposed out of the pelvis by laparoscopic surgery before radiation, there is no risk of premature ovarian failure. Most chemotherapeutic agents used for the eradication of malignancies are toxic to the ovaries and cause ovarian failure. Resumption of menses and pregnancy have been reported after radiotherapy or chemotherapy. By the same token, premature ovarian failure may occur years after chemotherapy or radiotherapy.

Finally, single gene defects may give rise to ovarian failure. These include mutations of FSH and LH receptors and galactosia. Mutations of galactose-1-phosphate uridyltransferase, for example, can lead to ovarian failure because of the accumulation of galactose-1-phosphate at toxic levels.

Diagnosis and Management of Premature Ovarian Failure

Premature ovarian failure should be suspected in a woman who is younger than 40 years who presents with amenorrhea, oligomenorrhea, or another form of menstrual irregularity. Menopausal serum FSH levels (40 IU/L) on at least two occasions are sufficient for the diagnosis of premature ovarian failure. Thus, these young women can be diagnosed with ovarian failure and infertility if gonadotropin levels are repeatedly elevated. There are, however, a number of case reports of pregnancies in affected women occurring during hormone replacement therapy. A randomized trial of hormone replacement in this setting showed that folliculogenesis occurred often but was less frequently followed by ovulation and even less frequently by pregnancy (up to 14%); estrogen therapy did not improve the rate of folliculogenesis, ovulation, or pregnancy. Therefore, the clinician should inform patients diagnosed with premature ovarian failure that there is a small but significant likelihood of spontaneous pregnancy in the future. Women desirous of achieving pregnancy are still best served by assisted reproductive technology employing donor oocytes because the probability of spontaneous pregnancy is low. Use of donor oocytes followed by in vitro fertilization with the partner's sperm and intratubal embryo transfer after synchronization of the recipient patient's endometrium with the donor's cycle using exogenous estrogen and progesterone are offered to the patient who wishes to carry a pregnancy in her uterus. This approach offers an excellent chance of pregnancy (>50% per donor oocyte in vitro fertilization cycle).

Patients with premature ovarian failure are at increased risk for having an abnormal complement of chromosomes. The risk of having an abnormal karyotype increases with decreasing age of onset of the ovarian failure. A chromosomal analysis is recommended for some of these patients because of increased risk of a gonadal tumor associated with the presence of a Y chromosome. The arbitrarily chosen age group for chromosomal analysis includes women 30 years of age or younger because it is extraordinarily rare to encounter a gonadal tumor in patients with premature ovarian failure after the age of 30.

The presence of mosaicism including a Y chromosome has been associated with a high incidence of gonadal tumors. These malignant tumors arise from germ cells and include gonadoblastomas, dysgerminomas, yolk sac tumors, and choriocarcinoma. In particular, the presence of secondary virilization in these patients with karyotypic abnormalities and premature ovarian failure significantly increases the risk of a dysontogenetic gonadal tumor. The precise risk of a tumor in various subsets of these patients is not well known because a significant number of women carrying a Y chromosome do not have symptoms of virilization. The frequency of Y-chromosome material determined by polymerase chain reaction is high in Turner's syndrome (12.2%), but the occurrence of a gonadal tumor among these Y-positive patients needs to be as low as 7% to 10%.

| TABLE 16-9 – Laboratory Evaluation of Premature Ovarian Failure |
|------------------|------------------|------------------|------------------|------------------|
| Karyotype (< 30 yr of age or sexual infantilism) | Cortisol after ACTH stimulation (adrenal insufficiency) | TSH (hypothyroidism) | Glucose (fasting and 2 hr after 75-g glucose load, diabetes mellitus) |
| FSH (to establish the diagnosis of premature ovarian failure) | | | |

637
Calcium and phosphorus (hypoparathyroidism)
Sedimentation rate, complete blood count with differential, antinuclear antibody, rheumatoid factor (autoimmune disease)
Pregnenolone (to evaluate 17-hydroxylase deficiency in sexually infantile women)
Galactose-1-phosphate (galactosemia)

ACTH, adrenocorticotropic hormone; FSH, follicle-stimulating hormone; TSH, thyroid-stimulating hormone.

Premature ovarian failure may also occur as an isolated autoimmune disorder or in association with hypothyroidism, diabetes mellitus, hypoadrenalism, hypoparathyroidism, or systemic lupus erythematosus. Therefore, the tests listed in Table 16-9 should be performed every few years because premature ovarian failure can be part of an autoimmune polyendocrine syndrome. Thyroid and adrenal insufficiency and diabetes mellitus are the endocrine disorders most frequently associated with premature ovarian failure. It should be noted, however, that overall it is fairly rare to encounter any endocrine disorder associated with premature ovarian failure.

Treatment of premature ovarian failure should be directed toward its specific cause, if this is possible. In most cases, however, it is not possible to identify a specific etiology if there are no karyotypic anomalies. If the patient desires pregnancy, she should be offered ovum donation and in vitro fertilization using her partner's sperm (see Fig. 16-29). If pregnancy is not desired, she should be treated with an oral contraceptive or with estrogen and progestin replacement. Estrogen therapy promotes and maintains secondary sexual characteristics and prevents premature osteoporosis.
Differential Diagnosis and Management of Anovulatory Uterine Bleeding

Acyclic production of estrogen during anovulatory cycles gives rise to irregular shedding of the endometrium. These bleeding manifestations of anovulatory cycles in the absence of uterine pathology or systemic illness are commonly referred to as dysfunctional uterine bleeding. Anovulatory uterine bleeding is the most common cause of chronic menstrual irregularities and is a diagnosis of exclusion. Pregnancy, uterine leiomyomas, endometrial polyps, and adenomyosis should be ruled out as anatomic causes of irregular uterine bleeding. Malignancies of the vagina, cervix, endometrium, myometrium, fallopian tubes, and ovaries should also be ruled out before a diagnosis of anovulatory uterine bleeding is made. Finally, coagulation abnormalities should be excluded.

Anovulatory uterine bleeding can be managed without surgical intervention by either restoring ovulation or mimicking the ovulatory hormonal profile by providing exogenous steroids. The rationale for using exogenous steroids is based on the knowledge of predictable responses of the endometrium to estrogen and progesterone. Physiologic responses of the endometrium to natural ovarian steroids have been uncovered by observing the gross and microscopic changes in the endometrium during thousands of normal ovulatory cycles in humans and other primates. The pharmacologic application of exogenous estrogens and progestins in women with anovulatory bleeding aims to correct the production of local tissue factors, which mediate physiologic steroid action, and thus reverse the excessive and prolonged flow typical of anovulatory cycles.

Clinical management of irregular uterine bleeding with exogenous hormones is a time-honored method and is also of diagnostic value. Failure to control vaginal bleeding with hormonal therapy, despite appropriate application and utilization, makes the diagnosis of anovulatory uterine bleeding considerably less likely. In this case, attention is directed to an anatomic pathologic entity within the reproductive axis as the cause of abnormal bleeding.

Heavy but regular menstrual bleeding (hypermenorrhea) can be encountered in ovulatory women. It may be due to anatomic causes such as a leiomyoma impinging on the endometrial cavity or the diffuse and pathologic presence of benign endometrial glands in the myometrium (adenomyosis). In the absence of a specific pathologic cause, however, it is presumed that hypermenorrhea reflects subtle disturbances in the endometrial tissue mechanism. In essentially all cases, evaluation and treatment are identical to the approach detailed in this section.

Characteristics of Normal Menstrual Periods

Normal menstruation takes place about 14 days after each ovulation episode as a consequence of postovulatory estrogen-progesterone withdrawal. The quantity and duration of bleeding are quite reproducible. This predictability leads many women to expect a certain characteristic flow pattern. Any slight deviations, such as plus or minus 1 day in duration or minor deviation from expected tampon utilization, are causes for major concern in the patient. Most women of reproductive age can predict the timing of their flows so accurately that even some instances of minor variability may require reassurance by the clinician. Although variability of menstrual cycles is a common feature during teenage years and the perimenopausal transition, the characteristics of menstrual bleeding do not undergo appreciable change between ages 20 and 40.

For ovulatory women, the changes in the length of menstrual cycles over the period of reproductive age are predictable. Between menarche and age 20, the cycle length for most ovulatory women is relatively longer. Between 20 and 40, there is increased regularity as cycles shorten. In the 40s, cycles begin to lengthen again. The highest incidence of anovulatory cycles occurs before age 20 and after age 40. In this age group, the average length of a cycle is between 25 and 28 days. Among ovulatory women, the frequency of a cycle less than 21 days long or a cycle greater than 35 days is extremely rare (less than 2%). Overall, most women have cycles that last from 24 to 35 days. Between ages 40 and 50, menstrual cycle length increases and anovulation becomes more prevalent. The average postovulatory bleeding lasts from 4 to 6 days. The normal volume of menstrual blood loss is 30 mL. More than 80 mL is considered abnormal. Most of the blood loss occurs during the first 3 days of a period, so excessive flow may exist without prolongation of flow.

During an ovulatory cycle, the duration from the ovulation to menses is relatively constant and averages 14 days. Greater variability in the length of proliferative phase, however, produces a distribution in the duration of a menstrual cycle. Menstrual bleeding more often than every 24 days or less often than every 35 days requires evaluation. Flow that lasts 7 or more days also requires evaluation. A flow that totals more than 80 mL per month usually leads to anemia and should be treated. In clinical practice, however, it is quite difficult to quantify menstrual flow because evaluation and treatment are based solely on the patient's perceptions regarding the duration, amount, and timing of her menstrual bleeding. Despite this difficulty in quantifying menstrual blood loss, the clinician should evaluate the cause of excessive uterine bleeding. Anemia should be ruled out by a complete blood count. A low hemoglobin value accompanied by microcytic and hypochromic red blood cells suggests excessive blood loss during menses. These patients should be provided with iron supplementation. The likely presence of coagulation defects, uterine leiomyomas, or adenomyosis underlying prolonged menses should also be evaluated in anemic patients through a meticulous history and physical examination followed by relevant laboratory tests.

Termination of Menstruation

Oligomenorrhea is defined as intervals between episodes of uterine bleeding greater than 35 days, and the term polymenorrhea is used to describe intervals less than 24 days. Hypomenorrhea refers to regular intervals (24 to 35 days) but excessive flow or duration of bleeding, or both. Hypomenorrhea refers to diminution of the flow or shortening of the duration of regular menses, or both.

Uterine Bleeding in Response to Steroid Hormones

Exogenous Estrogens

Uterine bleeding follows acute cessation of estrogen support to the endometrium. Thus, this type of uterine bleeding can occur after bilateral oophorectomy, radiation of the ovaries, or cessation of exogenous estrogen therapy. Similarly, the bleeding that occurs between cycles can be delayed by concomitant estrogen therapy. Flow occurs on discontinuation of exogenous estrogen. Thus, estrogen withdrawal by itself (in the absence of progesterone) almost invariably causes uterine bleeding.

Exogenous Progestins

Chronic exposure to varying quantities of estrogen stimulates the growth of endometrium continuously in the absence of progesterone, as in the case of excessive

639

Figure 16-39. Variation of the duration of the menstrual cycle in women with regular cycles. (From Cunningham FG, MacDonald PC, Gant NF, et al. The endometrium and decidua: menstruation and pregnancy. In Williams Obstetrics, 19th ed. Stamford, Conn, Appleton & Lange, 1993, pp 81109.)

639

In clinical practice, however, it is quite difficult to quantify menstrual flow because evaluation and treatment are based solely on the patient's perceptions regarding the duration, amount, and timing of her menstrual bleeding. Despite this difficulty in quantifying menstrual blood loss, the clinician should evaluate the cause of excessive uterine bleeding. Anemia should be ruled out by a complete blood count. A low hemoglobin value accompanied by microcytic and hypochromic red blood cells suggests excessive blood loss during menses. These patients should be provided with iron supplementation. The likely presence of coagulation defects, uterine leiomyomas, or adenomyosis underlying prolonged menses should also be evaluated in anemic patients through a meticulous history and physical examination followed by relevant laboratory tests.

Terminology Describing Abnormal Uterine Bleeding

Oligomenorrhea is defined as intervals between episodes of uterine bleeding greater than 35 days, and the term polymenorrhea is used to describe intervals less than 24 days. Hypomenorrhea refers to regular intervals (24 to 35 days) but excessive flow or duration of bleeding, or both. Hypomenorrhea refers to diminution of the flow or shortening of the duration of regular menses, or both.
extragonadal estrogen production in PCOS. After a certain point, the amount of estrogen produced in extraovarian tissue remains insufficient to maintain structural support for the endometrium. This gives rise to unpredictable episodes of shedding of the surface endometrium. Relatively low doses of estrogen yield intermittent spotting that may be prolonged but is generally light in quantity of flow. On the other hand, high levels of estrogen and sustained availability lead to prolonged periods of amenorrhea followed by acute, often profuse episodes of bleeding with excessive loss of blood.

**Progesterone Withdrawal Bleeding.**

The typical progesterone withdrawal bleeding occurs after ovulation in the absence of pregnancy. Removal of the corpus luteum is another example that leads to endometrial desquamation. Pharmacologically, a similar event can be achieved by administration and discontinuation of progesterone or a synthetic progestin. Progesterone withdrawal bleeding occurs only if the endometrium is initially primed by endogenous or exogenous estrogen. If estrogen therapy is continued as progesterone is withdrawn, the progesterone withdrawal bleeding still occurs. Only if estrogen levels are increased markedly is progesterone withdrawal bleeding delayed. Thus, progesterone withdrawal bleeding is quite predictable in the presence of previous or concomitant estrogen exposure.

**Progestin Breakthrough Bleeding.**

This is a pharmacologic phenomenon that occurs in the presence of an unfavorably high ratio of progestin to estrogen. In the absence of sufficient estrogen, continuous progestin therapy leads to intermittent bleeding of variable duration, similar to the low-dose estrogen breakthrough bleeding noted previously. This type of bleeding is associated with the combination oral contraceptives that contain low-dose estrogen and the long-acting progestin-only contraceptive methods such as Norplant and Depo-Provera. Progestin breakthrough bleeding is highly unpredictable and characterized by extensive variability between women.

**Causes of Irregular Uterine Bleeding**

Pregnancy and its complications represent one of the most common causes of irregular uterine bleeding (Table 16-10). Pregnancy should be ruled out by a urine test in any woman of reproductive age presenting with irregular bleeding (Table 16-11).

As pointed out earlier, anovulatory uterine bleeding arising from responses of the endometrium to inappropriate production of ovarian steroids has also been called dysfunctional uterine bleeding because treatments that restore ovulatory function potentially reverse the irregular bleeding pattern. Common examples of anovulatory bleeding include those associated with exercise-related anovulation, hyperprolactinemia, hypothroidism, or PCOS. In these cases, either restoring ovulatory menses by correction of the underlying disorder or use of exogenous hormones can achieve predictable uterine bleeding. On the other hand, various pathologic entities of the genital tract (ovaries, uterus, vagina, or vulva) or a coagulation abnormality may also cause deviation from normal menses (see Table 16-10).

Anovulatory uterine bleeding is a diagnosis of exclusion for the following reasons. Vulvar, vaginal, or uterine malignancies can give rise to irregular bleeding. Moreover, an estrogen- or androgen-secreting ovarian tumor may cause abnormal uterine bleeding (see Table 16-10). Pregnancy and pregnancy-related problems such as ectopic pregnancy or spontaneous miscarriage are extremely common causes of abnormal uterine bleeding. In fact, the most common cause of disruption of a normal menstrual pattern is pregnancy or a complication of pregnancy. Another

<table>
<thead>
<tr>
<th>Complications of Pregnancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Threatened miscarriage</td>
</tr>
<tr>
<td>Incomplete miscarriage</td>
</tr>
<tr>
<td>Ectopic pregnancy</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Anovulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physiologic</td>
</tr>
<tr>
<td>Uncomplicated pregnancy (amenorrhea)</td>
</tr>
<tr>
<td>Pubertal (postmenarcheal) anovulation</td>
</tr>
<tr>
<td>Premenopausal anovulation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hypothalamic (frequently presents as amenorrhea)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Functional (e.g., diet, exercise, stress)</td>
</tr>
<tr>
<td>Anatomic (e.g., tumor, granulomatous disease, infection)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mediations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hyperprolactinemia, other pituitary disorders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prolactinoma</td>
</tr>
<tr>
<td>Other pituitary tumors, granulomatous disease</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hypothyroidism</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Medications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Androgen excess</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCOS, hyperthecosis</td>
</tr>
<tr>
<td>Ovarian tumor (e.g., Sertoli-Leydig cell tumor)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nonclassical adrenal hyperplasia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cushing's syndrome</td>
</tr>
<tr>
<td>Glucocorticoid resistance</td>
</tr>
</tbody>
</table>

| Adrenal tumor (e.g., adenoma, carcinoma) |

| Medications (e.g., testosterone, danazol) |
| Other |

| Premature ovarian failure (frequently presents as amenorrhea) |
| Chronic illness |
| Liver failure |
Renal failure  
AIDS  
Other

Anatomic Defects Affecting the Uterus

Uterine leiomyomas
Endometrial polyps
Adenomyosis (usually presents as hypermenorrhea)
Intrauterine adhesions (usually presents as amenorrhea)
Endometritis
Endometrial hyperplasia, cancer
Chronic estrogen exposure (e.g., PCOS, medication, liver failure)
Estrogen-secreting ovarian tumor (e.g., granulosa cell tumor)
Advanced cervical cancer

Coagulation Defects (Usually Present as Hypermenorrhea)
Von Willebrand's disease
Factor XI deficiency

Extrauterine Genital Bleeding (May Mimic Uterine Bleeding)
Vaginitis
Genital trauma
Foreign body
Vaginal neoplasia
Vulvar neoplasia

AIDS, acquired immunodeficiency syndrome; LHRH, luteinizing hormone-releasing hormone; PCOS, polycystic ovary syndrome.

Commonly Used Tests

Urine hCG test
Serum hCG level (incomplete miscarriage, ectopic pregnancy)
Transvaginal pelvic ultrasonography (intrauterine or ectopic pregnancy, uterine leiomyoma, endometrial polyp or neoplasia, ovarian tumor)
Serum FSH, LH (anovulation; ovarian failure)
Serum prolactin, TSH (anovulation; hyperprolactinemia)
Complete blood count, PT, PTT (coagulation defect)
Liver and renal functions, HIV (anovulation; chronic disease)
Endometrial biopsy (endometrial disease; polyp, neoplasia, endometritis)

Less Commonly Used Tests

Evaluation for PCOS, ovarian or adrenal tumor, nonclassical adrenal hyperplasia, Cushing's syndrome and glucocorticoid resistance (androgen excess)
Head CT or MRI scan (hypothalamic anovulation, hyperprolactinemia)
Pelvic MRI scan (adenomyosis, uterine leiomyoma)
Hysterosonography with intrauterine saline installation (endometrial polyp, uterine leiomyoma)
Hysteroscopy (endometrial polyp, uterine leiomyoma)
Dilatation and curettage (endometrial disease not diagnosed by ultrasonography or biopsy)

CT, computed tomography; FSH, follicle-stimulating hormone; hCG, human chorionic gonadotropin; HIV, human immunodeficiency virus; LH, luteinizing hormone; MRI, magnetic resonance imaging; PCOS, polycystic ovary syndrome; PT, prothrombin time; PTT, partial thromboplastin time; TSH, thyroid-stimulating hormone.

common cause of irregular uterine bleeding is observed in oral contraceptive users in the form of progestin breakthrough bleeding. Progestin breakthrough bleeding during postmenopausal hormone replacement is also common (see later). Patients may be using other hormonal medications unknowingly with an impact on the endometrium. For example, the use of ginseng, an herbal root, has been associated with estrogenic activity and abnormal bleeding. Although uterine bleeding is a common benign side effect of various long-term hormonal treatments, the clinician should always be convinced that no other pathology is present. Anatomically demonstrable pathologies of the menstrual outflow tract include endometrial hyperplasia and cancer, endometrial polyps, leiomyomata uteri, adenomyosis, and endometritis. Irregular, serious bleeding may also be associated with chronic illness, such as renal failure, liver failure, and acquired immunodeficiency syndrome. Finally, careful examination is worthwhile to discover genital injury or a foreign object. At puberty, the most common cause of irregular uterine bleeding is anovulation. Approximately 20% of these adolescents with excessive irregular uterine bleeding, however, have a coagulation defect. Among all women of reproductive age with hypermenorrhea, the prevalence of a coagulation disorder was reported to be 17%. Von Willebrand's disease was the most common defect, and factor XI deficiency was the second common diagnosis. Bleeding secondary to a coagulation defect is usually a heavy flow with regular, cyclic menses (hypermenorrhea), and the same pattern can be seen in patients being treated with anticoagulants. Bleeding disorders are usually associated with hypermenorrhea since menarche and a history of bleeding with surgery or trauma. Hypermenorrhea may be the only sign of an inherited bleeding disorder.

Early pregnancy or its complications should always be ruled out first by a sensitive urine hCG measurement in any reproductive-age women presenting with irregular bleeding. Threatened or incomplete miscarriage and ectopic pregnancy are extremely common causes of irregular uterine bleeding. Other tests should be ordered if necessary on the basis of the initial clinical evaluation. These include tests to evaluate anovulatory disorders of various etiologies in patients with a history of prolonged heavy menses (hypermenorrhea) of pubertal origin, coagulation studies (e.g., prothrombin time, partial thromboplastin time, and bleeding time) and a complete blood count should be obtained.
Pelvic ultrasonography through a vaginal probe is an extremely useful test for the evaluation of normal or abnormal pregnancy, uterine leiomyomas, endometrial neoplasia, and ovarian tumors (see Table 16-11). Other imaging studies may be used judiciously to rule out pathologies of the hypothalamic, pituitary, or adrenal (see earlier). Of note is the use of pelvic MRI to rule out adenomyosis, a uterine disorder characterized by the abnormal presence of diffuse endometrial tissue in the myometrial layer (see Table 16-11). Advanced adenomyosis is associated with diffuse enlargement of the uterus, hypermenorrhea, and anemia.

Endometrial histology should be determined by an endometrial biopsy performed in the physician’s office in patients at risk for the development of endometrial hyperplasia or cancer (e.g., PCOS, liver failure, obesity, diabetes mellitus, hormone replacement). A benign endometrial polyp or a uterine leiomyoma protruding into the uterine cavity can be diagnosed by hysterosonography using intrauterine saline installation or hysteroscopy. Hysterosonography and hysteroscopy are not appropriate tests to evaluate endometrial hyperplasia or cancer because these procedures may cause dissemination of malignant cells. If malignancy is suspected, it should be ruled out by an office endometrial biopsy (see Table 16-11). Occasionally, an office endometrial biopsy cannot be performed or is not diagnostic of endometrial neoplasia. In these rare instances, endometrial curettage under anesthesia is performed for a reliable tissue diagnosis.

We should underscore again that a careful history and physical examination eliminate the need for most of these diagnostic tests. A useful question to ask oneself before ordering a certain diagnostic study is whether that particular test will alter the ultimate clinical management.

Management of Anovulatory Uterine Bleeding

The terms dysfunctional uterine bleeding and anovulatory bleeding are used interchangeably and denote inappropriate stimulation of the endometrium during dysfunctional states of the reproductive system. If ovulatory function can be restored, anovulatory bleeding usually gives way to cyclic predictable periods. Because restoring ovulatory function may not be possible or practical in a large number of these women, exogenous estrogen and progestin are administered for a number of purposes. The indications for hormonal treatment of uterine bleeding include the need to stop acute uterine bleeding, to maintain predictable bleeding episodes, or to prevent endometrial hyperplasia. A number of hormonal treatments are used to stop anovulatory uterine bleeding and to induce predictable bleeding episodes. We should reemphasize that anovulatory uterine bleeding is a diagnosis of exclusion. Various anatomically demonstrable pathologies of the genital tract as listed in Table 16-10 should be ruled out before administration of the following regimens.

Oral Contraceptives

Use of combination oral contraceptives in an acute or chronic fashion is the most common treatment for irregular uterine bleeding. The estrogen component of the combination pill stabilizes the endometrial tissue and stops shedding within hours and decreases ovarian secretion of sex steroids by suppression of gonadotropins within several days. The progestin component of the pill directly affects endometrial tissue to decrease shedding over days and potentiates ovarian suppression induced by estrogen. The progestin (in the presence of estrogen) induces differentiation of the endometrial tissue into a stable form termed pseudodecidual. Typically, a monophasic oral contraceptive preparation that contains 30 or 35 μg of ethinyl estradiol is preferred. Triphasic oral contraceptives or those with less than 30 μg of ethinyl estradiol are not suitable for the treatment of excessive anovulatory uterine bleeding. A combination oral contraceptive in high doses (two or three pills a day) can be used for short intervals (weeks) to treat an acute episode of excessive uterine bleeding. A usual dose (one pill per day) may be administered for years to manage chronic anovulatory bleeding associated with PCOS or hyperprolactinemia.

Oral Contraceptives and Acute Excessive Uterine Bleeding Associated with Amenorrhea

Unopposed estrogen exposure in women with anovulatory uterine bleeding is commonly associated with chronic endometrial buildup and heavy bleeding episodes. Therapy is administered as one pill twice a day for 1 week. In obese women, the oral contraceptive may be given three times a day. This therapy is maintained despite cessation of flow within 2 days. If flow does not abate, other diagnostic possibilities (polyps, incomplete abortion, and neoplasia) should be reevaluated. In case of anovulatory bleeding, the flow does diminish rapidly within 2 days after the beginning of high-dose (one pill two or three times a day) oral contraceptive treatment. Specific causes of anovulation and possible coagulation disorders are evaluated during the following few days. At this time, the physician also considers whether blood replacement or initiation of iron therapy is necessary. The hypothalamic-ovarian-progestin combination has produced the structural rigidity intrinsic to the compact pseudodecidual reaction for the moment. Continued random breakdown of formerly fragile tissue is avoided and blood loss stopped. A large quantity of tissue, however, remains to react to estrogen-progestin withdrawal. The patient must be warned to anticipate a heavy flow with severely cramping flow a few days after stopping this therapy. The patient should also be warned of possible nausea that may be caused by high-dose oral contraceptive treatment.

At the end of a week of high-dose oral contraceptive treatment, the pill is stopped temporarily. A heavy flow usually starts within a few days. On the third day of this withdrawal bleeding, a regular dose of combination oral contraceptive medication (one pill a day) is started. This is repeated for several 3-week treatments interrupted by 1-week withdrawal intervals. A decrease in volume with each successive cycle is expected. Oral contraceptives reduce menstrual flow by more than half in most women. Early application of the estrogen-progestin combination limits growth and allows orderly regression of excessive endometrial height to normal levels. Because oral contraceptives do not treat the underlying cause of anovulation but provide symptomatic relief by directly affecting the endometrium, cessation of oral contraceptives results in the return of erratic uterine bleeding. Regardless of the requirement for contraception, oral contraceptives represent the best choice for hormonal management of heavy anovulatory bleeding and should be offered as long-term management.

Oral Contraceptives and Chronic Irregular Uterine Bleeding

PCOS is a common form of anovulation associated with chronic steady-state levels of unopposed estrogen that may give rise to endometrial hyperplasia and cancer (see earlier). Hypothalamic anovulation and hyperprolactinemia, on the other hand, are associated with low estrogen levels, insufficient to prevent bone loss. A combination oral contraceptive is a suitable long-term treatment for both forms of chronic anovulation.

Oral contraceptives represent the most suitable long-term symptomatic management option for any kind of anovulatory uterine bleeding, including oligomenorrhea. Before the administration of an oral contraceptive, pregnancy should be ruled out. For this purpose, one pill per day is ordinarily administered for 3-week periods interrupted by 1-week hormone-free intervals. Withdrawal bleeding is expected during the hormone-free interval. The progestin component serves to prevent endometrial proliferation associated with withdrawal of estrogen. In patients with PCOS (see earlier). In cases of anovulation associated with hypothalamic amenorrhea (e.g., hypothalamic anovulation, hyperprolactinemia), on the other hand, the estrogen component of the pill provides sufficient replacement to prevent bone loss. The risk of thromboembolism, stroke, or myocardial infarction associated with long-term administration is extremely low in current nonsmokers and in the absence of a history of thromboembolism. Provided that an oral contraceptive controls the abnormal uterine bleeding effectively, a chronically anovulatory woman can continue this regimen until the menopause.

Synthetic Progestins

Synthetic progestins enhance endometrial differentiation and antagonize proliferative effects of estrogen on the endometrium (see Fig. 16-28). The effects of progestins or natural progesterone include limitation of estrogen-induced endometrial growth and prevention of endometrial hyperplasia. The absence of naturally synthesized prostegnerone in anovulatory states is the rationale for administering a progestin.

The most common indication for long-term cyclic progestin administration is to prevent endometrial malignancy in a patient with PCOS and unopposed chronic estrogen exposure of the endometrium. A combination oral contraceptive is the treatment of choice in these cases. If the patient cannot use an oral contraceptive for some reason (e.g., history of thromboembolism), a progestin can be administered in a cyclic fashion to prevent endometrial hyperplasia. Before the administration of a progestin (or oral contraceptive), pregnancy should be ruled out. In the treatment of oligomenorrhea associated with PCOS, orderly limited withdrawal bleeding can be accomplished by administration of a progestin such as medroxyprogesterone acetate, 10 mg/day for at least 10 days every 2 months. Alternatively, norethindrone...
Acetate at 5 mg/day or megestrol acetate at 20 mg/day may be administered for 10 days every 2 months. Absence of withdrawal bleeding requires further work-up.

In the treatment of excessive uterine bleeding (hypermenorrhea or polymenorrhea), these progestins at higher daily doses (medroxyprogesterone acetate 20 mg/day, norethindrone acetate 10 mg/day, or megestrol acetate 40 mg/day) are prescribed for 2 weeks to induce predecidual stromal changes in the endometrium. A heavy progestin withdrawal flow usually follows within 3 days after the last dose. Thereafter, repeated progestin treatment (medroxyprogesterone acetate 10 mg/day, norethindrone acetate 5 mg/day, or megestrol acetate 20 mg/day) is offered cyclically for at least the first 10 days of every other month to ensure therapeutic effect. Failure of progestin to correct irregular bleeding requires diagnostic reevaluation such as endometrial biopsy. On the other hand, predictable withdrawal bleeding within several days after each cycle of progestin administration suggests the absence of endometrial malignancy.

**High-Dose Estrogen for Acute Excessive Uterine Bleeding**

As already outlined, an oral contraceptive given two or three times a day is the treatment of choice to stop heavy anovulatory bleeding. A high-dose oral contraceptive regimen should be offered to women with heavy uterine bleeding plus or minus asymptomatic anemia after an anatomically demonstrable pathology of the genital tract has been ruled out (see Table 16-10). On the other hand, a patient with acute and severe anovulatory bleeding accompanied by symptomatic anemia represents a medical emergency. These patients should be hospitalized immediately and offered a blood transfusion. When genital tract pathology has been ruled by history, physical examination, and pelvic ultrasonography, intravenously administered high-dose estrogen is the treatment of choice to stop life-threatening bleeding. A well-established regimen is 25 mg of conjugated estrogen administered intravenously every 4 hours until bleeding markedly slows down or for at least 24 hours. Estrogen most likely acts on the capillaries to induce clotting. Before intravenous estrogen treatment is discontinued, an oral contraceptive pill is started three times a day. Oral contraceptive treatment is continued as described previously.

Because high-dose estrogen is a risk factor for thromboembolism, taking two or three oral contraceptives per day for a week or large doses of intravenous conjugated estrogens for 24 hours should also be regarded as significant risks. There are no data available, however, to evaluate any risk associated with this type of acute use of hormonal therapy for such short intervals. The physician and patient should make a decision regarding high-dose hormone therapy after considering its risks and benefits. Alternative treatment options may be offered to patients with significant risk factors. In women with a past episode of idiopathic venous thromboembolism or a strong family history, exposure to high doses of estrogen should be avoided. High-dose hormone treatment should also be avoided in women with severe chronic illness such as liver failure or renal failure. One alternative for these patients is dilatation and curettage, followed by an oral contraceptive at one pill per day until the uterine bleeding is under control.

**Luteinizing Hormone Releasing Hormone Analogues for Excessive Anovulatory Uterine Bleeding**

LHRH analogues (see Table 16-1) may be given to women with excessive anovulatory bleeding or hypermenorrhea related to severe chronic illness such as liver failure or coagulation disorders. It should be pointed out that monthly depot injections of LHRH agonists are not effective for acute excessive uterine bleeding and may increase uterine bleeding for the first 2 weeks. LHRH antagonists, on the other hand, down-regulate FSH and LH without a delay and achieve amenorrhea more rapidly. Depot formulations of LHRH antagonists, however, are not available at present. Long-term side effects of LHRH analogues including osteoporosis make this an undesirable choice for long-term therapy. If long-term treatment with LHRH analogues is chosen, a combination of 0.625 mg of conjugated estrogens and 2.5 mg of medroxyprogesterone acetate daily should be added back when excessive anovulatory bleeding is controlled. This add-back regimen is usually sufficient to prevent osteoporosis and does not ordinarily worsen the uterine bleeding.
Hormone-Dependent Benign Gynecologic Disorders

Endometriosis

Endometriosis is defined as the presence of endometrium-like tissue outside the uterine cavity, most often on the peritoneal surfaces of the pelvis and the ovaries. It is one of the most common causes of infertility and chronic pelvic pain and affects 1 in 10 women in the reproductive age group. 

Reliable diagnosis of endometriosis can be made only by direct visualization of these peritoneal lesions by laparoscopy or laparotomy. As in other common chronic diseases such as diabetes mellitus and asthma, endometriosis is inherited in a polygenic manner. Relatives of women with this disease have a sevenfold increase in the incidence of endometriosis compared with relatives of control subjects. Sampson proposed the most widely accepted mechanism for the development of endometriosis on pelvic peritoneal surfaces as the implantation of endometrial tissue on the peritoneum through retrograde menstruation. Because retrograde menstruation occurs in more than 90% of all women, endometriosis may be caused by genetic defects that favor survival and establishment of endometrial tissue in menstrual debris on the peritoneum. 

Endometriosis and normal endometrial tissues respond to estrogen and progesterone with similar histologic changes. Estrogen favors the growth of endometriosis, whereas progesterone may limit this mitogenic action of estrogen. Some endometriotic implants undergo atrophy in response to prolonged oral contraceptive therapy just as the normal endometrium does, the so-called pregnancy state. Yet, endometriotic tissue does not respond to progestins or native progesterone as predictably as normal endometrium does. Endometriotic tissue in ectopic locations such as the peritoneum or ovary is strikingly different from the eutopic endometrium within the uterus with respect to production of cytokines and prostaglandins, steroid biosynthesis and metabolism, steroid receptor content, and clinical response to progestins.

Although current hormonal therapy for infertility associated with endometriosis is not of proven value, it is somewhat successful for pelvic pain associated with endometriosis. The duration of relief provided by medical (hormonal) treatment, however, is relatively short. Various agents used are comparable in terms of efficacy. Most current medical treatments were designed to decrease estrogen secretion by the ovaries (e.g., LHRH agonists, oral contraceptives, danazol and progestins) or to antagonize the effects of estrogen on endometrial implants (e.g., oral contraceptives, danazol and progestins). A possible alternative mechanism of action of the androgenic steroid danazol or a progestin is a direct antiproliferative effect on endometriotic tissue.

Many patients and physicians do not favor danazol because of its anabolic and androgenic side effects of weight gain and muscle cramps and occasional irreversible virilization (e.g., clitoromegaly and voice changes). In fact, up to 50% of patients with endometriosis fail to complete 6 months of treatment with danazol. The rest of the hormonal agents, contraceptives, progestins, and LHRH agonists show comparable efficacy for the control of endometriosis-associated pain. A 6-month course using any one of these agents results in a significant reduction of pain in more than 50% of patients. Induction of pain relief with a continuously administered oral contraceptive or progestin takes longer than with an LHRH agonist. There is, however, a high incidence of persistence of the disease after all of these medical therapies. Six months after completion of a 6-month course of treatment with a progestin, oral contraceptive, or LHRH agonist, moderate to severe pain symptoms recurred in 50% of initial responders. The recurrence rate of pain in the rest of the patients was approximately 5% to 20% per year during a 5-year follow-up. A 6-month course of LHRH agonist treatment is currently the most popular regimen. The most serious side effect of the LHRH agonist treatment for endometriosis is bone loss related to estrogen deficiency, and oral estrogen-progesterin preparations or bisphosphonates are usually added back to minimize bone loss.

We are still far from the cure of endometriosis, and current treatments are not satisfactory for effective control of pain. The radical treatment is the removal of both ovaries, and even this was not found to be effective in a number of cases of postmenopausal endometriosis. New strategies are needed to offer women with endometriosis a reasonable chance to live without suffering from chronic pelvic pain for decades. There are two important caveats, which are not addressed by the LHRH agonist treatment. First, large quantities of estrogen can be produced locally within the endometriotic cells. This represents an intracellular mechanism of estrogen action, in contrast to ovarian secretion, which is an endocrine means of supplying this steroid to target tissues. Second, estradiol produced in peripheral tissue sites (e.g., adipose tissue and skin fibroblasts) may give rise to pathologically significant circulating levels of estradiol in a subset of women. LHRH agonists do not inhibit peripheral estrogen formation or local estrogen production within the estrogen-responsive lesion. As a further twist, endometriosis is resistant to selective effects of progesterone and currently used progestins. Thus, aromatase inhibitors and selective progesterone response modulators are candidate therapeutic agents for endometriosis. Preliminary evidence suggests that unusually aggressive endometriotic lesions resistant to other therapy can be treated successfully with aromatase inhibitors.

Uterine Leiomyomas

Uterine leiomyomas originate from the myometrium and are the most common solid tumor of the pelvis. Leiomyomas are responsible for over 200,000 hysterectomies per year in the United States. They are almost invariably benign and represent clonal expansion of individual myometrial cells. Leiomyomas can cause a variety of symptoms including irregular and excessive uterine bleeding, pressure sensation in the lower abdomen, pain during intercourse, pelvic pain, recurrent pregnancy loss, infertility, and compression of adjacent pelvic organs, or they may be totally asymptomatic. The prevalence rate of uterine leiomyomas is estimated to be 25% to 30%. Uterine leiomyomas are more common in black women and have a polygenic inheritance pattern. Diagnosis can be made by abdominal and transvaginal ultrasonography. Transvaginal ultrasonography is a sensitive method for determining the size, number, and location of uterine leiomyomas.

Uterine leiomyomas appear during the reproductive years and regress after menopause, indicating their ovarian steroid-dependent growth potential. The role of steroids or other growth factors in the initiation and growth of these tumors, however, is not well understood. The neoplastic transformation of myometrium to leiomyoma probably involves somatic mutations of normal myometrium and the complex interactions of sex steroids and growth factors. Traditionally, estrogen has been considered the major promoter of myoma growth. More recent biochemical, histologic, and clinical evidence suggests an important role for progesterone in the growth of uterine leiomyomas. Biochemical and clinical studies suggest that progesterone and progestins, acting through the PR, might enhance proliferative activity in leiomyomas.

The therapeutic choices depend on the goals of therapy, with hysterectomy most often used for definitive treatment and myomectomy when preservation of childbearing is desired. Intradecidual and submucous leiomyomas can be removed by hysteroscopic resection. Laparoscopic myomectomy is now technically possible but apparently involves an increased risk of uterine rupture during pregnancy. The overall recurrence rate after myomectomy varies widely (10% to 50%). Although LHRH agonist-induced hypogonadism can reduce the overall volume of the uterus containing leiomyomas and tumor vascularity, the severe side effects and prompt recurrences make LHRH agonists useful only for short-term goals such as reducing anemia related to uterine bleeding or decreased tumor vascularity before hysteroscopic resection.
MANAGEMENT OF THE MENOPAUSE

Consequences of the Menopause

The Climacteric

The menopause is the permanent cessation of menses as a result of the irreversible loss of a number of ovarian functions, including ovulation and estrogen production. The climacteric is a critical period of life during which striking endocrinologic, somatic, and psychologic alterations occur in the transition to the menopause. The climacteric is also referred to as the perimenopause. The climacteric encompasses the change from ovulatory cycles to cessation of menses and is marked by irregularity of menstrual bleeding.

The most sensitive clinical indication of the climacteric is the progressively increasing occurrence of menstrual irregularities. The menstrual cycle for most ovulatory women lasts from 24 to 35 days, whereas approximately 20% of all reproductive-age women experience irregular cycles. When women are in their 40s, anovulation becomes more prevalent; prior to anovulation, the menstrual cycle length increases, beginning several years before menopause. The median age for the onset of the climacteric transition is 47.5 years. Regardless of the age of its onset, the menopause (cessation of menses) is consistently preceded by a period of prolonged cycle intervals. Elevated circulating FSH marks this menstrual cycle change before menopause and is accompanied by decreased inhibin levels, normal levels of LH, and slightly elevated levels of estradiol. These changes in serum hormone levels reflect a decreasing ovarian follicular reserve and can be detected more reliably on day 2 or 3 of the menstrual cycle.

During the climacteric, serum estradiol levels do not begin to decline until less than a year before menopause. The average circulating estradiol levels in perimenopausal women are estimated to be somewhat higher than those in younger women because of an increased follicular response to elevated FSH. The decline in inhibin production by the follicle, allowing a rise in FSH, in the later reproductive years reflects diminishing follicular reserve and competence. Ovarian follicular output of inhibin begins to decrease after 30 years of age, and this decline becomes much more pronounced after age 40. These hormonal changes are parallel to a significant decrease in fecundity, which starts at around 35 years.

The climacteric is a transitional period during which postmenopausal levels of FSH can be observed despite continued menses, whereas LH levels still remain in the normal range. The perimenopausal woman is not beyond the realm of an unexpected pregnancy because there is occasional ovulation and functional corpus luteum formation. Thus, until complete cessation of menses is observed or FSH levels higher than 40 IU/L are measured on two separate occasions, some form of contraception should be recommended to prevent unwanted pregnancies.

The climacteric represents an optimal period to evaluate the general health of the mature woman and introduce the measures to prepare her for striking physiologic changes that come with the menopause. The patient and her clinician should attempt to achieve several important aims during the climacteric. The long-term goal is to maintain an optimal quality of physical and social life. Another immediate objective is the detection of any major chronic disorders that occur with aging. Finally, the clinician should counsel the perimenopausal woman about the symptoms and long-term consequences of menopause. The benefits and risks of lifelong hormone replacement should be discussed at great length at this time.

The Menopause

The median age of the menopause is approximately 51. The age of menopause is probably determined in part by genetic factors because mothers and daughters tend to experience menopause at the same age. A number of environmental factors may modify the age of menopause. For example, current smoking is associated with an earlier menopause, whereas alcohol consumption delays menopause. Oral contraceptive use does not affect the age of menopause.

The symptoms frequently seen and related to decreased estrogen production in menopause include irregular frequency of menses followed by amenorrhea, vasomotor instability manifest as hot flashes and sweats, urogenital atrophy giving rise to pain during intercourse and a variety of urinary symptoms, and consequences of osteoporosis and cardiovascular disease. The combination and the extent of these symptoms differ widely for each patient. Some patients experience multiple severe symptoms that may be disabling, whereas others have no symptoms or mild discomfort associated with the climacteric.

Biosynthesis of Estrogen and Other Steroids in the Postmenopausal Woman

No follicular units can be detected histologically in the ovaries after the menopause. In reproductive-age women, the granulosa cell of the ovulatory follicle is the major source of inhibin and estradiol. In the absence of these factors that inhibit gonadotropin secretion, both FSH and LH levels increase sharply after menopause. These levels peak a few years after menopause and decrease gradually and slightly thereafter. The postmenopausal serum level of either gonadotropin may be more than 100 IU/L. FSH levels are usually higher than LH levels because LH is cleared from the blood strikingly more quickly and possibly because the low levels of inhibin in the menopausal serum selectively lead to increased FSH secretion. Nevertheless, increased LH is a major factor that maintains significant quantities of androstenedione and testosterone secretion from the ovary, although the total production rates of both steroids decline after menopause.

The primary steroid products of the postmenopausal ovary are androstenedione and testosterone. The average premenopausal rate of production of androstenedione of 3 mg/day is decreased by half to approximately 1.5 mg/day. This decrease is primarily due to a substantial reduction in the ovarian contribution to the circulating androstenedione pool. Adrenal secretion accounts for most of the androstenedione production in the postmenopausal woman, with only a small amount secreted from the ovary. Both DHEA and DHEAS originate almost exclusively from the adrenal and decline steadily with advancing age independent of the menopause. The serum levels of both DHEA and DHEAS after menopause are about one fourth of those in young adult women.

Testosterone production is decreased by approximately one third after menopause. Total testosterone production can be approximated by the sum of ovarian secretion and peripheral formation from androstenedione (see Fig. 16-32). In the premenopausal woman, significant amounts of testosterone are produced by reduction of androstenedione in extracellular tissues. Because ovarian androstenedione secretion is substantially decreased after the menopause, the decrease in postmenopausal testosterone production is accounted for, in large measure, by a decrease in the relative contribution of extracellular sources. With the disappearance of follicles and decreased estrogen, the elevated gonadotropins drive the remaining stromal tissue in the ovary to maintain testosterone secretion at levels observed during the
premenopausal years. Thus, the contribution of the postmenopausal ovary to the total testosterone production is increased in the presence of seemingly unaltered ovarian secretion.  

The most dramatic endocrine alteration of the climacteric involves the decline in the circulating level and production rate of estradiol. The average menopausal level of circulating estradiol is less than 20 pg/mL. Both estradiol and estrone levels in postmenopausal women are usually slightly less than those in adult men. Circulating estradiol in postmenopausal women (and men) is derived from the peripheral conversion of androstenedione to estrone, which is, in turn, converted peripherally to estradiol (see Fig. 16-23).  

The mean circulating level of estrone in postmenopausal women (37 pg/mL) is higher than that of estradiol. The average postmenopausal production rate of estrone is approximately 42 pg per 24 hours. After menopause, almost all estrone and estradiol are derived from the peripheral aromatization of androstenedione. Thus, there is a drastic change in the androgen-to-estrogen ratio because of the sharp decrease in estradiol levels and slightly reduced testosterone. The frequent onset of a mild hirsutism after menopause reflects this striking shift in the hormone ratio. During the postmenopausal years, DHEAS and DHEA levels continue to decline steadily with advancing age, whereas serum androstenedione, testosterone, estrone, and estradiol levels do not change significantly.

The aromatization of androstenedione to estrone in extracervical tissues correlates positively with weight and advancing age (see Fig. 16-23 and Fig. 16-37). Body weight correlates positively with the circulating levels of estrone and estradiol. Because aromatase enzyme activity is present in significant quantities in adipose tissue, the aromatization of androstenedione to estrone in overweight individuals may reflect the increased aromatization of tissue containing the enzyme. In addition, there is a twofold to fourfold increase in the specific activity of aromatase per cell with advancing age. An increased overall number of adipose fibroblasts with aromatase activity and a decrease in the levels of TeBG cause an increased free estradiol level and contribute to the increased risk of endometrial cancer in obese women. The production rate and circulating levels of estradiol after menopause are clearly insufficient to provide support for urogenital tissues and bone. Thus, osteoporosis and urogenital atrophy are some of the most dramatic and unwanted consequences of estradiol deficiency during the menopause.

The consistent association between the onset of flashes and acute estrogen withdrawal is also supported by the effectiveness of estrogen therapy and the absence of hot flashes in postmenopausal women who are receiving estrogens. Under these circumstances of acute estrogen withdrawal, the lower frequency and intensity of hot flashes may vary from extremely rare to recurring every few minutes. At night, flashes are more frequent and severe enough to awaken a woman from sleep. They are also more intense during times of stress. In a cool environment, hot flashes are fewer, less intense, and shorter in duration than in a warm environment. 

The hot flash results from a sudden reduction of estrogen levels rather than from hypoestrogenism itself. Therefore, regardless of the cause of menopause, natural, surgical, or estrogen withdrawal caused by a long-acting LHRH agonist, hot flashes are associated with an acute and significant drop in estrogen level. 

The consistent association between the onset of flashes and acute estrogen withdrawal is also supported by the effectiveness of estrogen therapy and the absence of hot flashes in postmenopausal women who are receiving estrogens. Under these circumstances of acute estrogen withdrawal, the lower frequency and intensity of hot flashes may vary from extremely rare to recurring every few minutes. At night, flashes are more frequent and severe enough to awaken a woman from sleep. They are also more intense during times of stress. In a cool environment, hot flashes are fewer, less intense, and shorter in duration than in a warm environment. 

These flashes often occur in the menopause due to increased production of estrogens. The production rate and circulating levels of estradiol after menopause are clearly insufficient to provide support for urogenital tissues and bone. The clinical relevance of these findings was exemplified by the successful use of aromatase inhibitors as both first-line and second-line endocrine treatments for postmenopausal breast cancer.

Menopause and Uterine Bleeding

Perimenopausal or postmenopausal bleeding can be due to hormone administration or excessive extravascular estrogen formation. Irregular uterine bleeding is commonly observed during the perimenopausal transition as anovulatory cycles alternate with ovulatory cycles. Uterine bleeding after the menopause is less common if the patient is not receiving hormone replacement treatment (HRT). Obese patients are more likely to experience postmenopausal bleeding because of increased peripheral aromatization of androstenedione. Patients receiving continuous estrogen replacement therapy may experience unpredictable bleeding (see later). The major objective in these circumstances is to rule out endometrial malignancy. This can be best achieved by tissue diagnosis through an office endometrial biopsy using a plastic cannula. Transvaginal ultrasonographic measurement of endometrial thickness may be used in postmenopausal women to avoid unnecessary biopsies. A biopsy is required if an endometrial thickness greater than or equal to 5 mm is observed.

Before employing ultrasonography and endometrial biopsy to explore the etiology of bleeding that is assumed to arise from the intrauterine cavity, the clinician should rule out diseases of the vulva, vagina, and cervix as other potential causes of vaginal bleeding. Careful inspection of these organs and a normal cervical Pap smear within the past year are sufficient to rule out the vulva, vagina, and cervix as potential sources of bleeding. Postmenopausal uterine bleeding is the most common initial event for the patient and her physician to the possibility of endometrial cancer. On the other hand, the causes of postmenopausal uterine bleeding are benign most of the time. Endometrial malignancy is encountered in patients with bleeding in only about 1% to 2% of postmenopausal endometrial biopsies. Approximately three quarters of these biopsies reveal either no pathology or an atrophic endometrium. Other histologic findings include hyperplasia (15%) and endometrial polyps (3%). Persistent unexplained uterine bleeding requires repeated evaluation, biopsy, hysteroscopy, or dilatation and curettage.

Unpredictable irregular uterine bleeding is observed in approximately 20% of postmenopausal women receiving a long-term (>1 year) continuous estrogen-progesterone combination. This should also be evaluated appropriately with ultrasonography or biopsy, or both (see later).

Hot Flash

The most frequent and striking symptom in the climacteric is the hot flash. The hot flash typically occurs at the time of transition from perimenopause to postmenopause, is the climacteric. The flash is also a major symptom of the postmenopause and can last up to 5 years after menopause. More than four fifths of postmenopausal women experience hot flashes within 3 months after the cessation of ovarian function, whether natural or surgical in origin. Of these women, more than three fourths have them for more than 1 year and approximately half for up to 5 years. Hot flashes lessen in frequency and intensity with advancing age, unlike other sequelae of the menopause, which progress with time.

A hot flash is a subjective sensation of intense warmth of the upper body, which typically lasts for 4 minutes but may range in duration from 30 seconds to 5 minutes. It may follow a prodrome of palpitations or headache and is frequently accompanied by weakness, faintness, or vertigo. This episode usually ends in profuse sweating and a cold sensation. The frequency may vary from extremely rare to recurring every few minutes. At night, flashes are more frequent and severe enough to awaken a woman from sleep. They are also more intense during times of stress. In a cool environment, hot flashes are fewer, less intense, and shorter in duration than in a warm environment.

The hot flash results from a sudden reduction of estrogen levels rather than from hypoestrogenism itself. Therefore, regardless of the cause of menopause, natural, surgical, or estrogen withdrawal caused by a long-acting LHRH agonist, hot flashes are associated with an acute and significant drop in estrogen level. 

The consistent association between the onset of flashes and acute estrogen withdrawal is also supported by the effectiveness of estrogen therapy and the absence of hot flashes in postmenopausal women who are receiving estrogens. Under these circumstances of acute estrogen withdrawal, the lower frequency and intensity of hot flashes may vary from extremely rare to recurring every few minutes. At night, flashes are more frequent and severe enough to awaken a woman from sleep. They are also more intense during times of stress. In a cool environment, hot flashes are fewer, less intense, and shorter in duration than in a warm environment. 

Urogenital Atrophy

The urogenital sinuses give rise to the development of the lower vagina, vulva, and urethra during embryonic development, and these tissues are estrogen-dependent. The urogenital atrophy at menopause causes the vaginal walls to become pale because of diminished vascular supply and to thin down to only three to four cell layers. The vaginal epithelial cells in postmenopausal women contain less glycogen, which prior to menopause was metabolized by lactobacilli to create an acidic pH, thereby protecting the vagina from bacterial overgrowth. Loss of this protective mechanism leaves the thin, friable tissue vulnerable to infection and ulceration. The vagina also loses its rugae and becomes shorter and inelastic. Postmenopausal women may complain of symptoms secondary to vaginal dryness, such as pain during intercourse, vaginal discharge, burning, itching, or bleeding. Genitourinary atrophy leads to a variety of symptoms that affect the ease and quality of living.

Urethritis with dysuria, stress urinary incontinence, and urinary frequency are further results of mucosal thinning of the urethra and bladder. Intravaginal estrogen use of aromatase inhibitors as both first-line and second-line endocrine treatments for postmenopausal breast cancer.
treatment can effectively alleviate recurrent urinary tract infections and vaginal symptoms in the postmenopausal patient. Oral estrogen replacement also rapidly reverses vaginal atrophy and urethral symptoms caused by estrogen deficiency.

Cognitive Function and Estrogen

Although Alzheimer's disease affects both men and women, studies in many different populations show that 1.5 to 3 times as many women as men suffer from this disease. For women with this disease, one needs to consider behavioral and cognitive problems, therapeutic issues, and other gender-related risks. One obvious consideration is a possible link between estrogen deficiency and the increased incidence of Alzheimer's disease in postmenopausal women. A number of studies have suggested that the association is stronger among women without Alzheimer's disease. Treatment of women free of Alzheimer's disease with hormone replacement for signs and symptoms of menopause led to improvements in verbal memory, vigilance, reasoning, and motor speed but no enhancement of other cognitive functions. A meta-analysis of observational studies suggested that HRT was associated with a decreased incidence of dementia. However, possible biases and lack of control for potential confounders limit interpretation of these studies. Two small but randomized studies reveal the estrogen and short-term hormone therapy in women with Alzheimer's disease produced conflicting data. Thus, future large randomized trials should target the potential benefits of HRT in improving cognition in women with menopausal symptoms as well as in the prevention and treatment of Alzheimer's disease. The ongoing large trial called the Women's Health Initiative may provide some answers in several years.

Cardiovascular Disease

It has long been suggested that estrogen protects against atherosclerosis because the incidence of cardiovascular disease is lower in women than in men in all age groups. The gender gap is widest during the premenopausal years. For example, myocardial infarction is six times less common in premenopausal women than in men of the same age. This discrepancy has been attributed, in part, to the effects of estrogen. During reproductive life, women have lower LDL cholesterol than men, although these levels gradually increase with advancing age and rise rapidly after menopause. In contrast, despite consistently higher high-density lipoprotein (HDL) cholesterol levels in women than men throughout adulthood, this discrepancy persists after the menopause. HDL cholesterol becomes higher in premenopausal women but does not change greatly at menopause. The increases in LDL and total cholesterol levels at menopause are partially reversible with estrogen treatment. Therefore, the increased myocardial infarction rate after menopause may be a function of rising LDL cholesterol levels that seem to be related to decreased estrogen levels. Although HDL cholesterol levels do not decrease significantly in postmenopausal women, replacement with oral estrogen increases HDL cholesterol levels significantly; this may contribute to the cardioprotective effect of HRT.

A strong correlation of a cause-and-effect relationship has been established between cholesterol levels and coronary heart disease in postmenopausal women. Because menopausal postmenopausal women with elevated total cholesterol levels have a significantly increased risk of coronary heart disease compared with women with low levels. This risk is more pronounced in the first two decades after the menopause. The association becomes less pronounced with advancing age. In both men and women, an HDL cholesterol level is the most specific determinant of coronary heart disease. High levels of HDL cholesterol are protective, whereas a low level is a strong predictor of increased cardiovascular risk. Therefore, monitoring HDL cholesterol as well as total and LDL cholesterol levels is important in determining cardiovascular risk in postmenopausal women.

Replacement with estrogen decreases LDL and increases HDL cholesterol levels in postmenopausal women. Although estrogen also increases triglyceride levels, the impact of this effect on the cardiovascular system is unknown. If a progestin is added, the beneficial effects of estrogen may diminish. Overall, postmenopausal estrogen replacement with or without added progestin produces a favorable lipid profile. These favorable biochemical effects may provide some protection against cardiovascular disease. Many trials including a large, randomized study demonstrated a favorable impact on cardiovascular risk factors in women taking estrogen as well as a combination of estrogen and progestin. Moreover, the great majority of studies investigating coronary or cerebrovascular disease as an outcome concluded that postmenopausal use of estrogens protected against cardiovascular disease; although these studies have been observational rather than randomized and blinded trials. Thus, the opinion of the medical community in general was that exogenous estrogen given in postmenopausal replacement doses decreased cardiovascular risk for all women.

A later randomized study of estrogen-progestin users with preexisting coronary disease (Heart and Estrogen/progestin Replacement Study, or HERS) showed an early increase in mortality. Consequently, in 1999 the American College of Cardiology and American Heart Association revised their guidelines for providing HRT to patients with a history of acute cardiovascular infarction. The authors of HERS concluded that, over an average follow-up of 4 years, treatment with oral conjugated equine estrogen plus medroxyprogesterone acetate did not reduce the overall rate of coronary heart disease events in postmenopausal women with established coronary disease. However, it was associated with a small but statistically significant increase in the risks of deep vein thrombosis and pulmonary embolism. No significant effect of HRT on the risk of stroke was detected in these postmenopausal women with coronary disease. HRT in HERS also resulted in a marginally significant increase in the risk of symptomatic gallbladder disease and biliary tract surgery. It was recommended that hormone replacement therapy should not be initiated solely for prevention of cardiovascular disease in postmenopausal women with preexisting coronary heart disease but can be continued in patients with cardiovascular disease already receiving HRT for other reasons.

Another randomized trial showed that estradiol does not reduce mortality or the recurrence of stroke in postmenopausal women with preexisting cerebrovascular disease. Thus, HRT should not be prescribed for the secondary prevention of cerebrovascular disease.

Recent findings from the Women's Health Initiative reinforce and expand the concerns raised by these randomized trials. In the Women's Health Initiative, more than 16,000 women with intact uteri were randomized to receive either placebo or conjugated estrogens 0.625 mg and medroxyprogesterone 2.5 mg daily. The study was designed to investigate cardiovascular disease. However, possible biases and lack of control for potential confounders limit interpretation of these studies. Recent findings from the Women's Health Initiative reinforce and expand the concerns raised by these randomized trials. In the Women's Health Initiative, more than 16,000 women with intact uteri were randomized to receive either placebo or conjugated estrogens 0.625 mg and medroxyprogesterone 2.5 mg daily. The study was designed to investigate cardiovascular disease. However, possible biases and lack of control for potential confounders limit interpretation of these studies. For example, the estrogen plus progestin arm of the study was associated with an increased risk of coronary heart disease events, while the estrogen-only arm was associated with a decreased risk. These findings have been interpreted as evidence against the use of estrogen therapy in postmenopausal women, especially those with a history of cardiovascular disease.

Cardiovascular disease.

Postmenopausal Osteoporosis

Osteopenia and osteoporosis are extremely common in elderly postmenopausal women. Osteopenia indicates low bone mass measured by densitometry, whereas the term osteoporosis implies severely decreased bone mass associated with a significantly increased risk for fractures. The most frequent sites of fracture are the vertebrae, hip, and wrist. Despite the high prevalence of osteoporosis, the condition is often unrecognized. In postmenopausal women, it is currently an epidemic proportion in the United States, affecting over 20 million people. The majority of osteopathic patients are postmenopausal women.

Osteoporosis in postmenopausal women is a function of both advancing age and estrogen deficiency. Seventy-five percent or more of the bone loss in women during the first 15 years after menopause is attributed to estrogen deficiency rather than to aging itself. For the first 20 years after the cessation of ovarian estrogen secretion, postmenopausal osteoporosis accounts for a 50% reduction in trabecular bone and 30% loss of cortical bone. Vertebral bone is especially vulnerable because the trabecular bone of the vertebral bodies is metabolically very active and decreases dramatically in response to estrogen deficiency. Vertebral bone mass is already significantly decreased in perimenopausal and early postmenopausal women who have rising FSH and decreasing estrogen levels, whereas bone loss from the radius is not detected until at least a year after the menopause.

The risk of fracture depends on two factors: the peak bone mass achieved at maturity (at approximately age 30) and the subsequent rate of bone loss. An accelerated rate of bone loss after menopause strongly predicts an increased risk of fracture. The combination of low postmenopausal bone mass and accelerated loss of bone after menopause is additive, and these individuals are at the highest risk of fracture. An increased rate of average bone loss during menopause is an indicator of lower endogenous estrogen levels because postmenopausal bone loss is considerably slower in women with increased adipose tissue mass and thus elevated...
It has been shown conclusively by numerous studies that hormone replacement started at the climacteric prevents postmenopausal bone loss. Hormone replacement started at any age in a postmenopausal woman has potential beneficial effects by at least preventing additional bone loss. It should be noted, however, that the incidence of fractures or rate of height loss was not reduced during a 4-year follow-up in women starting HRT at a mean age of 66.7 (± 6.7) years. More recently, in the part of the Women's Health Initiative discussed above, a decreased number of hip and vertebral fractures were noted in the group of postmenopausal women receiving estrogen/progestin; this important evidence was the first from randomized trials suggesting that estrogen prevents fractures.
Postmenopausal Hormone Replacement

The most common current practice is to treat all women disturbed by the symptoms of hormone deprivation (hot flashes and urogenital atrophy) with estrogen and to use long-term hormonal prophylaxis against osteoporosis and cardiovascular disease on the basis of an informed decision by the patient. Patients who have undergone a hysterectomy can be given estrogen therapy alone. A progestin is added to estrogen in the postmenopausal woman with a uterus in order to prevent endometrial hyperplasia or cancer. The recent findings from the Women's Health Initiative trial are too new to allow prediction of how the negative findings from this trial will affect practice. Modest increases in breast cancer, coronary heart disease, and stroke were balanced by modest decreases in fractures and colon cancer. How individual patients and physicians will balance the immediate and common benefits in the treatment of hot flashes and urogenital atrophy against the predominantly negative long-term consequences that affect a modest fraction of patients (at least over 5 years) remains to be determined. It should be noted that the part of the Women's Health Initiative involving the use of estrogen without progestin to treat women after hysterectomy has not been halted; those results will become available in 2005. The decision for using long-term hormone replacement should be made by the patient based on accurate information. The primary role of the physician is to provide scientific information to a patient, using understandable language. This is not an easy task, given the complexity of the existing data and occasionally opposite opinions.

Estrogens and progestins used for postmenopausal hormone replacement are among the most commonly prescribed medications in the United States. Currently, 46% of women who have experienced a natural menopause and 71% of women who have had bilateral oophorectomy report having used postmenopausal HRT. The average duration of use in the United States as of 1992 was 6.6 years, but only 20% of users had maintained treatment for at least 5 years. Emphasizing the education of the patient and primary care physician on the basis of appropriate interpretation of epidemiologic studies and clinical trials will ensure appropriate use of long-term postmenopausal HRT.

Target Groups for Hormone Replacement

In women with gonadal dysgenesis and surgical menopause, the duration of estrogen deprivation is prolonged. Estrogen replacement is recommended for these patients for the reduction of hot flashes and for long-term prophylaxis against cardiovascular disease, osteoporosis, and target organ atrophy. A low-dose contraceptive may be offered to nonsmoking women until the age of 45. After this age, doses of estrogen equivalent to 0.625 mg of conjugated estrogens may be more appropriate because of a sharp age-related increase in risk for thromboembolic events. The physician should recommend a continuous estrogen-progestin combination to those with a uterus and an estrogen-only regimen to women without a uterus.

During the climacteric, hot flashes can be suppressed with an estrogen-progestin combination. Because bone loss related to estrogen deprivation also begins during this period, starting hormone replacement therapy during the climacteric is of paramount importance for minimizing osteoporosis. In climacteric women, unexplained uterine bleeding should be evaluated with an endometrial biopsy before the start of hormone replacement.

The lifelong use of hormone therapy after menopause is dependent on the informed decision of the woman based on balanced and evidence-based advice provided by her physician. The benefits of hormone replacement with respect to bone metabolism, cognitive function, urogenital health, and sexual function are substantial and for many women outweigh the increased breast cancer and cardiovascular disease risk.

Estrogen Preparations and Beneficial Dose of Estrogen

The amount of estrogen that is optimally effective in maintaining the spine and femoral neck bone mass is equivalent to 0.625 mg/day of conjugated estrogens. The effective doses of oral estrogen that reduce the incidence of fracture are 0.625 mg/day of conjugated estrogens and 1.25 mg/day of estrone sulfate. Also, transdermal estradiol at a rate of 0.05 mg/贴 is also presumed to lower fracture risk based on equivalent doses of various preparations that provide similar average circulating estradiol levels. Oral intake of 0.625 mg of conjugated estrogen, 1.25 mg of estrone sulfate, or 1 mg of micronized estradiol results in similar average serum levels of estrogens: estradiol, 30 to 40 pg/mL, and estrone, 150 to 250 pg/mL. Transdermal administration of estradiol with patches releasing 0.05 mg/day gave rise to similar average serum estradiol (30 to 40 pg/mL) but much lower estrone (40 pg/mL).

Short-term cardiovascular or hemodynamic effects of estrogen vary according to blood estrogen levels. Improvements in left ventricular contraction and function are associated with levels achieved by 0.625 mg of oral conjugated estrogens. Extremely high estradiol levels achieved with large doses of estrogen, on the other hand, decrease left ventricular function and aortic blood flow. The beneficial effect of postmenopausal estrogen in preventing the hyperinsulinemia associated with aging is present with a dose of 0.625 mg of conjugated estrogens but lost with a dose of 1.25 mg.

The effect of estrogen on arterial thrombosis is also dose-related. For example, oral contraceptives with high doses of estrogen significantly increase the risks of myocardial infarction and stroke, especially in smokers. Numerous studies suggest that doses of conjugated estrogens greater than 0.625 mg are, in fact, not as beneficial in terms of cardiovascular disease and mortality. However, the number of patients receiving these high doses was not large enough to achieve statistical significance. Thus, it is imperative to achieve and maintain the lowest beneficial levels of circulating estradiol and avoid higher levels associated with unfavorable hemodynamic effects or thrombosis.

Estrogen-Progestin Regimens

The addition of a progestin, either cyclically or continuously, to concomitant estrogen replacement reduces the risk of estrogen-induced endometrial hyperplasia or carcinoma but poses additional problems. These problems include regular withdrawal bleeding in up to 90% of women treated with cyclic therapy and irregular spotting in 20% of women treated with continuous estrogen plus progestin. Furthermore, progestins appear to reduce the beneficial effects of estrogen on HDL and LDL cholesterol and possibly cardiovascular risk.

A time-honored sequential regimen involves oral administration of 0.625 mg of conjugated estrogens or the equivalent dose of a variety of available products from day 1 to 25 of each month. A daily dose of 10 mg of medroxyprogesterone acetate is added from day 12 to 25 or from day 16 to 25. Withdrawal bleeding is expected on or after day 26 of each month. Another common cyclic regimen involves continuous oral administration of 0.625 mg of conjugated estrogens or the equivalent daily dose. A daily dose of 5 to 10 mg of medroxyprogesterone acetate is added for the first 10 to 14 days of each month. One-year randomized trial data indicate that the 5-mg dose protects the endometrium as well as the 10-mg dose. Progestin withdrawal bleeding occurs in 90% of women with a sequential or cyclic regimen. These regimens can also cause adverse symptoms related to the relatively high daily doses of progestin, such as breast tenderness, bloating, fluid retention, and depression. Thus, the

[Larsen: Williams Textbook of Endocrinology, 10th ed., Copyright © 2003 Elsevier]
The lowest possible dose of a progestin is recommended.

The continuous combined method of treatment, on the other hand, has the potential benefit of reduced bleeding and amenorrhea but is occasionally complicated by breakthrough bleeding (see Fig. 16-41). In this regimen, a combination of 0.625 mg of conjugated estrogens and 2.5 mg of medroxyprogesterone acetate is given orally every day. The continuous combination regimen is simple, convenient, and associated with a higher incidence of amenorrhea in 80% of patients after at least 6 months of use. The rest of the patients continue to experience some degree of unpredictable spotting. Thus, overall compliance is much better in users of the continuous combination regimen. Moreover, the lower daily dose of medroxyprogesterone acetate is associated with a lower incidence of breast tenderness with this regimen. Other estrogen-progestin combinations are also available for similar continuous use.

Most postmenopausal women stop HRT within several years after the initiation of therapy because hot flashes decrease significantly or disappear at this time. Upon the cessation of HRT that had lasted several years, hot flashes usually do not return or are less severe than during the climacteric. Consequently, long-term compliance with HRT remains poor. The fear of malignancy and irregular uterine bleeding are two major factors that stop women from continuing HRT indefinitely. The current data on breast cancer are indicative of a slightly increased breast cancer risk for estrogen-only and estrogen plus progestin regimens. The addition of a progestin to the estrogen-only regimen, on the other hand, has effectively prevented endometrial cancer. Irregular uterine bleeding is another major reason for discontinuation of HRT. The incidence of persistent uterine bleeding with the traditional sequential regimen can be as high as 90%, which deters women from taking lifelong men. By switching to the continuous combination regimen, this bleeding can be reduced to a 20% incidence of spotting with long-term (>1 year) use. Thus, more clinicians start women with the continuous combination treatment with the hope of improving compliance for long-term use (see Fig. 16-41).

Selective estrogen receptor modulators are compounds that act like estrogen in some target tissues but antagonize estrogenic effects in others. One of the first selective estrogen receptor modulators was tamoxifen, for which estrogen-like agonist activity on bone was observed to occur simultaneously with estrogen antagonist activity on the breast. An unwanted effect of tamoxifen is its estrogen-like action on the endometrium. Second-generation compounds have since been developed, notably raloxifene, which has estrogen-like actions on bone, lipids, and the coagulation system; estrogen antagonist effects on the breast; and no detectable action in the endometrium.

In randomized placebo-controlled studies involving postmenopausal women or patients with osteoporosis, raloxifene at 60 to 150 mg/day was effective in increasing bone mineral density over 12 to 36 months. Raloxifene also decreased the risk of vertebral fractures. Raloxifene is similar to placebo in its endometrial effects and similar to estrogen in causing a twofold to threefold increase in the risk of venous thromboembolism. Raloxifene lowers total and LDL cholesterol. HDL cholesterol and triglycerides are virtually unaffected. Patients receiving either a cyclic or daily combination hormone replacement regimen who have an endometrial thickness less than 5 mm can be managed conservatively. The propensity of raloxifene to cause hot flashes precludes its use in women with vasomotor symptoms. On the other hand, the lack of stimulatory effects on the endometrium and the reduction in the incidence of invasive breast cancer indicate that raloxifene is an alternative to tamoxifen for the management of postmenopausal osteoporosis, especially for patients reluctant to use estrogen. It should be emphasized that, at this time, postmenopausal hormone replacement with estrogen remains the "gold standard" for health benefits with respect to bone, cardiovascular system, urogenital organs, and possibly the central nervous system.

Approximately 90% of women receiving estrogen plus cyclic administration of a progestin have monthly progestin withdrawal bleeding in a predictable fashion, whereas continuous combined estrogen-progestin therapy causes breakthrough bleeding in approximately 40% of women during the first 6 months. (The rest of the women with a continuous combination regimen are amenorrheic.) The pattern of vaginal bleeding in the continuous regimen is unpredictable and causes anxiety in most patients. Fortunately, the incidence of breakthrough bleeding with the continuous combined regimen decreases to 20% after 1 year of treatment.

Breakthrough bleeding with the combined continuous regimen remains the most important reason for discontinuance of this therapy. Most patients find it unacceptable and prefer to switch to a cyclic progestin regimen or discontinue hormone replacement altogether. There is no effective pharmacologic method to manage the breakthrough bleeding associated with continuous combined estrogen-progestin regimens. One can only reassure the patient that the bleeding is likely to subside within a year from the start of HRT. If breakthrough bleeding continues beyond a year, the regimen should be changed to daily estrogen plus cyclic progestin monthly for 10 days.

HRT can be started in the amenorrheic postmenopausal patient at any time. Perimenopausal women with oligomenorrhea, hot flashes, or other associated symptoms should also be given HRT. In the oligomenorrheic patient, a hormone replacement regimen can be initiated on day 3 of one of the infrequent menses. If the candidate for hormone replacement does not have irregular uterine bleeding, it is not essential to perform endometrial biopsies routinely before beginning treatment. Studies indicate that asymptomatic postmenopausal women rarely have endometrial abnormalities. Pretreatment biopsies using a thin plastic biopsy cannula in the office may be limited to patients at higher risk for endometrial hyperplasia (e.g., unpredictable uterine bleeding, history of PCOS or chronic anovulation, obesity, liver disease, and diabetes mellitus).

Giving a woman a combined estrogen-progestin regimen does not preclude the development of endometrial cancer. It is, therefore, necessary to rule out endometrial malignancy in women receiving HRT who are experiencing irregular uterine bleeding. The important task is to differentiate breakthrough bleeding from bleeding induced by hyperplasia or cancer. Because breakthrough bleeding is extremely common, a large number of biopsies would have to be performed to detect a rare case of endometrial abnormality during HRT. In order to decrease the number of endometrial biopsies, a screening method using transvaginal ultrasonography has been introduced. The thickness of the postmenopausal endometrium as measured by transvaginal ultrasonography in postmenopausal women correlates with the presence or absence of pathology. Patients receiving either a cyclic or daily combination hormone replacement regimen who have an endometrial thickness less than 5 mm can be managed conservatively. An endometrial thickness equal to or greater than 5 mm requires biopsy. Following this algorithm, it is estimated that 50% to 75% of bleeding patients receiving HRT and evaluated by ultrasonography require biopsy.

Hypertension

This rare side effect is observed in patients with severe familial hypertriglyceridemia. An oral estrogen regimen can hasten severe hypertriglyceridemia or pancreatitis in women with severely elevated triglyceride levels. Therefore, estrogen replacement is a relative contraindication in women with substantially increased triglyceride levels.

Gallbladder Disease

There is a minimally increased risk of gallbladder disease with estrogen use during the menopause. There is a marginally significant increase in the risk of cholecystectomy in past and current users of HRT. Preexisting gallbladder disease is a relative contraindication for estrogen replacement.

Breast Cancer

It has been debated whether continuous combination hormone replacement using conjugated estrogens (0.625 mg/day) plus medroxyprogesterone acetate (2.5 mg/day) increases the risk of venous thromboembolism. A recent randomized clinical trial has shown that this regimen of HRT significantly increases the risks of deep venous thrombosis and pulmonary embolism, although these increases were modest. Even if the relative risk is significantly increased, the actual risk remains extremely low because of the low frequency of this event in the general population.
Breast tissue is a major target for estrogen, and most breast tumors are estrogen-responsive. A number of case-control and cohort studies concluded that 5 or more years of current use of postmenopausal HRT is associated with a slight increase in the risk of breast cancer, a risk that is less than that associated with postmenopausal obesity or daily alcohol consumption.\footnote{648} \footnote{649} Many observational studies, however, have failed to develop evidence that long-term postmenopausal HRT increases the risk of breast cancer.\footnote{648} \footnote{649} Moreover, none of the epidemiologic studies found an increased risk of breast cancer associated with less than 5 years of use or past use of postmenopausal HRT.\footnote{649} The addition of a progestin to the treatment regimen slightly increased the risk observed in estrogen-only users.\footnote{649} In this context, the recent findings from the Women's Health Initiative support and extend the previous work, in that the combined use of estrogen and progestin led to a statistically significant increase in invasive breast cancer over the 5.2 years of the study.\footnote{593A} Because breast cancer was uncommon in this group of postmenopausal women, this increase represented an absolute increase of 8 cases per 10,000 women receiving hormone replacement.

Epidemiologic data indicate that a positive family history of breast cancer should not be a contraindication to postmenopausal estrogen use. Moreover, postmenopausal women in whom the cancer develops during HRT have a reduced risk of dying from breast cancer.\footnote{649} This reduced risk may be due to an increased rate of early detection and development of less aggressive tumors in association with HRT. In conclusion, there may be a minimally increased real risk of developing breast cancer in long-term HRT users. This risk, however, is extremely small compared with the clear-cut benefits of estrogen such as osteoporosis prevention and urogenital tissue support.

Hormone Replacement Therapy after a Diagnosis of Breast Cancer.

HRT is typically withheld from women with breast cancer because of concerns that estrogen may stimulate recurrence. Surprisingly, a number of relatively small studies showed either unaltered or lower risks of recurrence and mortality in women who used HRT after a diagnosis of breast cancer compared with nonusers.\footnote{650} \footnote{651} HRT in most of these small studies was started after at least a 5-year disease-free interval.\footnote{650} \footnote{651} On the basis of these insufficient but encouraging data, a decision to provide HRT is dependent on the choice of the individual patient. In these patients, tamoxifen or raloxifene represents a viable alternative to estrogen replacement for long-term prophylaxis against osteoporosis.
References


73. Witschi E. Migration of the germ cells of human embryos from the yolk sac to the primitive gonadal folds. Contrib Embryol 1948; 32:67.


References


179. Waterman M, Simpson E. Regulation of steroid hydroxylase gene expression is multifactorial in nature. Recent Prog Horm Res 1989; 45:533566.


655


References


315. Kovacs C, Chik C. Hyperprolactinemia caused by lactation and pituitary adenomas is associated with altered serum calcium, phosphate, parathyroid hormone (PTH), and PTH-related peptide levels. J Clin Endocrinol Metab 1985; 80:3033042.


440. Velazquez E, Mendoza S, Wang P, et al. Metformin therapy is associated with a decrease in plasma plasminogen activator inhibitor-1, lipoprotein (a), and immunoreactive insulin levels in patients with the polycystic ovary syndrome. Metabolism 1997; 46:454457.

680
References


References


cite{Alexandersen P, Ris B, Christiansen C. Monofluorophosphate combined with hormone replacement therapy induces a synergistic effect on bone mass by dissociating bone formation and resorption in postmenopausal women: a randomized study. J Clin Endocrinol Metab 1999; 84:30133020.}


cite{Archer D, Pickar J, Bottiglioni F. Bleeding patterns in postmenopausal women taking continuous combined or sequential regimens of conjugated estrogens with medroxyprogesterone acetate. Menopause Study Group. Obstet Gynecol 1994; 83:686692.}


cite{Ravnikar V. Compliance with hormonal therapy. Am J Obstet Gynecol 1987; 156:13321334.}


cite{Diez J. Skeletal effects of selective oestrogen receptor modulators (SERMs). Hum Reprod Update 2000; 6:255258.}

cite{Clemett D, Spencer C. Raloxifene: a review of its use in postmenopausal osteoporosis. Drugs 2000; 60:379411.}

cite{Burger H. Selective oestrogen receptor modulators. Horm Res 2000; 53(suppl):2529.}


Chapter 17 - Fertility Control: Current Approaches and Global Aspects

Michael Kafriessen
Eli Adashi

GLOBAL ASPECTS OF FERTILITY CONTROL

Regulation of human fertility encompasses social, legal, and health care measures, including the employment of medical technologies and procedures that result in limiting the number of offspring. This can be achieved by spacing the birth of individual children in the sense of family planning or by terminating procreation when the desired family size has been achieved. The aim of fertility control is to achieve a family or population size that is compatible with a reasonable quality of life and is culturally and economically supported.

Growth of the Human Population

History and Current Status

The explosion of the world population is a relatively recent phenomenon (Fig. 17-1). By AD 1, 200 million people lived on this planet. By 1650, more than 1600 years later, this number had increased to 500 million. Thereafter, the population-doubling time became progressively shorter. By 1850, the human population numbered 1 billion, and within 80 years, in 1930, the population reached 2 billion. In 2000 the world population crossed the 6 billion milestone, and the growth continues.

Consequences of Overpopulation: From Malthus to Ehrlich

During the industrial revolution, Thomas Malthus, an English political economist, noticed the soaring population of the British Isles, a geographic area with limited arable land, and concluded that the human population would outrun the available food supplies. Between 1778 and 1816, Malthus published a series of essays in which he predicted that a collision of population growth and lack of world food supplies would result in worldwide disasters such as famines, epidemics, and wars. Malthus believed that strict limits on reproduction were essential to the betterment of humankind.

In 1968, at a time when humankind experienced the most prominent population increase (see Fig. 17-1), Paul Ehrlich, a biologist at Stanford University, brought the issue of overpopulation to the public consciousness in his disquisition *The Population Bomb*. Ehrlich pointed out the negative effects of population growth on the environment, nonrenewable natural resources, and general economic progress.

Coping with the Population Growth

Human ingenuity has largely disproved the predictions of both Malthus and Ehrlich. The extraordinary demographic history of the 20th century that occurred without global malthusian disasters can be explained mainly by two phenomena: innovations in agriculture and advances in medicine.

Agriculture: From the Neolithic to the Green Revolution

Historically, the phases in which the growth of the world population took an upward trend have coincided with improvement in agricultural techniques. In the Neolithic era, humans started to grow food supplies and this transition from hunting and gathering resulted in the first substantial population growth. However, further agricultural progress was slow, and so was the population growth.

Not until the 18th and 19th centuries were more advanced agricultural methods applied. The industrialization of agriculture dates only from the beginning of the 20th century, when the farm tractor, introduced in Iowa in 1901, led to mechanization of the farmer’s work. Further improvements, known collectively as the green revolution, followed. They included the use of chemical fertilizers, herbicides, and insecticides; the development of hybrid grains; and, more recently, the genetic manipulation of rice and grains. Farmers in the United States planted hybrid seeds and increased yields of grains by one third in only one decade, between 1930 and 1940.

During the first 35 years of the green revolution, global grain production doubled.

Famine, which historically has been a worldwide and perennial problem, became much less of a threat in the 20th century. Although the last 100 years have seen devastating famines in numerous areas of the world and millions of undernourished people have died, the demographic impact of these famines was relatively local and short-term. Famines are endemic on the Indian subcontinent, for example, but its population increased from 300,000 in the year 1900 to 1.3 billion in 2001. Despite predictions that the world population would outstrip food production, food production has risen a full 16% above population growth.

Today, there are fewer hungry people than ever before in history. In 1996, the number of hungry people of the world was 17%, whereas in 1970, 25 years earlier, 35% of the world's people fit in the “hungry” classification.

The increased productivity of industrialized agriculture has demographic consequences. Currently, 2% of the world's farmers, most of them in developed and rapidly developing countries, produce one fourth of the world's food. The reduced need for workers has liberated the industrial farmer from the pressure to have a large family. By contrast, traditional farmers, mostly in the underdeveloped countries, still feel the need to secure the necessary agricultural workforce through having more children.

The success of industrial farming, however, came at a price. The expansion of arable land disturbed the balance of the ecosystem. For example, deforestation and reduction of natural pastures led to droughts and floods in certain areas. During the next 50 years, the growing global population will require further advances in agricultural production to ensure a sufficient, secure, and equitable food supply. To control population growth...
reversethe environmental impacts of agricultural expansion, intensive scientific efforts must be implemented along with regulatory, technologic, and policy changes.10

Medicine: From Art to Science

During the 20th century, major advances in the practice of medicine and the application of preventive medicine increased worldwide population survival rates. A number of factors contributed to enhanced survival: an increase in the number of infants born alive, a decrease in infant and child mortality, and the containment of the spread of major epidemics. In the developed countries, the perinatal mortality rate fell from 225 per 1000 live births to under 20 per 1000. In the less developed countries, the perinatal mortality rate also decreased, but the disparity between underdeveloped and some industrialized nations was staggering. In 1988, the perinatal mortality was 5 per 1000 for Japan but it was 118 per 1000 for Bangladesh.11

Globally, life expectancy increased from approximately 30 years at the beginning of the 20th century to 47 years in the middle of the century and 65 years by the end of the century. It is projected that by the year 2050, life expectancy will be 76 years. The current 180,000 living centenariansmost of them in Europe, Japan, China, and the United Statesbest attest to the improved health conditions of humankind.

The kinds of epidemics that decimated populations in the past have become less of a peril. Historians estimate that in the 1350s, in medieval England, the epidemic of plague known as the "black death" reduced the total English population by 20% and decreased life expectancy to under 18 years.12 The acquired immunodeficiency syndrome (AIDS), a modern-day parallel to a medieval plague, has taken its toll mainly in Africa, where 28 million of the 36 million AIDS-affected people live. Nevertheless, even in the sub-Saharan nations, where the incidence of AIDS is as high as 30% and life expectancy has been reduced to 37.2 years, the population continues to expand, although at a slower rate than before the outbreak of the epidemic. However, in some countries of sub-Saharan Africa, notably in South Africa and Zimbabwe, a negative growth rate is expected.13

In summary, malthusian predictions of global catastrophes caused by overpopulation did not materialize. However, the balance of a steadily growing population and potential technologic limitations on resources remains precarious.
Population Projections

Current projections for the growth of the world population by the middle of the 21st century vary from 7.9 billion to 10.9 billion. [See Fig. 17-1]. With respect to individual continents, Asia will remain the most populous. The most significant growth is expected to occur in Africa, with the population rising from the current 900 million to 2 billion. Notable population growth is also projected for countries in both North America and Latin America. The United States will probably be the only industrialized nation with a population increase. The population of Europe is expected to decrease [Fig. 17-2]. However, there is a broader question related to the world population growth, namely what is the optimal number of people the world can support? In 1994, a serious attempt to answer this question was made at the International Conference on Population and Development, now known as the Cairo Conference. The conference forecasted a world population of 10 billion in 2050 and recommended holding the population at that level.
The Future of the World Population

Medical progress has affected human demographics by another innovation: the development of effective means of preventing pregnancy. Introduced in the 1960s when the population was increasing, modern contraception was immediately recognized as a potential instrument for large-scale family planning. Indeed, available demographic data show that since the 1980s, population growth rates, along with fertility rates, have been falling (Fig. 17-3: see Fig. 17-1). This signifies a tapering off of the most rapid phase of world population growth. However, the data also show that the momentum of growth will persist into the middle of this century, when the world population is projected to be 10 billion.

The goal of holding the world population at 10 billion can be accomplished if the fertility rate is limited to an average of 2.1 children per woman, the essential replacement rate; this can be achieved by effective fertility control.
Historically, most societies rewarded families for having children by reducing their tax burden and giving awards to mothers of numerous children. In the second half of the last century, however, certain societies, pressured by overpopulation, installed strong disincentives for families that had more than a prescribed number of children. In China, for example, a family that has more than the allowable one child is castigated. In Singapore, a highly developed but crowded country, the acceptable number of children for a family used to be determined, among other things, by the parents' educational level. In other cultural environments, the government resorted to positive incentives by providing gifts to individuals who underwent voluntary sterilization. Coercive methods of family planning have been criticized and are globally unacceptable.

Modern family planning must respect the freedom of reproductive choices, free access to family planning facilities, and availability of up-to-date methods of fertility control. Individual families as well as entire nations must be educated to comprehend that effective family planning makes good social and economic sense. Currently available data show that the decrease of birth rates is inversely proportional to the percentage of the population practicing contraception. The data also show that nations enjoying the highest living standards or those striving to overcome economic hurdles have the lowest birth rates. Examples of the latter are the countries of the former Eastern European bloc.

Effective fertility control can be achieved only when the society supports it. Initially, after their introduction in the 1960s, modern methods of human fertility control were accessible only to the affluent. Today, in the milieu of globalization, contraception has become a matter of governmental policy in many nations. It is also of primary concern to global organizations such as the World Health Organization (WHO) and the United Nations Educational, Scientific, and Cultural Organization as well as private institutions such as the Population Council and the Alan Guttmacher Institute of Family Planning.
The Role of the Physician and Other Health Care Givers in Fertility Control

Health care givers in nearly all specialties have become increasingly involved in issues of fertility control. Practitioners ranging from pediatricians who care for teenage girls to internists who care for premenopausal women are frequently asked to provide contraceptive advice.

The role of medical specialists in reproductive health care is also undergoing a transition. Two major challenges have arisen. The first is a result of the existing and widening cultural diversification in developed countries, where physicians frequently face problems that not only challenge their medical skills but also test their ability to deal with the family planning needs of communities with diverse ethnic, cultural, and religious backgrounds. The second challenge has to be met by all those working in reproductive health care and family planning: the increasing emphasis on state-of-the-art medical practice. There is a global need for the rapid incorporation of appropriate technologies into daily clinical practice, substantially increasing the quality of rational medical care, worldwide.

The individual methods and technologies of fertility control are discussed in the following sections of this chapter.
**Methods of Fertility Control: Efficacy, Continuation of Use, Changing Trends**

When making a recommendation to a candidate for contraception, three aspects of each method have to be examined: efficacy, tolerability, and whether the method is suitable for temporary or permanent cessation of procreation. Table 17-1 lists the principal methods of fertility control currently available and indicates their contraceptive efficacy as well as the average time patients adhere to an individual method.

The estimates of contraceptive failure during "perfect" use are derived from studies conducted for research and registration purposes; they include volunteers or self-selected groups who are highly motivated to adhere to the study protocol. Therefore, the results are more favorable than those obtained during "typical" use, that is, use in the general population. Data on the efficacy for typical use are generated by the National Surveys of Family Growth, among others.

**Efficacy and Continuation**

If the outcome was left to chance—that is, no method of fertility control was practiced—5 of 100 women of reproductive age would become pregnant within a year. Surgical sterilization, for both men and women, remains one of the most effective methods of contraception, although it is not failure proof. The small proportion of early failures may be associated with suboptimal surgical techniques, such as mistaken anatomical structures for the fallopian tube or vas deferens; insufficient electrocoagulation when this method is used in women; or unprotected early intercourse after vasectomy while live spermatozoa are still present in the part of the vas deferens that is distal to the interruption. Late failures, between the third and tenth years after surgery, also occur and are discussed in the section on sterilization. The major advantage of surgical sterilization is that it is permanent and can be a welcome solution to parents with large families. Chemical sterilization of women aims at producing tubal occlusion by injection of a substance (e.g., quinacrine) into the junction between the tube and the uterus. Data from a large study conducted in Vietnam show that this method is inferior to surgical sterilization.

Among the hormonal contraceptive methods, long-acting approaches such as implants, injectables, and copper- or hormone-containing intrauterine devices (IUDs) achieve a higher level of efficacy than oral contraceptives (OCs) that require a daily conscientious action by the user. According to revised data of 1995, the failure rates of OC formulations are higher than previously thought. Emergency postcoital contraception, when properly used, is also effective; however, it is recommended only as an emergency provision.

Hormonal methods of fertility control also encompass non-surgical termination of pregnancy by pharmacologic means ("contragestion"). Mifepristone and prostaglandins administered not later than the seventh week of pregnancy achieve complete abortion in a high proportion of cases. The procedure is less effective when employed after the seventh week of pregnancy.

The efficacy of IUDs has been improved by incorporating either progesterone or a synthetic progestogen, levonorgestrel. The copper IUD has staged an impressive revival, principally outside the United States. Its efficacy and safety have been improved and are now comparable to those of hormonal contraception. Copper- and progesterin-bearing IUDs are likely to continue to grow in popularity.

Other methods of fertility control are less effective. For the barrier methods, there is a large discrepancy between the efficacy of ideal use and that of typical use. With the cervical cap and the vaginal sponge, failure rates are high even with perfect use. The same applies to the "natural" methods of fertility control. They rely, one way or another, on prediction or detection of ovulation and require continuous watchfulness and motivation of both partners. Therefore, their efficacy does not compare favorably with that of the hormonal or intrauterine

---

**TABLE 17-1 – Fertility Control Methods: Failure Rates and Continuation of Use (United States Data)**

<table>
<thead>
<tr>
<th>Method</th>
<th>Percent Pregnant during First Year of Use</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimates 1987 and 1990, All Methods</td>
</tr>
<tr>
<td></td>
<td>Perfect use</td>
</tr>
<tr>
<td>Chance</td>
<td>85</td>
</tr>
<tr>
<td>Sterilization</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.1</td>
</tr>
<tr>
<td>Female</td>
<td>0.2</td>
</tr>
<tr>
<td>Surgical</td>
<td></td>
</tr>
<tr>
<td>Chemical (quinacrine)</td>
<td></td>
</tr>
<tr>
<td>Women &lt;35 y</td>
<td>13.0</td>
</tr>
<tr>
<td>Women 35 y</td>
<td>7.0</td>
</tr>
<tr>
<td>Hormonal contraception, emergency contraception, and contragestion</td>
<td></td>
</tr>
<tr>
<td>Combination pill</td>
<td>0.1</td>
</tr>
<tr>
<td>Progestagen-only pill</td>
<td>0.5</td>
</tr>
<tr>
<td>Norplant</td>
<td>0.05</td>
</tr>
<tr>
<td>Depo-Provera</td>
<td>0.3</td>
</tr>
<tr>
<td>Emergency contraceptionhormonal</td>
<td>0.1</td>
</tr>
</tbody>
</table>
Contraception

Pharmacologic abortion

1.05.0 (up to 7 wk) 9.0 (> 7 wk) ND ND

Intrauterine devices (IUDs)

- IUD-progesterone T 1.5 2.0 80
- IUD-levonorgestrel 20 0.1 0.1 ND 81
- IUD-T 380 (copper) 0.6 0.8 78

Barrier methods

- Condom
  - Male 3.0 14.0 14.0 63
  - Female 5.0 21.0 ND 56
- Diaphragm
  - 6.0 20.0 18.0 58
- Cervical cap
  - Parous women 26.0 40.0 12.0 42
  - Nulliparous women 9.0 20.0 56
- Sponge
  - Parous women 20.0 40.0 42
  - Nulliparous women 9.0 20.0 ND 56
- Spermicides
  - 6.0 26.0 26.0 40
- Withdrawal
  - 4.0 19.0 24.0 ?
- Periodic abstinence
  - Calendar 9.0 ? 63
  - Ovulation method 3.0 ? 21.0
  - Postovulation 1.0 ?
  - Symptothermal 2.0 ?

Lactational amenorrhea provides an effective but temporary method of contraception

ND, no data.

*Data from references [17][18][19].

New estimates of contraceptive failure according to correction for abortion underreporting, from 1995 National Survey of Family Growth. [20]

Data on quinacrine sterilization from reference [21].

§Periodic abstinence methods. Calendar: The woman records the length of 6 to 12 cycles and determines the beginning of the fertile period by subtracting 18 days from the shortest cycle. The end of the fertile period is estimated by subtracting 11 days from the longest cycle. Ovulation method: Women are taught to recognize the characteristic of the cervical mucus during the fertile period. Sexual abstinence begins on the day when the mucus becomes clear, slippery, and stretchy (usually a few days before ovulation). At the peak of the fertile period, the mucus stretches between the finger and thumb to the maximum (spinnbarkeit). Intercourse is resumed on the third or fourth day after the peak mucus. Symptothermal method: Cervical mucus (ovulation) method supplemented by calendar in the preovulatory phase and basal body temperature in the postovulatory phase. Postovulatory method: Ovulation is estimated by the basal body temperature or cervical mucus methods; unprotected intercourse is allowed 3 to 4 days thereafter. Note: Sophisticated electronic devices are currently available to estimate ovulation more precisely, some of them measuring the estrogen and luteinizing hormone concentrations in urine.

Lactational amenorrhea affords protection against pregnancy; however, breakthrough ovulations do occur, particularly when breast-feeding is not exclusive and is interrupted by other types of baby nourishment.

Women frequently discontinue the use of contraception. Overall, 31% of women discontinue use of reversible contraceptives for a method-related reason within 6 months of use; 44% do so within 12 months. However, 68% of women overall resume use of a method within 1 month and 76% do so within 3 months. High rates of method-related discontinuations reflect dissatisfaction with available methods or management, or both. [22]

Trends in Fertility Control in the United States

In the third quarter of the last century, the use of fertility control methods in the United States underwent a major change (Table 17-2) [23][24][25].

Among women, hormonal contraception has remained the most popular method. Two events positively influenced the employment of hormonal methods: the advent of hormonal implants and the approval of medroxyprogesterone acetate (MPA) depot injections for contraception.

As physicians became versatile in laparoscopic surgeries, the
frequency of female sterilization jumped from 12% in 1973 to 23% in 1982; 1995 brought a further increase to nearly 30%. The proportion of male sterilization remained stable at 11%.

The condom now ranks as the third method of choice. Between 1982 and 1995, the frequency of its use rose from 12% to 27%, most markedly among unmarried women in the age group 15 to 29 years. Fear of human immunodeficiency virus (HIV) and other sexually transmitted diseases (STDs) impelled this shift.

Possibly the most dramatic change in the use of contraceptive methods was the virtual abandonment of IUDs in the United States. This was triggered by the Dalkon Shield episode ending in 1973—the occurrence of excess pelvic infection in women using the device. Another negative factor was the recall of the "copper 7" IUD from the market in 1986 because of putative infertility of women who had discontinued its use. A newly designed copper IUD has surfaced, although U.S. physicians are still more hesitant to prescribe it than their colleagues outside the United States. Uptake of the hormone-bearing IUD is similarly slow.

Contraceptive foams alone are used rarely; currently, manufacturers and physicians recommend that they be used simultaneously with barrier methods.

### TABLE 17-3 -- Fertility Control in the Developing World (Percent Users), Trend 1980-1993

<table>
<thead>
<tr>
<th>Method</th>
<th>1980</th>
<th>1993</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female sterilization</td>
<td>24</td>
<td>39</td>
</tr>
<tr>
<td>Intrauterine device</td>
<td>32</td>
<td>26</td>
</tr>
<tr>
<td>Pill</td>
<td>13</td>
<td>8</td>
</tr>
<tr>
<td>Male sterilization</td>
<td>13</td>
<td>8</td>
</tr>
<tr>
<td>Injectables</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Vaginal</td>
<td>0</td>
<td>0.3</td>
</tr>
<tr>
<td>Condom</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Traditional</td>
<td>12</td>
<td>9</td>
</tr>
</tbody>
</table>


Surgical female sterilization is the most frequently used method of effective fertility control in the developing countries (Table 17-3). The biggest difference between the United States and developing countries is the high use of IUDs; use by approximately 30% of women in developing countries makes IUD insertion the second most frequent method of contraception. Hormonal contraception ranks third, possibly because of its high cost and problems with access. In Sri Lanka, for instance, some women buy only five pills at a time because their financial situation does not allow them to acquire a full month’s supply.

### Contraceptive Failures

Given the length of time that most women practice a reversible method of fertility control, experiencing at least one contraceptive failure is likely. By sheer statistical probability, failure can happen even if the most effective methods are used perfectly. Between the ages of 16 to 30, a hormonal contraception user has a more than 50% chance of becoming pregnant. During a lifetime of use of reversible methods, the typical woman experiences 1.8 contraceptive failures. This applies particularly when the health status of a woman prohibits the use of a highly effective method, such as the pill, and the couple has to resort to methods that have fewer adverse effects but may also be less effective.

A National Survey of Family Growth analyzed the outcome of unplanned pregnancies in the United States that occurred despite contraception between 1994 and 1995. The analysis took into account the total number of births and therapeutic abortions in the year 1994. Miscarriages were excluded from analysis because it was difficult to calculate the proportion of women who did not plan the pregnancy. For purposes of the analysis, the authors assumed that all therapeutic abortions resulted from unintended pregnancies.

In 1994, there were 5.38 million pregnancies in the United States; of those, 2.73 million (51%) were intended and resulted in births of live children. A full 2.65 million (49%) pregnancies were unintended, of those, 1.22 million or 46% resulted in births and 1.43 million or 54% ended as therapeutic abortions. Alternatively, of the total 3.95 million births (100%), the majority of 2.75 million or 70% were intended and 1.22 million or 30% were unintended.

The survey further revealed that 48% of women from 15 to
### Table 17-4

<table>
<thead>
<tr>
<th>Method</th>
<th>0.2</th>
<th>0.9</th>
<th>1.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sterilization</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Implant</td>
<td>0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intrauterine device</td>
<td>0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Injectable</td>
<td>0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pill</td>
<td>11.7</td>
<td>20.3</td>
<td></td>
</tr>
<tr>
<td>Male condom</td>
<td>32.4</td>
<td>56.6</td>
<td></td>
</tr>
<tr>
<td>Withdrawal</td>
<td>5.9</td>
<td>10.3</td>
<td></td>
</tr>
<tr>
<td>Other*</td>
<td>6.4</td>
<td>11.3</td>
<td></td>
</tr>
<tr>
<td>No contraception</td>
<td>42.5</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>100.0</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

NA, not applicable.

*Female condom, diaphragm, sponge, foam, suppository, periodic abstinence.

---

From Henshaw SK, Kost K. Abortion patients in 19941995: characteristics and contraceptive use. Fam Plann Perspect 1998; 28:140147, 158.

44 years of age had had one or more unintended pregnancies, either an unplanned birth or an abortion, or both. The percentage increased with age to a high of 60% among women in the age group 35 to 39 years.

The information on the use of contraceptive methods at the time when an unplanned pregnancy was conceived and later resulted in therapeutic abortion is based on a representative subset of 9885 patients who had abortions. Of these, 42.5% did not use any contraception at the time of conception and 57.5% practiced some method of fertility control. The male condom was most frequently associated with contraceptive failure. More than half of the women who resorted to abortion reported the use of condoms.

Among women who experienced contraceptive failure, 76% of those younger than 18 years preferred the condom. However, only 46% of women older than 30 years used this method. Pill use peaked at 25% among women aged 20 to 29 and decreased to 16% after the age of 30. On the other hand, the use of methods classified as other rose from 1% in the age group younger than 18 years to 24% in the age group 30 years or older.

The Need for Future Improvements

The high discontinuation rate, the number of unintended pregnancies, and the number of abortions testify to the fact that present reproductive health care is still deficient in providing adequate fertility control options to all women. The health care system alone cannot be blamed for the large number of unplanned pregnancies. Patients frequently choose methods of low efficacy, the primary example being the use of the male condom. When effective methods are recommended, they are frequently applied incorrectly or their use is interrupted, even if it is resumed later on.

These problems can be partially remedied by appropriate information about the available contraceptive choices and by intensive counseling adjusted to the level of the individual patient. Informing the general public about new findings and advances in reproductive health care has been largely facilitated by access to electronic sources of information; the medical profession should take full advantage of this progress.

The proportion of women who did not use any contraception at the time of an unplanned conception is disturbing. This problem could be partially addressed by the development of a highly effective postcoital method. An alternative solution would be implementation of an effective system that would make the presently available postcoital contraceptives promptly accessible. Finally, there must be renewed interest in improving present methods and designing new effective and safe approaches to fertility control.
Hormonal Contraception

The knowledge that estrogens and progestagens can inhibit ovulation has been the foundation of the modern hormonal methods of contraception. In 1940, for the first time, Sturgis and Albright achieved inhibition of ovulation in women by estrogens in order to relieve dysmenorrhea.

The foundation of modern contraception was laid in Manhattan in 1950 at a lengthy conference. Among the participants were Gregory Pincus, director of the Worcester Foundation, and a small group of fertility control advocates including Margaret Sanger, the founder of the Planned Parenthood movement. The conference ended by granting Pincus seed money of $2100 to initiate his research on hormonal contraception. In 1953, working with a Boston gynecologist, John Rock, Pincus started to test oral progesterone for ovulation inhibition. In 1956, Pincus’ group of scientists and physicians published results of the first successful contraceptive clinical trial, which took place in Puerto Rico using a synthetic progestagen, norethynodrel, provided by the G.D. Searle company. Norethynodrel, however, was not the first orally active progestagen. This priority belongs to norethindrone, synthesized in 1951, which became the lead compound in the development of clinically important oral progestagens.

Starting in the 1930s, chemists searched for orally active steroids. In 1938, a group at the Schering Corporation in Berlin, Germany, developed the first orally active steroidal estrogen by attaching the ethinyl group (-CCH) to the 17th carbon of the estradiol molecule (Fig. 17-6). Ethinylestradiol still constitutes the estrogenic component of nearly all combined OCs. The development of orally active progestagens was triggered by the discovery that when the C-19 methyl group of testosterone is split off, the resulting molecule loses androgenicity and acquires progestagenic properties. In 1951, Carl Djerassi, leading a team of chemists in Mexico City, reasoned that attaching the ethinyl group (-CCH) to the 17th carbon of the nortestosterone molecule would greatly enhance its progestagenic activity and make the compound orally active (Fig. 17-7 and Fig. 17-8). Djerassi and colleagues produced 17-ethinyl-19-nortestosterone. The compound, known by its generic names norethindrone and in Europe norethisterone (commonly abbreviated NET), had 10 times the activity of natural progesterone when orally administered. To this day, norethindrone has remained the progestagenic component of many combined OCs (Table 17-6). In a parallel development, manipulation of the progesterone molecule produced a group of orally highly active progestagens, such as medroxyprogesterone acetate (MPA) (Provera) (see Fig. 17-16 later).

Originally, hormonal contraception included only the pill, that is, an orally active combination of an estrogen and a progestagen. The pill offered several advantages over other available contraceptive methods. For the first time, its high efficacy brought confidence in a contraceptive method and freedom from the fear of pregnancy. It produced only temporary and reversible infertility; therefore, it was ideal for family planning. Because the method was not linked to coitus, the spontaneity of the sexual act could be preserved. The pill has become popular among women in both developed and developing countries, and today 10 million women in the United States and 60 million women worldwide adhere to this method.

However, use of the pill requires a conscious daily action on the part of the user, and it has been postulated that missing pills could be the reason for the discrepancy between pregnancy rates during perfect and typical use (see Table 17-1). To close this gap, long-term methods of hormonal contraception have been invented that require only one or a limited number of actions on the part of the user. Hormonal implants, depot injections, and hormonal IUDs are examples of such long-term methods with high efficacy during typical use.

Unintended pregnancies can result from failure of contraceptives or from exposure to an unplanned sexual contact. Such events necessitate short-term preventive steps. Two important developments took place in this direction: emergency post-coital contraception and the use of anti progestagenic steroids for contraception.

The hormonal contraceptives available today can be categorized by the way they are administered and by the duration of their action (see Table 17-5).
Oral Contraception

Steroidal Components

In the following paragraphs we describe the pharmacologic and biologic properties of the two hormonal components of OCs, the estrogens and the progestagens.

Contraceptive Estrogens

Structure and Function

In the combined estrogen-progestagen OCs, one type of estrogen prevails, 17-ethinylestradiol. A limited number of OC preparations contain a derivative of ethinylestradiol, mestranol.

The ethinyl group protects ethinylestradiol and mestranol (MEE) from oxidation and conversion into less active estrogens such as estriol. Mestranol is a prodrug that does not bind to estrogen receptors and has to be converted into ethinylestradiol in order to become biologically active. MEE has been replaced by ethinylestradiol in virtually all OCs worldwide.

Pharmacokinetics and Metabolism

After ingestion, ethinylestradiol is rapidly absorbed from the gastrointestinal tract. The time to the maximum ethinylestradiol concentration in plasma ($T_{\text{max}}$) is 1 to 2 hours and the elimination half-life is wide, ranging from about 9 to 27 hours. In the intestine and in the liver, ethinylestradiol is readily conjugated with sulfuric and glucuronic acids and undergoes enterohepatic circulation. Intestinal bacteria possessing the appropriate enzymes can hydrolyze ethinylestradiol sulfates, and some of the deconjugated estrogen is reabsorbed. One could speculate that the use of oral antibiotics that affect intestinal flora may influence blood levels of ethinylestradiol; so far, clinical proof of this assumption is lacking. The pharmacokinetics and metabolism of mestranol are similar to those of ethinylestradiol except that $T_{\text{max}}$ is longer than might have been expected because mestranol has to be converted to ethinylestradiol.\[37\]

Individual subjects vary considerably in the amount of absorbed ethinylestradiol and the circulating concentrations of ethinylestradiol as well as in the elimination time. Intrasubject variations can also be prominent. These variations may explain why adverse effects and contraceptive failures occur in only certain individuals and why the same woman can experience side effects during some treatment cycles but remains symptom-free during others. The basis of the intersubject and intrasubject variations has not been explained satisfactorily.

As we are witnessing another wave of reduction of the ethinylestradiol content in the combination pills to 20 µg, bioavailability becomes an issue. As these and ever lower doses of ethinylestradiol become more common, we must be certain to provide OCs in formulations that deliver the digested amounts of steroids at appropriate concentrations for the individual user.

Contraceptive Progestagens

In contrast to estrogen synthesis, the synthesis of progestagens has been prolific. The reasons for this dichotomy are several. Because ethinylestradiol exhibited potent oral activity and was not protected by patents, there was little incentive to search for other OC estrogens. New progestagens were synthesized in order to secure proprietary rights for the developers of OCs. Also, synthesis of new compounds was prompted by the desire to produce a highly effective progestagen that would minimize the dose of the progestagen in the OC combination and thus reduce the incidence of progestagen-related adverse events. Currently, about a dozen progestagens are being used clinically in established contraceptive preparations and several compounds are in various stages of preclinical and clinical testing.

According to the classical definition, progestagens transform the estrogen-primed endometrium into a secretory one and support the development and maintenance of pregnancy. The advent of molecular biology has defined progestagens as compounds that bind to and activate progesterone receptors within the target cells.

However, binding to progesterone receptors does not preclude the progestagen molecule binding with other receptors or expressing effects other than progestagenic effects, or both. For example, norethindrone, under certain circumstances, can stimulate the proliferation of the atrophic endometrium.\[38\]\[39\] Besides being a potent progestagen, cyproterone

### Table 17-5 — Hormonal Contraceptives in Clinical Use

| Oral contraception (see details in Table 17-6) |
| Cyclic estrogen-progestagen combinations |
| Continuous progestagen-only oral contraception |
| Long-acting preparations |
| Injectable preparations |
| Progestagen-only preparations |
| Depo-medroxyprogesterone acetate, 150 mg, q3mo, IM |
| Norethindrone enanthate, 200 mg, q2mo, IM (not used in the United States) |
| Medroxyprogesterone combination injectables |
Norplant: 6 capsules with total amount of 216 mg of levonorgestrel
Norplant II (Jadelle): 2 rods with total amount of 150 mg of levonorgestrel

Hormonal intrauterine systems

Progestasert: T-shaped; vertical arm releases progesterone, 65 µg/d
Mirena (levonorgestrel-20): T-shaped; vertical arm releases levonorgestrel, 15 µg/d
Vaginal rings releasing contraceptive hormones
NuvaRing approved in 2001
Transdermal patch releasing contraceptive hormones
Ortho EVRA: 20 cm² patch, delivers 150 µg noretisteron, 20 µg/d ethinylestradiol

Emergency methods of fertility control

Postcoital hormonal contraception
a. 1 mg norgestrel and 100 µg ethinylestradiol within 72 h of unprotected intercourse, repeated after 12 h
b. 0.75 mg levonorgestrel within 72 h of unprotected intercourse, repeated after 12 h

Contraception

Mifepristone, 200-600 mg, orally, followed by 400-800 µg misopristol orally

---

Figure 17-9
Classification of contraceptive steroids. Hybrid progestagens are in gray boxes. In parentheses are the active progestagenic metabolites of norgestimate and desogestrel.

Figure 17-8
Classification of contraceptive progestagens. Hybrid progestagens are in gray boxes. In parentheses are the active progestagenic metabolites of norgestimate and desogestrel.

Classification of Contraceptive Progestagens

Contraceptive progestagens have been classified in various ways, for example, according to the amount of ethinylestradiol with which they are combined, the type of progestagen they contain, and the time when they became available for clinical use.

According to the chemical structure of the steroids, the classification of contraceptive progestagens presented recognizes three basic groups, 

1. pregnanes,
2. estranes, and
3. gonanes.

These chemical reactions led to the synthesis of norethindrone, the first orally highly active progestagen, which enabled the development of oral contraception. All estranes are 19-norsteroids.

Gonanes

The gonane structure lacks both the C-18 and the C-19 angular methyl radicals. However, all gonane progestagens bear an ethyl group between rings C and D at C-13 (Fig. 17-12 and Fig. 17-13; see Fig. 17-8). This chemical modification makes them more active progestational agents than estranes.

---

Norplant and levonorgestrel are derivatives of 13-ethylgonane. Because they were developed in the 1960s, they are categorized as early gonanes. Only the
dl-Norgestrel and levonorgestrel were synthesized in order to acquire a more potent progestagen than norethindrone, but they also increased the compound's androgenic properties with undesired clinical and metabolic effects. Thus, the next task of the steroid chemist was to produce highly active progestagens without the androgenic effects. The synthesis of the later gonanes met this challenge.

Later Gonanes.

Three progestagens belong to this family, norgestimate, desogestrel, and gestodene (see Fig. 17-13).

**Norgestimate.**

The key difference between norgestimate and other progestagens is the oxime group (-C=N-OH) in position 3 of the molecule instead of the keto group (C=O). Norgestimate is solely the biologically active levor form. Because the C-3 keto group is typical of androgenic compounds, its replacement by the oxime may contribute to the reduced androgenicity of norgestimate as compared with levonorgestrel.

**Desogestrel.**

This advanced gonane is interesting in that its progestagenic activity has been increased by substitution on C-11. Manipulations at C-11 give the steroid molecule the ability to bind to the progesterone, glucocorticoid, and mineralocorticoid receptors. The intensity of the binding depends on the structure of the substituting groups.

**Gestodene.**

Gestodene differs from norgestrel in a single feature, namely the double bond between C-15 and C-16. This seemingly simple change has a profound effect on the configuration of the molecule, principally on the spatial arrangement of the D ring and C-17. One can speculate that these changes affect the conformation of the molecule in a way that affects its binding to hormone receptors. In vitro studies have shown that gestodene binds to the heme of the P450 enzymes that inactivate estrogens, with consequent increased concentrations of these hormones. The clinical relevance of this finding is unknown.

**Some Aspects of the Clinical Pharmacology of Contraceptive Progestagens.**

The orally active contraceptive progestagens are rapidly absorbed from the digestive tract and are transported to the liver through the portal circulation. Thereafter, they enter the general circulation, where they form a hormonal pool. In the general circulation, progestagens are present either in the free form or bound to albumins or to the sex hormone-binding globulin (SHBG), or both. Only the free form reaches receptors in the respective target tissues and becomes biologically active. Progestagens can easily be released from the steroid-albumin complex; the bond with SHBG is firmer.

During first-pass liver metabolism, progestagens can be structurally modified with consequent changes of their biologic activity. Glucuronidation and sulfuration of progestagens facilitate their excretion by the kidneys. A variable fraction of progestagens is excreted in the feces. During first-pass liver metabolism, progestagens act on hepatocytes in ways that can alter their metabolism (see later).

The pharmacologic properties of progestagens determine their clinical use. Progestagens of the pregnane series can bind strongly to progestagenic as well as to androgenic receptors and can act as antiestrogens. Oral MPA is used for treatment of various gynecologic disorders, for example, for the management of dysfunctional uterine bleeding, and in the menopause. The compound has been developed as the first injectable long-term contraceptive. In the past, medroxyprogesterone and other pregnane derivatives were proscribed for contraceptive use in the United States because preclinical toxicology had shown an accelerated development of benign and malignant breast nodules in beagle dogs as breed that suffers spontaneously from a high incidence of breast tumors, including carcinoma. Extensive clinical trials conducted by the WHO have lifted the cloud of potential carcinogenicity hanging over MPA, and it is used as an injectable contraceptive globally.

The accelerated growth of breast nodules in beagle dogs was not observed in toxicologic studies with progestagens of the estrane and gonane series. The reasons for this difference in animal carcinogenic potential of pregnanes versus estranes and gonanes have not been adequately elucidated. It is noteworthy that long-term toxicologic studies in monkeys have not shown the formation of any type of breast nodules. In monkeys, spontaneous breast carcinoma is unknown, and the induction
of hormone-associated breast pathology would have been particularly important.

Of the other pregnanes, acetates of chlormadinone, cyproterone, and megestrol are used in OCs in some countries outside the United States. They are also used for management of breast, endometrial, and prostatic carcinomas. Because they are also highly effective antiandrogens, they have been part of the management of benign prostatic hypertrophy, prostatic carcinoma, precocious puberty, and certain hyperandrogenic symptoms in women.

In the estrane series, the four derivatives of norethindrone are rapidly converted to norethindrone. Within 30 minutes after ingestion of any of the derivatives, only norethindrone can be detected in the general circulation.

The earliest OCs contained norethindrone and norethynodrel, originally combined with up to 150 µg of ethinylestradiol. Today, OCs containing more than 50 µg of ethinylestradiol are considered "high-dose" estrogen preparations and have been removed from the OC market in the United States. "Mid-dose"

The early gonanes, norgestrel and levonorgestrel, display certain unique features. The time required for the circulating levels of levonorgestrel to decline by 50% is about 15 hours, and for norethindrone it is about 7 hours. The difference in the elimination time is one of the reasons that contraceptive doses of levonorgestrel can be lower than those of norethindrone. Norgestrel and levonorgestrel are strong progestagens with antiestrogenic properties; however, they also show some androgenicity. Levonorgestrel also decreases the plasma concentration of SHBG by 50% and, in combined OCs, suppresses the estrogen-induced formation of SHBG. Consequently, less SHBG is available for binding testosterone. A combined OC composed of 150 µg of levonorgestrel and 30 µg of ethinylestradiol increases the levels of SHBG slightly by about 20% from baseline. In contrast, women using norgestimate or desogestrel-ethinylestradiol OCs have a three-fold increase in circulating levels of SHBG, which results in a 50% decrease of free testosterone.

Compounds of the late gonane series, norgestimate and desogestrel, are metabolized extensively. After ingestion, norgestimate is rapidly converted into noregestromin and desogestrel is converted into etonogestrel. These metabolites account for the biologic activity of the parent compounds. Noregestromin has been synthesized as a specific hormone for the contraceptive patch, and etonogestrel is used in silastic implants for contraception.

Gestodene is the only compound of the gonane series that is not rapidly metabolized. It is a highly potent progestagen; however, at the time of this writing the compound has not been approved for clinical use in the United States.

**Hybrid Progestagens**

**19-Norpregnanes**

The 19-norpregnanes are a cross between the pregnanes and estranes. These compounds are derived from 17-hydroxyprogesterone acetate but lack the C-19 methyl radical and in that respect are related to estranes. *Noregestrol*, an important member of this series, is currently being investigated as a contraceptive implant.

*Nestorone* is another 19-norpregnane of the 17-acetoxyprogesterone series, which has a methylene group on C-16. In receptor assays, nestorone had progestational effects equal to or better than those of levonorgestrel without estrogenic, androgenic, and anabolic activities. However, nestorone binds to glucocorticoid receptors. Nestorone has low oral but high parenteral progestational activity. Therefore, the Population Council is studying nestorone as a contraceptive implant in subdermal implants, vaginal rings, and transdermal formulations. The compound may be suitable for use in nursing mothers because of its low oral bioavailability.

**Other Hybrid Progestagens**

Dienogest is a new addition to the estranes. In this compound, the cyanomethyl group (CN) has replaced the C-17 ethinyl group (CH₂). Dienogest is 100% orally available. Some evidence suggests that the compound lacks androgenic activity and produces less glucocorticoid antagonism than mefiopristone.

The compound suppresses endometrial growth and is being tested in the management of endometriosis. The antiestrogenic activity of Dienogest is relatively low, mandating the use of 2 mg in combination with 30 µg of ethinylestradiol in a 21-day cyclic regimen.

**Drospirenone**

An addition to contraceptive progestagens has been *drospirenone*. It is a progestagen derived from spironolactone, a potent steroid with antimineralocorticoid activity, which also has progestagenic properties. Drospirenone is relatively weak progestagen; a daily dose for a contraceptive regimen is 3 mg combined with 30 µg of ethinylestradiol. The compound has antiandrogenic and antiimineralocorticoid properties and causes potassium retention. Therefore, it should not be taken by patients with kidney, liver, or adrenal gland disease or with other drugs that increase potassium concentrations in the circulation. Such drugs include nonsteroidal anti-inflammatory drugs, potassium-sparing diuretics (spironolactone and others), potassium supplementation, angiotensin-converting enzyme inhibitors, angiotensin II receptor antagonists, and heparin.

**Progestagenic Potency, Androgenicity, and Comparative Metabolic Effects of Oral Contraception**

**Multiple Steroid Actions**

Significant determinants of the biologic activity of a steroid include its bioavailability to the target tissues and its affinity for relevant receptors. The multifaceted nature of the steroid molecule is illustrated by its capacity to bind to several different receptors and activate them to various degrees. The final biologic activity depends on the proportion of activated receptors. Progestagenic, estrogenic, androgenic, and glucocorticoid actions. Therefore, the most appropriate progestagens are found in the progestogen androgen receptors. Moreover, some progestagens bind to glucocorticoid receptors.

**Progestagenic and Androgenic Activity**

It is difficult to establish the relative progestagenic and androgenic potency of progestagens because systematic testing of all progestagens by one method has not been performed. Different progestagens have been tested by different methods. The relative binding affinities of progesterone and gonanes for rabbit uterine
progestagen receptors are shown in Figure 17-16. The binding affinity of norgestimate and its 17-deacetylated metabolite (norelgestromin) was similar to that of progesterone; levonorgestrel was at about five times and gestodene and 3-ketodesogestrel were about nine times more active than progesterone.

The relative binding activities of progestogene and gonane progestagens for rat prostatic androgen receptors are depicted in Figure 17-16. Norgestimate and norelgestromin, which have the same progestagenic activity as progesterone, display low androgenic activity. Levonorgestrel, a 5 times more potent progestagen than progesterone, has 44 times higher androgenic activity than progesterone. Although 3-ketodesogestrel and gestodene are more active progestagens than levonorgestrel, they have less androgenic activity than levonorgestrel. In this assay, the androgenic activity of dihydrotestosterone is 200-fold greater than that of progesterone (not shown in Fig. 17-16). The androgenic activity of levonorgestrel is one fourth the activity of dihydrotestosterone.

Biologic tests of androgenicity were conducted on castrated rats, the end point being the weight increase of the prostatic gland. In this assay, the androgenic potency of testosterone equals 100%, that of levonorgestrel is 15% and that of norethindrone is only 1.6%. Medroxyprogesterone and chlormadinone acetate displayed no androgenic action in this test.

NonReceptor-Mediated Action

In skin and some other tissues, testosterone is a prohormone and becomes biologically active only after conversion into dihydrotestosterone by 5-reductase. In vitro experiments have demonstrated that norgestimate and desogestrel inhibit the action of 5-reductase. This may partially explain the beneficial effects of these progestagens in the management of androgenic skin lesions such as acne.

Lipid Metabolism

Important differences exist among progestagens in their effect on lipid metabolism, particularly on the cardioprotective lipoproteins. Estrogens increase total cholesterol, but they also increase the high-density lipoprotein (HDL) fraction and decrease the low-density lipoprotein (LDL) fraction of cholesterol. Progestagens exert an antagonistic effect on these positive actions of estrogens by various mechanisms, including an increase in the activity of hepatic lipase, which degrades HDL. There are quantitative differences among individual progestagens, however, and the net metabolic effect depends on an intricate interplay between the two components of the combined contraceptives, the type and dose of the progestagen, and the treatment regimen, whether it is monophasic or triphasic. In general, the higher the androgenic properties of a progestagen, the more pronounced are the negative effects on cardioprotective lipoproteins.

Because norethindrone and estrane progestagens are derived from testosterone, and testosterone is part of the chemical name of the compounds, it is sometimes assumed that norethindrone and its analogues have androgenic properties. In doses and combinations used in current clinical practice, norethindrone and its analogues do not exhibit substantial clinical or metabolic androgenic effects. A monophasic combination OC (0.5 μg of norethindrone plus 35 μg of ethinylestradiol per day) has been associated with a 10% increase in HDL and a 10% decrease of LDL.

With respect to gonanes, levonorgestrel increases LDL only slightly but reduces HDL significantly. Norgestimate, a later gonane, significantly elevates HDL with a nonsignificant effect on LDL. (Fig. 17-17)

Protein Binding

Natural and synthetic sex steroid hormones enter the target cells by passive diffusion. The capacity of sex steroids to reach receptors in these target cells is modulated by SHBG and other proteins, such as albumins. Natural sex steroids and some synthetic progestagens bind to SHBG with higher affinity and specificity than they bind to albumin. As long as a sex steroid is bound to SHBG, it cannot affect its biologic action, which is accomplished by the free or non-SHBG-bound fraction of the steroid. The binding of sex steroids to albumin is less tight, and albumin-bound steroids are more readily available to target cells, ensuring a rapid pharmacologic response.

Estrogens stimulate hepatocytes to produce SHBG, and androgens and some progestagens interfere with this action. Depending on their composition, OCs are associated with increased formation of SHBG and reduced levels of free.

Insulin and Carbohydrate Metabolism

The original high-dose OCs were associated with insulin resistance and glucose intolerance. Glucose tolerance tests showed a significant increase of blood glucose and insulin after 1 year of OC use. It was initially thought that these changes in glucose tolerance were related to the estrogenic component of the pill. However, after tests of estrogens alone, even in high doses, demonstrated no such negative effects, it was found that high doses of many progestagens can impair carbohydrate metabolism. With the advent of low-dose OCs, the effect of contraceptive hormones on carbohydrate metabolism has been minimized.

Mechanism of Contraceptive Action

Originally, it was assumed that the contraceptive action of combined steroid hormones primarily involved inhibition of the pituitary gonadotropins with consequent blocking of ovulation. This is certainly the case with the combined estrogen-progestagen OC. However, blood levels of progesterone and pituitary gonadotropins indicated that during progestagen-only contraception, ovulatory function has been preserved during many cycles. Therefore, other mechanisms of contraceptive action were explored. Among these were changes in the consistency and increases in the thickness of the cervical mucus that impair the ability of the sperm to penetrate it. Low doses

Figure 17-15 Relative binding affinities of contraceptive progestagens for progesterone receptors. The assay measures displacement of H-labeled R5020 from progestagen receptors isolated from the rabbit uterus. The binding affinity of contraceptive progestagens for progesterone receptors is shown in Figure 17-15. The androgenic activity of levonorgestrel is about one fourth the activity of dihydrotestosterone. (Data from Phillips A, Demarest K, Hahn DW, et al. Progestational and androgenic receptor binding affinities and in vivo activities of norgestimate and other progestins. Contraception 1990; 41:399.)

Figure 17-16 Relative binding affinity of contraceptive progestagens for androgen receptors. The assay measures displacement of H-labeled dihydrotestosterone from rat prostatic androgen receptors. (Data from Phillips A, Demarest K, Hahn DW, et al. Progestational and androgenic receptor binding affinities and in vivo activities of norgestimate and other progestins. Contraception 1990; 41:399.)
of progestagens also alter the function of the endometrium, leading to an intruterine milieu hostile to pregnancy.\textsuperscript{119}

**Oral Contraceptive Treatment Regimens**

**Combined Oral Contraception: Cyclic Estrogen-Progestagen Combinations**

In this method both hormonal components are given in a cyclic fashion, usually from the 5th through the 25th day of the menstrual cycle. With most of the currently used preparations, a 7-day placebo period follows the 21-day hormonal treatment so that women do not need to keep track of when to start a new cycle of contraceptive pills. There is growing interest in developing combined oral contraception (COC) for continuous use to avoid monthly withdrawal bleeding.

Since the inception of COC, the developmental trend aimed at reduction of the amounts of hormonal components in the combination to a level that would make the pill safer and still provide high contraceptive protection and cycle control.

The current COC preparations can be classified according to the amount of ethinylestradiol in the daily dose: (1) 20 µg of estrogen, the lowest dose used in the United States; (2) more than 20 µg to less than 50 µg, the dose most frequently recommended today; and (3) 50 µg, a dose that is rarely recommended. Preparations with an estrogen content greater than 50 µg/day have been removed from the market in the United States. A further distinction is made according to whether the dosage regimen is monophasic, biphasic, or triphasic.

The monophasic dosage regimens consist of contraceptive steroids given in a fixed estrogen-progestagen combination from the 1st through the 21st day of treatment.

With the biphasic preparations, the daily dose of ethinylestradiol is usually constant throughout the entire 21 days of use but the initially low progestagen dose increases at the middle of the cycle, from day 11 on. A special case of biphasic dosage is a preparation (Mircette) employing a constant dose of 20 µg of ethinylestradiol combined with 150 µg of desogestrel given from day 1 through 21. This is followed by 2 days of placebo and then 5 days of ethinylestradiol at 10 µg/day.

In the triphasic treatment regimens, the daily doses of one or both steroid components are modified three times during the treatment period. The individual dosage regimens for OC are listed in Table 17-6.

The development of the biphasic and triphasic treatment regimens was motivated by the desire to mimic the hormonal events of the normal menstrual cycle and to decrease the total load of contraceptive steroids per month. These dosage modifications have contributed to the variety of OC choices. Rigorous comparative studies have not been conducted to show whether the phasic regimens offers any clinically meaningful advantages over the monophasic schedules. However, similar performance at lower hormonal doses has theoretical appeal.

**The Sequential Method**

In this method, estrogens were given in a fixed dose throughout the 21-day treatment period; however, progestagens were given only during the last 5 to 9 days. The use of this regimen was discontinued in the United States in the 1970s because of concerns that repeated exposure of the endometrium to unopposed estrogens may induce atypical hyperplasia. The method is mentioned here because it is described in earlier literature.

**Continuous Progestagen-Only Oral Contraception**

The continuous progestagen-only method, known as the minipill, was developed to eliminate the estrogens entirely from OC preparations. The preparations currently available in the United States contain norethindrone, 350 µg/day; norgestrel, 75 µg/day; and levonorgestrel, 30 µg/day. Outside the United States, preparations with other progestagens are available: lynestrenol, 500 µg/day; ethynodiol diacetate, 500 µg/day; and desogestrel, 75 µg/day. The pregnancy rate for a typical user is higher than with the estrogen-progestagen OC, and the cycle control is less satisfactory.

The method is well suited for breast-feeding mothers. The amounts of progestagen that are excreted into the milk of breast-feeding mothers are negligibly low and do not affect the quantity and the composition of the milk. The combination of the ovulation-suppressive effect of prolactin and the contraceptive effects of the progestagen-only pill offers excellent protection from pregnancy. Another group of women who could benefit from this method are women around the age of 40 with naturally decreased fecundity. Good candidates for progestagen-only contraception are women who do not tolerate estrogens and who reject the use of an IUD.

Some preparations using the continuous progestagen-only method (minipill) have been associated with ectopic pregnancy. For this reason, when pregnancy occurs, all efforts must be made to rule out ectopic pregnancy. The physician must be attentive to complaints of pelvic pain by users of the progestin-only method; sometimes the diagnosis of ectopic pregnancy is delayed because symptoms of ectopic pregnancy such as irregular uterine bleeding, prolonged cycles, and amenorrhea resemble typical side effects.

The decreased efficacy of the minipill compared with combined OCs is probably due to its mechanism of action. The minipill does not consistently inhibit ovulation (about 40% of cycles are ovulatory); thus, other mechanisms become operative. Prevention of fertilization is largely due to changes in viscosity of the cervical mucus in addition to other changes that are inhospitable to impregnation of the ovum.

**Benefits of Oral Contraception**

Physicians, patients, and the general public have been made well aware of adverse effects of hormonal contraception. Data obtained over the last two decades brought evidence that the
<table>
<thead>
<tr>
<th>Brand Name</th>
<th>Type</th>
<th>Levonorgestrel Dose</th>
<th>Norethindrone Dose</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td></td>
<td>0.075 mg</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>0.125 mg</td>
<td></td>
</tr>
<tr>
<td>Yasmin</td>
<td>Combination</td>
<td>0.125 mg</td>
<td>3.0 mg</td>
</tr>
<tr>
<td>Bristol-Myers Squibb</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ovcon-35</td>
<td>Combination</td>
<td>0.125 mg</td>
<td>30 mg</td>
</tr>
<tr>
<td>Ovcon-50</td>
<td>Combination</td>
<td>0.500 mg</td>
<td>50 mg</td>
</tr>
<tr>
<td>Organon</td>
<td></td>
<td>0.125 mg</td>
<td>3.0 mg</td>
</tr>
<tr>
<td>Desogen</td>
<td>Combination</td>
<td>0.125 mg</td>
<td>35 mg</td>
</tr>
<tr>
<td>Micette</td>
<td>Biphasic, special combination</td>
<td>0.125 mg</td>
<td>30 mg</td>
</tr>
<tr>
<td>Cyclessa</td>
<td>Combination, triphasic</td>
<td>0.15 mg</td>
<td>20 mg</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>0.15 mg</td>
<td>10 mg</td>
</tr>
<tr>
<td>Ortho-McNeil Pharmaceutical</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Micronor</td>
<td>Progestagen-only</td>
<td>0.15 mg</td>
<td>35 mg</td>
</tr>
<tr>
<td>Micronor</td>
<td>Combination</td>
<td>0.15 mg</td>
<td>35 mg</td>
</tr>
<tr>
<td>Micronor</td>
<td>Combination</td>
<td>0.15 mg</td>
<td>35 mg</td>
</tr>
<tr>
<td>Micronor</td>
<td>Combination</td>
<td>0.15 mg</td>
<td>35 mg</td>
</tr>
<tr>
<td>Micronor</td>
<td>Combination</td>
<td>0.15 mg</td>
<td>35 mg</td>
</tr>
<tr>
<td>Micronor</td>
<td>Combination</td>
<td>0.15 mg</td>
<td>35 mg</td>
</tr>
<tr>
<td>Ortho-Novum</td>
<td>Combination, triphasic</td>
<td>0.5 mg</td>
<td>35 mg</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>0.75 mg</td>
<td>35 mg</td>
</tr>
<tr>
<td>Ortho-Novum</td>
<td>Combination, biphasic</td>
<td>1.0 mg</td>
<td>35 mg</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>1.0 mg</td>
<td>35 mg</td>
</tr>
<tr>
<td>Ortho-Novum</td>
<td>Combination, triphasic</td>
<td>0.5 mg</td>
<td>35 mg</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>0.75 mg</td>
<td>35 mg</td>
</tr>
<tr>
<td>Ortho-Novum</td>
<td>Combination, triphasic</td>
<td>1.0 mg</td>
<td>35 mg</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>1.0 mg</td>
<td>35 mg</td>
</tr>
<tr>
<td>Parke-Davis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estrostep</td>
<td>Combination, triphasic</td>
<td>1.0 mg</td>
<td>20 mg</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>1.0 mg</td>
<td>35 mg</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>1.0 mg</td>
<td>35 mg</td>
</tr>
<tr>
<td>Loestrin 1/20</td>
<td>Combination</td>
<td>1.0 mg</td>
<td>35 mg</td>
</tr>
<tr>
<td>Loestrin 1.5/30</td>
<td>Combination</td>
<td>1.5 mg</td>
<td>35 mg</td>
</tr>
<tr>
<td>Norlestrin 1/50</td>
<td>Combination</td>
<td>1.0 mg</td>
<td>50 mg</td>
</tr>
<tr>
<td>Norlestrin 2.5/50</td>
<td>Combination</td>
<td>2.5 mg</td>
<td>50 mg</td>
</tr>
<tr>
<td>Watson Laboratories</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brevicon</td>
<td>Combination</td>
<td>0.5 mg</td>
<td>35 mg</td>
</tr>
<tr>
<td>Norinyl 1 + 35</td>
<td>Combination</td>
<td>1.0 mg</td>
<td>35 mg</td>
</tr>
<tr>
<td>Norinyl 1 + 50</td>
<td>Combination</td>
<td>1.0 mg</td>
<td>35 mg</td>
</tr>
<tr>
<td>Nor-Q.D.</td>
<td>Combination</td>
<td>0.5 mg</td>
<td>35 mg</td>
</tr>
<tr>
<td>Tri-Norinyl</td>
<td>Combination, triphasic</td>
<td>0.5 mg</td>
<td>35 mg</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>1.0 mg</td>
<td>35 mg</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>1.0 mg</td>
<td>35 mg</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>0.5 mg</td>
<td>35 mg</td>
</tr>
<tr>
<td>Wyeth-Ayerst</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alesse</td>
<td>Combination</td>
<td>0.100 mg</td>
<td>20 mg</td>
</tr>
<tr>
<td>Lo/Ovral</td>
<td>Combination</td>
<td>0.300 mg</td>
<td>30 mg</td>
</tr>
<tr>
<td>Nordette</td>
<td>Combination</td>
<td>0.150 mg</td>
<td>30 mg</td>
</tr>
<tr>
<td>Ovral</td>
<td>Combination</td>
<td>0.500 mg</td>
<td>50 mg</td>
</tr>
<tr>
<td>Ovrette</td>
<td>Progestagen-only</td>
<td>0.075 mg</td>
<td>50 mg</td>
</tr>
<tr>
<td>Triphasil</td>
<td>Combination</td>
<td>0.050 mg</td>
<td>30 mg</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>0.050 mg</td>
<td>30 mg</td>
</tr>
</tbody>
</table>
The decreased risk is related to the length of use of OCs. A decrease is evident after only 1 year of use; women who have been practicing hormonal contraception for 4 years or more have risk reduced by more than 50%. It is important to note that hormonal contraception offers long-term protection and that the residual protective effects persist for 20 years or longer. It is also important that the protective effect has been associated with the use of all OCs for which data have been gathered.

Data on progestagen-only contraception and preparations with 20 µg of ethinylestradiol are not yet available, but it is likely that they would offer the same protection as the other preparations.

The mechanism of the protective effect on the endometrium most likely involves the direct antiestrogenic effects of the progestagen component of the OCs. Estrogens stimulate synthesis of both estrogen and progestagen receptors, and progestagens inhibit this synthesis. Part of the antiestrogenic effect of progestagens may also involve the stimulation of estradiol 17-hydrolases in the endometrial cell and accelerated conversion of estradiol to estrone, a less potent estrogen than estradiol. Consequently, the proliferation of the endometrial epithelium, both endometrial glands and stroma, proceeds at a reduced rate with less mitotic activity.

With an estimated 26,700 new cases per year, ovarian carcinoma occurs less frequently than endometrial carcinoma; however, it is more deadly. With 14,800 deaths per year, ovarian cancer is the fourth leading cause of death from cancer, after lung, breast, and colon and rectal carcinomas. The 5-year survival rate is less than 45% after the diagnosis has been established. The protective effect of hormonal contraception is evident with only 3 to 6 months of use. It becomes highly significant after 7 years of use (relative risk 0.2 to 0.4), and the residual protective effect extends at least 10 to 15 years after termination of use.

Hormonal contraception reduces the risk of functional ovarian cysts by 60%. This can be important because functional ovarian cysts are the fourth leading reason for hospitalization of women in the United States, with 160,000 admissions per year. We can speculate that the decreased relative risk of both ovarian carcinoma and functional ovarian cysts is probably due to inhibition of monthly proliferation of the graafian follicles.

Each year, more than 1 million women in the United States experience an episode of pelvic inflammatory disease (PID). Epidemiologic data suggest that women taking OCs have a reduced risk of being hospitalized for PID, but further work is needed to assess the effects of OCs on the incidence of PID.

With respect to ectopic pregnancy, studies from the early...
1990s showed that the incidence rate of ectopic pregnancy per 1000 woman-years is 3.0 for women who do not practice contraception at all. The incidence rate is reduced to 0.005 in women using combination OCs. In women using a Cu-T IUD or the levonorgestrel-containing implant Norplant and in women who have undergone tubal sterilization, the incidence rate is still low (0.2 to 0.3). 684 685

Uterine Leiomyomas

High doses of norethindrone only, or norethindrone given simultaneously with gonadotropin-releasing hormone (GnRH) agonists, can achieve reduction of the size of leiomyomas. Because leiomyoma is the most frequently encountered tumor of the female genital tract, there has been great interest in determining whether OCs protect against the occurrence of this tumor. The question has been addressed by several studies that provided opposing results, and the problem remains unresolved. At least, there does not appear to be an increased risk. 686 687

Endometriosis

OCs have been recommended for mild forms of endometriosis. However, well-designed studies proving a substantive effect are lacking. OCs may protect against the occurrence of endometriosis, 688 and some clinicians use OCs as a follow-up for the prevention of recurrence of endometriosis after a completed course of GnRH agonists.

Other Reproductive Noncontraceptive Benefits

Table 17-7 also shows the beneficial effect of OCs on conditions that are not always serious but negatively affect the quality of life. Such conditions include dysmenorrhea, mittelschmerz, menorrhagia and irregular menses, premenstrual syndrome, and iron deficiency anemia. The beneficial effects are achieved by suppression of ovulation and by influencing the endometrium.

Nonreproductive Benefits of Oral Contraception

Management of Hyperandrogenism

OCs have been used in the management of hyperandrogenic conditions such as acne, seborrhea, and hirsutism. Triphasic norgestimate-ethinylestradiol regimens and other OC combinations of ethinylestradiol with norethindrone acetate, levonorgestrel, or desogestrel have been shown in randomized, blinded, placebo-controlled trials to be effective in the treatment of acne. The efficacy of other OCs has also been supported by studies conducted under less rigorous protocols. 689 690 For example, a single-blind randomized study demonstrated alleviation of acne by treatment with a progestagen, chloromadinone acetate, combined with ethinylestradiol. 691

Several modes of action have been considered to explain the efficacy of OCs in hyperandrogenism. OCs can suppress production of ovarian androgens by inhibiting pituitary gonadotropins. An increase of circulating levels of SHBG is associated with a decrease of bioavailable testosterone, and inhibition of 5-reductase in the skin tissues can also contribute to this antiandrogenic effect.

Medical treatment of hirsutism with OCs is more difficult. In addition to a number of supportive case series reports, two well-controlled clinical trials using OCs with or without GnRH agonists have been reported. 692 693 The first one employed a randomized, double-blind, placebo-controlled study design. Neither the OC Norinyl 1/35 (1 mg of norethindrone plus 35 µg of ethinylestradiol) nor placebo had a beneficial effect on hirsutism. In the second study, which was investigator-blind but not placebo-controlled, the contraceptive Demulen (1 mg of ethynodiol diacetate plus 35 µg of ethinylestradiol) had no effect on hirsutism. However, OC complements GnRH agonist analogues in the management of hirsutism. In these two studies, only the combination of OC and a GnRH agonist had a clinically and statistically significant beneficial effect on hirsutism.

Bone Mineral Density

Well-designed studies demonstrate that the use of OCs increases bone mineral density so that users enter menopause with higher bone mass than nonusers, by 12% on the average. This beneficial effect depends on the duration of OC use; the greatest protection is afforded to women who use OCs for 10 years or more. The ultimate question, whether previous OC users suffer fewer bone fractures during menopause than nonusers, has been addressed by a large case-control study. The results have shown a 25% reduction in hip fractures in previous users of OCs. 694 695

Rheumatoid Arthritis

The relationship between OC use and rheumatoid arthritis has been of interest in countries where this disease affects larger segments of the population. A case-control study from Holland reported 60% protection in ever-users of OCs. 696 Other studies and meta-analyses of various clinical trials have not provided an unequivocal conclusion. 697

Colorectal Cancer

Several studies employing epidemiologic methodology have provided evidence that OCs afford about 50% protection from colorectal cancer and that this effect is directly proportional to the duration of OC use. The subject remains controversial because other clinical observations failed to confirm the protective effect. The mechanism by which OCs would exert a protective effect against colorectal cancer is not clear. 698 699

Adverse Events Associated with the Use of Oral Contraception

The frequency of adverse events associated with oral contraception has decreased continuously since hormonal contraception was first introduced for general clinical use. There are several reasons for this favorable development. The amount of estrogens in the pill has been gradually reduced since estrogens were identified as the culprit in the most severe adverse events, principally cardiovascular and cerebrovascular complications. This reduction was paralleled by decreases in the daily doses of the progestagenic component of OC. In addition, candidates for OCs are being selected more carefully as risk factors for potential complications have been identified, notably smoking among women older than 35. Researchers realized that more frequent follow-up of new OC users could detect early signs and symptoms of complications. For example, measurements of blood pressure before therapy and during the first 3 months of OC use identify individuals predisposed to hypertension. Finally, physicians defined appropriate contraceptive options for patients with medical problems.

Most Commonly Reported Adverse Events

The use of hormonal contraception is associated with a number of less serious adverse events that are nonetheless important because they constitute reasons for discontinuation.

Breakthrough Bleeding.

This event can be expected in 10% to 30% of women during the first 3 months of use of hormonal contraception; thereafter, it is much less frequent. Depending on its intensity and duration, it can be handled by reassuring the patient or by short-term administration of a supplementary estrogen such as micronized estradiol.
Amnenorrhea.

Amnenorrhea, that is, lack of bleeding during the active pill-free period, is always a cause for anxiety because of the possibility of pregnancy. Pregnancy should be ruled out by a sensitive urine or blood pregnancy test. If pregnancy is ruled out and the patient is comfortable with amenorrhea, she may resume the same OC. Alternatively, she may be switched to another OC, usually with more estrogen dominance. Sometimes supplementation by estrogens is recommended. Repeated episodes of amenorrhea can be bothersome and irritating, and it is sometimes best to recommend another method of contraception. A positive pregnancy test should alert the physician to the unlikely but critical possibility of an ectopic pregnancy.

Other Adverse Effects.

One study compared two groups of patients, randomly assigned to receive either a tricyclic combination of ethinylestradiol and norgestimate (Ortho Tri-Cyclen) or a placebo. Symptoms that are usually attributed to the use of OCs, such as common headaches, nausea, breast tension and tenderness, weight gain, and mood change, were assessed before and during treatment. The difference in the incidence of side effects between the two groups was not statistically significant. Only the higher incidence of breast tenderness and mood change in the OC group approached significance (P = .07).

Women who discontinue OC use conceive later than women practicing nonhormonal methods of contraception. "Postpill amenorrhea" develops in 1% of users.

Caution.

Despite the reassuring reports with respect to adverse events, the prescribing physician must be aware that certain conditions warrant caution or constitute a frank contraindication to hormonal contraception. These conditions are discussed in the sections "Contraindications" and "Contraception for Women with Health Problems."

Cardiovascular and Cerebrovascular Adverse Events

From the inception of their clinical use in the 1960s, OCs have been associated with cardiovascular and cerebrovascular complications, namely idiopathic venous thromboembolism (VTE), stroke, and myocardial infarction (MI). These adverse events merit reevaluation in the light of epidemiologic studies that were prompted by the introduction of low-dose OCs into clinical practice.

Venous Thromboembolism

The Role of Estrogens.

Since the first reports of VTE in OC users, clinicians suspected that the noxious agent is the estrogenic component of the combination pill. A dose-response relationship between the estrogen dose and VTE was demonstrated in an epidemiologic study of 234,218 women between 1980 and 1986, when both high-dose and low-dose estrogen pills were being prescribed. The highest incidence of VTE, 10 per 10,000 woman-years, occurred among women who used OCs with an ethinylestradiol content of more than 50 µg/day. With preparations containing the medium dose of ethinylestradiol, 50 µg/day, the incidence of VTE decreased to 7 per 10,000 woman-years; with pills having an ethinylestradiol content less than 50 µg/day, the rate of VTE decreased to 4.2 events per 10,000 woman-years. The difference between the

<table>
<thead>
<tr>
<th>TABLE 17-8 – Relative Risk of Idiopathic Venous Thromboembolism in Pregnancy and during the Use of Contraceptive Hormones</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Population</strong></td>
</tr>
<tr>
<td>Unexposed</td>
</tr>
<tr>
<td>Emergency contraception</td>
</tr>
<tr>
<td>Hormone replacement therapy</td>
</tr>
<tr>
<td>Progestagen-only contraception</td>
</tr>
<tr>
<td>Levonorgestrel-containing oral contraceptives (second generation)</td>
</tr>
<tr>
<td>Third generation oral contraceptives (gestodene, desogestrel)</td>
</tr>
<tr>
<td>Progestogens for menstrual disorders</td>
</tr>
<tr>
<td>Pregnant, postpartum</td>
</tr>
</tbody>
</table>


incidence of VTE at 50 µg/day ethinylestradiol and the incidence at less than 50 µg/day ethinylestradiol was statistically significant.

A subsequent study analyzed idiopathic VTE in a cohort of 74,086 women during the 5-year period from 1993 to 1997. In this epidemiologic study, the relative risk of VTE was defined for pregnancy, various forms of oral contraception including combination OC and the progestagen-only method, therapeutic use of progestagens, and hormonal replacement therapy. The results are summarized in Table 17-8, to which we have added an analysis of the relative risk of VTE associated with emergency contraception. The highest risk of VTE is associated with pregnancy13 times higher than in nonpregnant women. Emergency contraception is not associated with a substantive risk. There is a slight but nonsignificant association between progestagen-only contraception and VTE, the relative risk being 2.4 with a 95% confidence interval (CI) of 0.8 to 6.5. A similar association has been reported for women using hormonal replacement therapy, the relative risk being 2.3 (CI 0.4 to 15.0). With the therapeutic use of progestagens for menstrual disorders, the relative risk for development of VTE rose to 5.3 (CI 1.6 to 18.7). The daily doses in progestagen-only contraception are normally less than 0.5 mg, whereas the therapeutic doses range from 5.0 to 30.0 mg/day.

An ongoing controversy involves reports that the risk of VTE associated with combination OCs containing desogestrel and gestodene is more than twice that for OCs with norgestrel and norethindrone (3.4 versus 8.0). This finding merits a discussion of the role of progestagens in the genesis of VTE.

The Role of Progestagens.

In 1995, several independent clinical epidemiologic studies presented evidence that OCs containing desogestrel and gestodene are associated with double the risk of nonfatal VTE compared with earlier OCs containing norethindrone or levonorgestrel. The results were surprising because thrombotic phenomena have not been conventionally associated with the progestagenic component of OCs. The data were immediately questioned, and the relationship of the various progestagens to VTE became controversial. However, a study analyzing OC use for the 7-year period between 1993 and 1999 confirmed that the risk for development of

VTE of women using OCs containing desogestrel and gestodene is twice that of women using OCs with levonorgestrel. The controversy concerning the third-generation pills containing gestodene and desogestrel continued into 2001, when a meta-analysis established a 1.7-fold risk for third-generation versus second-generation pills.

Norgestimate has not been included in these analyses. An analysis based on postmarketing surveillance of adverse events associated with the use of COCs
Venous Thromboembolism and Coagulation Factors.

Studies of the effects of hormonal contraceptive agents on blood coagulation factors reveal that OCs increase the synthesis of globulins in the liver, including many clotting factors. Consequently, circulating concentrations of many clotting factors are affected but not to a clinically significant level. Most frequently affected are fibrinogen and factors dependent on vitamin K (prothrombin and factors VII, IX, and X) and factor XII. At the same time, a decrease in the levels of antithrombin III, an anticoagulant factor, was noted. Despite the fact that these changes had occurred in virtually all OC users tested, VTE remains a rare event.

In discussing these findings, several issues have to be taken into account. Even under physiologic conditions, concentrations of the clotting factors are excessive in the circulation of healthy women, in some cases reaching 200% of the "normal" values. For hemostasis, however, only a fraction of this activity is needed. The coagulation factors are proenzymes that are present in the circulation in their inactive form. Damage to the blood vessel must occur to activate the coagulation cascade. With respect to antithrombin III, the OC-induced decrease is about 10%, far short of the profound reduction needed to form a clot.

Changes in blood coagulation depend on the dose of estrogen. The decrease of the ethinylestradiol content to below 50 µg/day, common in current OC preparations, has considerably limited the changes in blood coagulation factors that were observed in OCs with a higher ethinylestradiol content.

Venous Thromboembolism and Leiden Factor V.

During the normal coagulation process, protein C and its cofactor S prevent hypercoagulation by inhibiting the activity of coagulation factors V and VII. The Leiden mutation, a genetic mutation of factor V consisting of an alteration of a single amino acid, makes factor V resistant to the action of protein C. The Leiden mutation of factor V occurs in 5% of the U.S. white population and is less frequent in black and Hispanic women. Its presence predisposes the carrier to VTE. (Table 17-9).

In women of reproductive age, the rate of VTE increases to 5.7 VTE events per 10,000 woman-years, and in OC users the increase amounts to 28.5 events per 10,000 woman-years. The identification of the Leiden factor V mutation is the first instance in which increased VTE events in OC users could be linked to a concrete defect in a coagulation cascade. However, screening for the Leiden mutation would be impractical because examination of 1 million potential OC users would detect only 50 women at risk. In addition, 62,000 women would have false-positive results.

Physicians generally consider a personal history of venous thrombosis an absolute contraindication to the use of OCs. Screening

| TABLE 17-9 – Oral Contraception and Factor V Leiden: Risk of Idiopathic Venous Thromboembolism |
| Population | Relative Risk | Incidence per 10,000 Woman-Years |
| Controls | 1 | 0.8 |
| Oral contraception only | 3.8 | 3.0 |
| Factor V Leiden mutation | 7.9 | 5.7 |
| Factor V Leiden mutationoral contraceptive users | 34.7 | 28.5 |


for factor V Leiden may be justified in women with a strong family history of venous thrombosis. The presence of superficial varicose veins that are not a consequence of previous venous thrombosis is not a contraindication to the use of oral contraception.

In conclusion, although significant strides have been made in accumulating knowledge about VTE and OCs, the phenomenon remains as enigmatic as before. Reducing the dose of ethinylestradiol in the combination pill to less than 50 µg/day, along with other preventive measures, has substantially lowered the risk of VTE although it has not been eliminated entirely. There are no substantive data supporting increased safety for products containing 20 µg versus 30 to 35 µg despite the logical appeal. In-depth molecular biologic research is needed to understand VTE and pave the road to its rational prevention.

Stroke and Oral Contraception.

Stroke has been recognized as one of the serious complications of OC use, although its incidence has been rare. In 1976, Vessey and Doll reported 41 to 45 strokes per 100,000 woman-years in OC users, a fourfold to fivefold increase over the rate of stroke in nonusers. Later studies had more reassuring outcomes. In 1996, a large epidemiologic study demonstrated that women using OCs with a low estrogen content (<50 µg of ethinylestradiol) are not at increased risk for stroke.

The published findings were based on investigations of a large California health organization, Kaiser Permanente, and included an analysis of 1.1 million women, 14 to 40 years of age, who were observed during the 5 years 1991 to 1994 for a total of 3.6 million woman-years. During this time period, 408 women suffered a proven stroke. This confirms the fact that stroke occurs rarely in young women, with an incidence rate of only 11.3 cases per 100,000 woman-years. With respect to OCs, 1.93 (0.874.29) strokes per 100,000 woman-years in OC users, a fourfold to fivefold increase over the rate of stroke in nonusers. Later studies had more reassuring outcomes. In conclusion, although significant strides have been made in accumulating knowledge about VTE and OCs, the phenomenon remains as enigmatic as before.

<p>| TABLE 17-10 -- Relative Risk of Stroke in Women Using Low-Estrogen Oral Contraception (&lt;50 µg Ethinylestradiol) |</p>
<table>
<thead>
<tr>
<th>Use of oral contraception</th>
<th>Ischemic</th>
<th>Hemorrhagic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never or past users only</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Current users</td>
<td>1.18 (0.542.59)</td>
<td>1.14 (0.602.16)</td>
</tr>
<tr>
<td>Current users plus smoking</td>
<td>0.74 (0.173.87)</td>
<td>3.64 (0.9513.87)</td>
</tr>
</tbody>
</table>

Health status and lifestyle:

| Hypertension | 7.78 (3.5177.31) | 4.64 (2.1410.06) |
| Diabetes | 7.15 (3.1716.13) | 2.5 (0.620.08) |
| >3 Drinks per week | 1.93 (0.874.29) | 2.02 (1.063.85) |

The association of stroke, OCs, and smoking is not as strong as that for MI, although an increased risk of stroke for smokers using OCs was demonstrated in significantly higher, 6.88 (95% CI 5.05 to 9.36). Nonsmokers showed a strong dose response, from 2.47 (95% CI 1.12 to 5.45) in smokers of 1 to 5 cigarettes per day to 74.6 (95% CI 33.0 to 169.0) in smokers of 40 cigarettes per day.

Between 1997 and 1999, three important studies addressed this issue. A multicenter study organized by the WHO found a 5-fold increased risk of MI in current users of OC in Europe but only a 2.6-fold increase in women who had their blood pressure checked before prescription of an OC. When a subset of patients from England was analyzed, the odds ratio was 2.10 (CI 0.63 to 7.07); in other countries the odds ratio was much higher, possibly indicating different practices in selecting patients for OC. Another international study pointed out that any risk of MI is confined to women with known cardiovascular risk factors.

The third study was a case-control study conducted in 1993 to 1995. A group of 448 women of childbearing age, all of whom had suffered an incident of MI, was matched with 1728 control subjects without a history of MI. It is important to note that during the time period of the analysis, OCs with an ethinylestradiol content of less than 50 µg, combined with norgestrel or levonorgestrel as well as with norgestimate, desogestrel, and gestodene, were already in clinical use. This study had sufficient power to examine the effects of OCs with different progestagens on the incidence of MI.

The study has shown that the risk of MI in OC users is not increased and that there is no difference between OCs with an ethinylestradiol content less than 50 µg combined with norgestrel or levonorgestrel or with norgestimate, desogestrel, and gestodene. Progestagen-only contraception was found to have no effect on the genesis of MI. The study confirmed the high risk of MI associated with hypertension, diabetes, hyperlipidemia, angina, diabetes mellitus, and smoking. Of these, the highest risk factor was smoking. On the other hand, modest alcohol intake (once a week) and physical exercise for 1 hour or more per week were found to be beneficial. These study conclusions were strengthened by the observation that of women younger than 45 years who suffered MI, 87% were not taking OCs but most of them had one or more known cardiovascular risk factors.

Oral Contraception and Smoking

Perhaps in no other condition have the adverse effects of smoking been so well defined as in oral contraception. Cardiovascular and cerebrovascular complications of OC use have been recognized since the early surveillance studies that also pointed out the negative effects of smoking. Whereas the probability of developing these complications has been minimized for nonsmokers by reducing the amount of estrogens, smoking OC users are still exposed to a considerable risk of MI and stroke.

A long-term follow-up of women by the Royal College of General Practitioners contributed considerably to our understanding of the relation between OC, smoking, and MI. In 1998, data for 10,073 women were evaluated to determine whether changes in smoking habits have an effect on the risk estimates for MI. When the information on smoking supplied by the women at their entrance into the study was taken into consideration, the relative risk of MI was 3.6 (95% CI 2.2 to 5.9). However, during the study, 53% of women who had ever smoked regularly stopped smoking and only 4.5% of nonsmokers started smoking. Therefore, the authors performed another analysis in which they took into consideration the smoking status of the women at the occurrence of MI. As expected, women who continued smoking throughout the study were at a much higher risk for MI than those who did not smoke at the occurrence of the event. The relative risk for smokers was 5.1 (95% CI 3.0 to 8.7).

The increased risk of smokers for MI has also been underscored in the Myocardial Infarction Causality case-control study conducted between 1993 and 1995 in the United Kingdom. Odds ratios for risk of MI in smokers versus nonsmokers showed a strong dose response, from 2.47 (95% CI 1.12 to 5.45) in smokers of 1 to 5 cigarettes per day to 74.6 (95% CI 33.0 to 169.0) in smokers of 40 cigarettes per day. The relative risk of MI for all users of combination OCs was at 1.40 (95% CI 0.78 to 2.52), but the relative risk for women who ever smoked was significantly higher, 6.88 (95% CI 5.05 to 9.36).

The association of stroke, OCs, and smoking is not as strong as that for MI, although an increased risk of stroke for smokers using OCs was demonstrated in...
The landmark epidemiologic study by Petitti and colleagues in 1996 showed that for young women smokers in the United States using OCs with less than 50 µg of ethinylestradiol, the risk of hemorrhagic stroke is three times higher than for nonsmokers. In this study, smoking did not increase the risk of ischemic stroke.

Two WHO collaborative studies in 1996 evaluated the effect of smoking in a case-control design. With respect to hemorrhagic stroke in women of all ages, the odds ratios for current OC users who were also current cigarette smokers were greater than 3. For female smokers younger than 35 years of age, the odds ratios were 2.6 in Europe and developing countries; for female smokers older than 35 years the odds ratios were 3.9 for Europe and 5.4 for developing countries. With respect to ischemic stroke, the WHO collaborative study showed that the OC-associated odds ratios were higher among current smokers in Europe (7.2 smokers versus 2.1 nonsmokers) as well as in developing countries (4.8 smokers versus 2.6 nonsmokers), suggesting a synergistic effect of both factors.

Currently, mechanisms by which smoking exerts its negative effects are being investigated. Attention is focused on nicotine’s interference with endothelial cells, principally with their ability to secrete vasodilating substances and to regenerate.

The data presented indicate that OCs can be safely prescribed to women of all ages as long as they do not smoke and have no other risk factor. OCs can be prescribed to smokers younger than 35 years, but after this age limit OCs must be avoided unless the women stop smoking. For patients of any age, smoking cessation may be the single most effective preventive action any health care provider can recommend. Women seeking contraceptive advice must be thoroughly questioned about their smoking history, and the dangers of smoking must be clearly delineated.

Breast Cancer

Breast cancer is a frightening malignancy. In the United States, 182,000 new cases are detected annually. It is the most common carcinoma in women, and with 46,000 deaths a year it is the second most lethal cancer in women. Because of fear of breast cancer, many women do not start OCs or may discontinue their use. Fortunately, a recent study has convincingly demonstrated that present or past use of oral contraceptives is not associated with an increased risk of breast cancer.

This clinical trial was a population-based case-control study conducted under an extremely rigorous protocol. The *cases*4575 women newly diagnosed with breast cancer during July 1994 through April 1996, were matched and compared to 4682 control women without breast carcinoma. The probands were 3564 years of age, a period of life during which breast cancer strikes most frequently. The study enrolled both Caucasian as well as African Americans and was conducted in five regions of the United States. Among current OC users the relative risk of breast cancer was 1.0 (95% confidence interval 0.8 to 1.3); for former users the risk was 0.9 (95% confidence interval 0.8 to 1.0). The study has also failed to detect any significant association between the risk of breast cancer and the amount of estrogen in the combination, duration of OC use, the initiation of use during adolescence, the race of the probands, or family history of breast cancer.

An earlier study from 1996 was a meta-analysis of 54 small studies conducted over a period of 25 years. However, the study had the disadvantages of a meta-analysis, for example, non-uniform study designs and protocols, different quality of study conduct and analyses, and follow-up. The study showed only a slightly increased risk of breast cancer among current OC users and a gradual decline of the risk during the postpill period, until the risk completely disappeared 10 years after stopping the pill. Why breast cancer in OC users is less advanced than in nonusers has not been explained.

Whether women with a family history of breast cancer are at a higher risk for the disease while using OCs was analyzed in more detail in another case-control study. Study participants included 394 sisters and daughters and 3002 granddaughters and nieces of 436 families of breast cancer probands. OC use was associated with an increased risk of breast cancer among daughters and sisters (relative risk 3.3, CI 1.6 to 6.7) but not among granddaughters and nieces. Other studies have been consistent with regard to family history as an OC risk factor. The study underscores the importance of obtaining a thorough family history not only before prescribing OCs but for women in general.

Cervical Carcinoma and Oral Contraception

In the United States, 12,800 cases of invasive cervical cancer are diagnosed annually and 4800 women die from this disease. Cervical cancer used to be the most frequently diagnosed malignancy in women in the United States. Over the last 30 to 40 years, its incidence has been declining, most likely as a result of early treatment of precursor lesions. Today, it has assumed the fifth place among the most frequently encountered cancers in women. Nevertheless, concerns about its association with the use of OCs remain.

Squamous Cell Cervical Carcinoma

Epidemiologic studies of the relation between oral contraception and squamous cell carcinoma of the uterine cervix originate from the late 1980s and early 1990s. No association between cervical carcinoma and OCs was found during the first 5 years of use, but the incidence of cervical carcinoma doubled at and after 10 years of use. Squamous cell carcinoma arises from the columnar epithelium of the endocervical canal and constitutes 10% to 15% of all cervical cancers. It has been linked to the use of OCs for several reasons. A doubling of its incidence along with a decrease in the frequency of squamous cell cervical cancer was noted in the 1970s and 1980s. This time period coincides with the rise of OC use. Notably, the malignancy affected women between the ages of 20 and 35 years from higher socioeconomic backgrounds, a group that was inclined to practice hormonal contraception.

Episodic studies of OCs and endocervical carcinoma showed that OC use was associated with a risk of adenocarcinoma of the cervix twice that in never-users and that the risk increased with the duration of use, being highest after 12 years of use.

As with its squamous cell counterpart, the incidence of endocervical cancer depends on the sexual habits of affected women, such as age at first sexual intercourse, number of sexual partners, and the use of barrier methods of contraception including spermicides. However, the most confounding factor is exposure to the human papillomavirus (HPV) epidemic that occurred at the time when the incidence of adenocarcinoma of the cervix began to rise. It is not clear whether OC use and the coinciding change in sexual behavior, including displacement of barrier methods, may have provided conditions suitable for the virus to flourish.

A study from the National Cancer Institute in 1999 has shown that both types of cervical cancers are associated with OC use. However, the positive association was weak for squamous cell carcinoma but strong for adenocarcinoma. After accounting for HPV infection, sexual history, and cytologic screening, the association between OCs and squamous cell in situ carcinoma, squamous cell invasive carcinoma, and invasive adenocarcinoma disappeared. A positive association remained between current use of OCs and adenocarcinoma in situ. These results are reassuring, but they underscore the necessity of properly collecting endocervical cells and
taking cervical samples for cytologic screening. Women at risk for STD should rely on an effective method of contraception to reduce the likelihood of unintended pregnancy. They should combine that use with a condom or diaphragm to reduce the risk of infection.

Liver Tumors

Contraceptive steroids have been associated with an increased incidence of benign liver adenomas. However, the association has been questionable at best. Because of the rarity of liver adenomas, their incidence is difficult to establish, even in the general population. Earlier estimates for OC users varied from one adenoma in 300,000 to one in 1 million women.

An increased risk of liver cancer among users of OC has not been proved. The incidence of this malignancy and the rate of death have remained relatively stable in the United States over the decades when generations of women have used OCs. Nevertheless, the provider should be alert to the possibility of liver cancer in this population.

Mortality Associated with Oral Contraceptive Use

The most comprehensive analysis of mortality and OCs has been performed by the Royal College of General Practitioners in Britain. A total of 46,000 women, half of them OC users, were observed for 25 years from 1968 to 1993. Over the entire period, 1599 deaths were reported. Oral contraception did not increase or decrease total mortality. The relative risk of death from all causes combined did not differ significantly between ever-users and never-users. The relative risk of death from ovarian and colorectal cancer was significantly lower in ever-users, and the risk of dying from circulatory disease was higher in the OC group. During the use of OC and up to 10 years after discontinuation, the risk of cervical carcinoma was higher in the OC group (see also discussion of oral contraception and neoplasia). It should be noted that the study was done during a time when most of the OCs contained 50 µg or more of ethinylestradiol. Nevertheless, the overall data are reassuring.

Contraindications

With the current types of low-dose OCs, there are few contraindications to their use. Those that exist include presence of or a history of thromboembolic disease, thrombophlebitis, coronary occlusion, atherosclerosis and stroke, uncontrolled hypertension, most cases of systemic lupus erythematosus, and diabetic retinopathy or nephropathy, although uncomplicated diabetes mellitus is not an absolute contraindication to the use of OCs. For patients with known or suspected breast carcinoma, the use of OCs is contraindicated. Therefore, the prescribing physician must pay close attention to a personal and family history of breast carcinoma and examine potential OC users for breast nodules and breast carcinoma.

Any type of amenorrhea or abnormal uterine bleeding, or both, should be understood before prescribing OCs. Proper examinations and tests should rule out pregnancy as well as premalignant and malignant changes of the uterus or vagina. Amenorrhea may be the first sign of prolactinomas, and their presence must be ruled out if the patient’s amenorrhea cannot be plausibly explained. Smokers older than 35 years should not use OCs. Women with migrainoid headaches, severe depression, diabetes mellitus, and a history of gallbladder disease should be carefully observed while taking OCs. The presence of leiomyomas is not a contraindication, but it warrants monitoring.

With respect to the liver, hormonal contraception is contraindicated for women with active liver disease (i.e., currently abnormal liver function tests, infiltration, severe cirrhosis). Patients with acute viral hepatitis should not use OCs; however, after liver function tests have been normal for at least 3 months, combined OCs can be started or resumed. Further, the use of OCs is contraindicated in women with a history of cholestatic jaundice during prior pregnancy or OC use, the presence of or a history of a liver adenoma or malignant liver tumors, and the presence of hepatic porphyrias. In compensated liver cirrhosis and in less serious liver diseases, low-dose OCs can be used provided the patient is monitored by regular examinations including liver function tests. The section “Contraception for Women with Health Problems” includes additional information on indications and contraindications for the use of contraceptive methods.

Drug Interactions

Certain drugs can accelerate the biotransformation of contraceptive hormones, particularly of estrogens to metabolites that are less active. This was first reported for rifampin, which is used in the treatment of tuberculosis. Other drugs such as phenobarbtlone, promethazine, chlorpromazine, and phenytoinmost of them used in psychotherapy were also implicated. The basis of this drug interaction is probably induction of liver microsomal enzymes. It is not clear whether the accelerated metabolism of OCs by these compounds would be sufficient to cause reduced ovarian suppression or uterine breakthrough bleeding, or both. As with all effective methods of contraception, the product label provides up-to-date guidance as the field evolves.
LONG-ACTING PREPARATIONS

Hormonal Subdermal Implants

Contraception by means of hormonal implants was developed for women who seek a reliable and reversible long-term method of fertility control that requires less than daily compliance on their part. Hormonal implants are suitable for women who do not wish to take a pill daily or to return regularly for depot injections and who desire or need protection from pregnancy for 1 to 5 years.

Types of Implants and Hormones Used

Table 17-12 summarizes hormonal implants that are currently available for general clinical use as well as preparations that are still in research. There are essentially three types of hormonal implants.

The first type consists of rod-like silicone capsules that are filled with the active steroid in microcrystalline form. The capsule determines the rate of drug release from the system. Norplant is an example of this type of implant.

Levonorgestrel was chosen as the progestagen for Norplant, the first contraceptive hormonal implant introduced into general clinical practice. Levonorgestrel is a highly active progestagen, which means that the necessary amount can be loaded into capsules of relatively small volume. The original version required six capsules to be implanted to achieve a contraceptive concentration of the hormone in the circulation. The number was later reduced to two implants. Norplant affords protection from pregnancy for 5 years. The number of implants in Norplant has been a source of numerous problems (see later); therefore, simplified methods have been pursued.

The second form of implants consists of rods with a matrix that carries the active hormone. The matrix is covered with a thin membrane that controls the release of the drug from the matrix into the surrounding tissues and ultimately into the circulation. The system is encased in silicone tubing. One version consists of two rods in which levonorgestrel is embedded.

Table 17-12 -- Hormonal Contraceptive Implants

<table>
<thead>
<tr>
<th>Commercial Name</th>
<th>No. of Implants and Size (mm)</th>
<th>Hormone, Total Amount</th>
<th>Release Characteristics</th>
<th>Years of Protection</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Silastic Implants Currently Available</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norplant</td>
<td>6 capsules, 34 × 2.4 mm</td>
<td>Levonorgestrel, 36 mg each capsule (216 mg)</td>
<td>First 2 mo: 80 µg/d; mo 26: 34 µg/d; gradual decrease to 25 µg/d at 5 yr of use</td>
<td>5</td>
</tr>
<tr>
<td>Norplant 2 or Jadelle</td>
<td>2 rods, 43 × 2.5 mm</td>
<td>Levonorgestrel, 75 mg each capsule (150 mg)</td>
<td>Similar to Norplant</td>
<td>5</td>
</tr>
<tr>
<td>Implanon</td>
<td>1 rod, 40 × 2.5 mm</td>
<td>Etonogestrel, 60 mg</td>
<td>Years 12; 60 µg/d; year 3: 30 µg/d</td>
<td>3</td>
</tr>
<tr>
<td>Uniplant</td>
<td>1 rod, 40 × 2.5 mm</td>
<td>Nomegestrol, 36 mg</td>
<td>100 µg/d</td>
<td>1</td>
</tr>
<tr>
<td>Elcometrin</td>
<td>1 capsule</td>
<td>Nestorone, 50 mg</td>
<td>ND</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>1 rod</td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td><strong>Biodegradable Implants in Research</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Capronor (capsule)</td>
<td>1 caprolactone rod</td>
<td></td>
<td>Blood levels of levonorgestrel 0.203 mg/mL</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>40 × 2.4 mm</td>
<td>Levonorgestrel 26 mg</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>25 × 2.4 mm</td>
<td>Levonorgestrel, 16 mg</td>
<td>Blood levels of levonorgestrel 0.304 mg/mL</td>
<td></td>
</tr>
<tr>
<td>Annuelle (pellet)</td>
<td>25 pellets each the size of a grain of rice</td>
<td>Norethindrone</td>
<td>In research</td>
<td>3</td>
</tr>
</tbody>
</table>

ND, no data.

*Plans to make the system clinically available have been deferred by the company holding the progestagen patent. Registered in Brazil for treatment of endometriosis.

in a silastic matrix covered by a membrane and encased in silastic tubing. Implanon is a single-rod system that employs etonogestrel (3-ketodesogestrel), the biologically active metabolite of desogestrel. It is designed to provide protection for 3 years. Uniplant is a single-rod system that contains nomegestrol acetate, a hybrid progestagen of the 19-norpregnane series (see “Classification of Contraceptive Progestagens”). It is effective for 1 year; however, the corporation holding the patent for the compound has deferred introduction of this system into clinical practice. Clinical research studies are also being conducted with Elcometrin, a system that releases nestorone and is effective for 2 years with a 6-month safety margin. In research. In the third system, the matrix is made of a biodegradable material that is firm enough to be implanted. The drug is released from the matrix, which is slowly degraded until it is completely used. Two prototypes of these systems have been in clinical research: (1) a caprolactone rod-shaped matrix containing levonorgestrel that provides protection for 1 year and (2) norethindrone or norgestimate embedded in pellets that are injected subcutaneously.

Mechanism of Action

Levonorgestrel exerts multiple contraceptive actions. When in sufficient concentrations, levonorgestrel suppresses ovulation; up to 90% of cycles are anovulatory during the first 2 years after insertion. The number of ovulations increases with time so that during the fifth year of use 50% of women ovulate, particularly those with a larger body mass. When ovulation occurs, secondary contraceptive mechanisms remain operative. The ovulatory cycles are frequently associated with an insufficient luteal phase, but the main contraceptive effect involves changes in the viscosity and consistency of the cervical mucus that make it difficult for the sperm to penetrate. Levonorgestrel also has a profound effect on the uterine mucosa that is out of phase, and researchers speculate that it becomes hostile to fertilization. Segal and colleagues have proved that fertilization with subsequent interruption of early pregnancy (menstrual abortion) does not occur during the use of Norplant.
Efficacy and Properties of Individual Steroids

The contraceptive efficacy of Norplant is high. Extensive data show pregnancy rates of less than 0.5 per 100 woman-years even after 5 years of use. Bleeding irregularities may be a problem, particularly during the first year of use. The contraceptive efficacy of Implanon, the system using etonogestrel, is also impressively high; no pregnancies were reported during more than 53,000 cycles of use. This high efficacy is probably related to ovulation suppression throughout the projected time of effective use; only during the last 6 months of use has an occasional ovulation been observed. The system is designed for 3 years of protection.

So far, data on the nestorone implant have shown high contraceptive efficacy, with only one pregnancy during 4000 cycles of use. Nestorone is a unique steroid that is not orally active because of rapid disintegration in the alimentary tract. Therefore, the implant is especially well suited for breast-feeding mothers.

Return to Fertility

After removal of the implant, ovulatory cycles occur promptly with rapid return of fertility. This is the main difference between the implant and MPA depot injections (Depo-Provera), in which residual deposits release the compound for a prolonged period after treatment has been discontinued, thus delaying pregnancy.

Table 17-13 -- Advantages and Disadvantages of Hormonal Implants

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>No estrogen</td>
<td>Minor surgery needed for initiation and termination of method</td>
</tr>
<tr>
<td>Long-term use not necessitating repeated effort on the part of the patient</td>
<td>Dependence on trained health care personnel to insert and remove the implants</td>
</tr>
<tr>
<td>Efficacy independent of compliance</td>
<td>High prevalence of endometrial bleeding irregularities</td>
</tr>
<tr>
<td>Exposure to progestagen lower than with the pill or injectables</td>
<td>High costs if discontinued early</td>
</tr>
<tr>
<td>In spite of menstrual irregularities, total blood loss is decreased</td>
<td></td>
</tr>
</tbody>
</table>

Adverse Events and Continuation Rates

During the first year after implantation, the major complaint has been irregular uterine bleeding, which caused some patients to discontinue the method. During later months of use, amenorrhea becomes more prominent. Despite these complications, the continuation rate after the first year has been close to 80% and after 5 years it has been 35% to 70%.

Because in some cases Norplant capsules were difficult to insert and even more difficult to remove, formation of scar tissue could not be avoided. The ensuing medical-legal problems and diminished demand led to voluntary withdrawal of the preparation from some markets, notably in Great Britain.

In conclusion, the advantages and disadvantages of hormonal implants are summarized in Table 17-13.
Injectable Progestagen-Only Preparations

The progestagens most widely used in long-acting injectable preparations are MPA, under the name Depo-Provera, and norethindrone enanthate, which is not available in the United States.

**Depot Medroxyprogesterone Acetate**

Depot MPA is an aqueous suspension on microcrystals of the hormone. A dose of 150 mg is injected every 3 months intramuscularly, preferably deep into the gluteal or deltoidal muscle. There is at least a 2-week safety margin after the prescribed 90 days when a new injection should be administered. A 300-mg dose is given every 6 months. The efficacy of depot MPA is high; the pregnancy rate is 0.5 per 100 woman-years for 12 months of use. This result presumes that women are able to obtain their injection at appropriate intervals.

**Norethindrone Enanthate**

Norethindrone enanthate is given intramuscularly in a dose of 200 mg. The first two injections should be given at 60-day intervals (±5 days), followed by injections every 12 weeks or 84 ± 7 days. The contraceptive efficacy of norethindrone enanthate is high, usually less than 1 pregnancy per 100 woman-years.

Long-acting injectable preparations are usually administered within the first 5 days of the menstrual cycle; when given later, a backup method usually a barrier contraceptives should be used for the first 14 days after dosing.

The main disadvantages of progestagen-only contraception are bleeding irregularities including metrorrhagia and amenorrhea. Another disadvantage is the delayed return to fertility; women do not become pregnant until about 6 months after they have discontinued treatment. Appropriate pretherapy counseling reduces requests for discontinuation. Depot MPA has been associated with decreased circulating levels of estradiol and a small and reversible decrease of bone mineral density. This finding is still debated, particularly with regard to use of depot MPA as a contraceptive method for teenage women.
Monthly Injectable Combination Contraception

To circumvent the disadvantages of progestagen-only contraception, monthly combination injectable contraception has been developed. Injection of 25 mg of MPA combined with 5 mg of estradiol cypionate (a long-acting derivative of estradiol) is performed within the first 5 days of the menstrual cycle and repeated every 28 ± 5 days. This method presents several improvements over the progestagen-only injectables for women who may use estrogen. Improvements include regular cyclic bleeding and, after discontinuation of this method, prompt resumption of fertility as with combination OC. Unlike use of OC, monthly injectable combination contraception does not require daily actions on the part of the users; however, the user must return to the health care practitioner every month to receive the injection. This presents a certain disadvantage compared with an injection of depot MPA or norethindrone enanthate every 3 months. However, the combination injections provide a high degree of protection; the reported pregnancy rate is 0.1 per 100 woman-years for women who return for injection at appropriate intervals.
Contraceptive Vaginal Rings

Contraceptive vaginal rings use another route of administration of contraceptive steroids. Vaginal mucosa has been found to absorb steroid hormones in amounts that are sufficient for contraception. In the United States, Mishell’s group has devoted considerable effort to the development of hormone-releasing vaginal rings since the late 1960s.

The current vaginal contraceptive rings are of the core design, in which the steroid is formulated into silicone rods that are placed within polysiloxane tubing. The release rate is directly proportional to the surface of the silicone core and inversely proportional to the thickness of the polysiloxane tubing.

The vaginal ring system was studied for progestagen-only contraception using two strategies: intermittent application, with monthly removal and reinsertion of the device, or continuous use for 3 months, after which the ring is exchanged. Nearly all available progestagens have been tested in the vaginal rings.

Other clinical studies were conducted with vaginal rings containing an estrogen-progestagen combination. In 2001, the Food and Drug Administration (FDA) approved the use of a vaginal ring releasing etonogestrel and ethinylestradiol over 21 days (NuvaRing, Organon, West Orange, New Jersey).
Transdermal Contraceptive Patch

Skin has been recognized as an excellent tissue for delivery of drugs. Hormonal patches for menopausal women were developed many years ago and have been well accepted.

The development of the contraceptive hormonal patch was motivated, in part, by the need to increase the compliance of patients who do not wish to take the pill daily. At present, only one contraceptive patch is available: Ortho EVRA. The patch is about 4 cm in diameter (20 cm$^2$), and it is designed to deliver 150 µg of norelgestromin and 20 µg of ethinylestradiol daily for 7 days. Effectiveness and compliance have been comparable to those of a triphasic OC regimen in clinical studies. Like monophasic OCs, the patch may one day be explored as a continuous versus cyclic method.
EMERGENCY METHODS OF FERTILITY CONTROL

Hormonal Method

Despite the wide availability of birth control methods, many women become pregnant unintentionally. In the United States, this number reaches an annual total of close to 3.5 million pregnancies; of these, over 1 million end in abortion. The following situations may lead to unintended pregnancies:

1. Unprotected intercourse.
2. Intercourse during failure or inadequate use of a barrier contraceptive method, such as a slipped or broken condom; incorrectly inserted, dislodged, or expelled diaphragm, cervical cap, or IUD; incorrectly placed female condom.
3. Intercourse after having missed progestagen-only contraceptive pills at any time of the cycle or after having missed combined contraceptive pills, in particular at the beginning or at the end of the pack, so that the pill-free interval is more than 7 days.
4. Unprotected intercourse or improper contraception shortly after a vasectomy when viable sperm are still in the vas deferens distal to the ligation (see "Vasectomy").

The need for emergency contraception, popularly known as the "morning-after pill," had already been perceived in the development of hormonal contraceptive methods. Morris and van Wagenen[145] coined the term "interception" and conducted early studies with diethylstilbestrol. Although diethylstilbestrol successfully prevented unintended pregnancy, it was quickly abandoned because of unacceptable adverse effects such as nausea, vomiting, and metrorrhagia.

More successful was the Yuzpe method, named after the Canadian gynecologist who invented it. The method consists of taking two tablets, each containing 0.5 mg of norgestrel combined with 50 µg of ethinylestradiol. This dose of 1 mg norgestrel and 100 µg ethinylestradiol must be repeated after 12 hours. Under ideal circumstances, the first dose should be taken within 72 hours of unprotected intercourse.[146]

In order to eliminate the estrogen-related adverse events of the combination pills, emergency contraception with a progestagen only has been developed. The treatment regimen consists of taking two doses of 0.75 mg of levonorgestrel 12 hours apart. As with the Yuzpe method, the dosing must start within 72 hours after the unprotected intercourse. The protective effect of the regimen gradually decreases when the pills are started later than 72 hours after unprotected intercourse. [147]

Successful clinical studies have been conducted with mifepristone in twice-daily doses of 5.0 and 25 mg. Compared with the methods described previously, mifepristone (see "Contragestion")

<table>
<thead>
<tr>
<th>Side effects (%)</th>
<th>/mg ethinylestradiol plus</th>
<th>/mg dl-norgestrel × 2</th>
<th>/mg levonorgestrel × 2</th>
<th>/mg mifepristone</th>
<th>/mg mifepristone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nausea</td>
<td>50</td>
<td>20</td>
<td>20</td>
<td>17</td>
<td>Delay of menses</td>
</tr>
<tr>
<td>Vomiting</td>
<td></td>
<td></td>
<td>6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

at the lower dose is associated with somewhat decreased side effects, principally vomiting, and the higher dose is associated with postponement of menses (Table 17-14)[148]

The mechanism of action has not been completely elucidated. The hormonal intervention may affect the timing of ovulation mostly postponement and produce an out-of-phase endometrium. The administered hormones may interfere with fertilization by affecting the ovum or the transport of sperm, or both.

The emergency contraception just described provides about 98% protection from an exposure to pregnancy risk when taken as prescribed. One of the major disadvantages of emergency contraception is not the method itself but its availability. Two states, Washington and California, have initiated a pilot program in which selected pharmacies following a strict protocol can distribute the morning-after pill directly to women who ask for it. Similar programs are being prepared in Oregon and Alaska.

When resorting to emergency contraception, patients are advised to take certain precautions that are specific for this kind of contraception. Along with combined estrogen-progestagen emergency pills, the patients may be given an antiemetic drug to minimize the adverse gastrointestinal events. After taking emergency hormones, ovulation and the following menstruation may be delayed. Patients should be advised to schedule an examination if menstrual bleeding does not occur within 3 weeks after the treatment. Patients should also be instructed to start to use a regular method of contraception. The use of emergency contraception as a primary method must be discouraged because of its side effects, including cycle and ovulation irregularities, and its relative unreliability with cumulative failure rates if used as the routine method.
Intrauterine Device as an Emergency Contraceptive

An IUD inserted 5 to 7 days after unprotected intercourse can provide an effective means of emergency fertility control. The method is contraindicated if an STD is present or suspected. Consequently, it is not suitable after rape, after coerced intercourse, or for women with multiple partners. In addition, the presence or a history of PID and suspected ectopic pregnancy are important contraindications to the method. The mechanism of IUD emergency contraception has not been defined. Also, unlike hormonal methods that are FDA approved, the IUD approach is not currently an approved method for emergency contraception.
Contragestion

The attempts to terminate early pregnancy by pharmacologic means have culminated in the development of a hormonal method for which the name contragestion has been suggested. The key hormone is mifepristone, a synthetic derivative of norethindrone, characterized by a complex structure (dimethylaminophenyl group) attached to C-11 (Fig. 17-19). It is recognized that manipulation of the C-11 position of the steroid molecule leads to compounds with glucocorticoid activity as well as to compounds with an increased capacity to bind to progesterone receptors. Desogestrel, a C-11 substituted steroid, is a potent progestagen. The modification on C-11 of mifepristone gives the molecule a unique ability to compete for specific binding sites on progesterone receptors. In fact, mifepristone has twice as much affinity for progesterone receptors as progesterone itself but without exercising any progestagenic activity. Thus, mifepristone is a true progestrogen antagonist. Mifepristone also binds to glucocorticoid receptors. However, in order to exercise any antiglucocorticoid activity, the amounts must be larger than those needed for its antiprogestagenic action.

It has been assumed that brief interruption of progesterone action leads to irreversible damage to the decidualized endometrium and interferes with nidation and the development of the early conceptus. Progesterone has a quieting effect on myometrial contractility. By antagonizing progesterone, mifepristone releases the myometrium from the suppressive action of progesterone, and myometrial activity intensifies. However, complete expulsion of the conceptus follows in only 40% of cases. Prostaglandins are potent stimulators of myometrial contractions, but when administered by themselves, they were able to terminate pregnancy in only 20% of the cases. With the availability of orally active prostaglandins, it was logical to combine the effect of an antiprogestagen with a prostaglandin. With administration before the 49th day of pregnancy, the success rate of complete abortion has been 92% to 98%.

Most experience with contragestion has been in England, France, and China. The treatment regimens consist of a single dose of mifepristone, 200 to 600 mg orally. Thirty-six to 48 hours thereafter, the women return to the clinic to receive an oral dose of 400 to 800 µg of misoprostol, a synthetic analogue of the natural prostaglandin E₃.

Prostaglandin vaginal suppositories can also be used. Prostaglandins of the prostaglandin E₂ series exert a dilatory effect on the uterine cervix, and it is possible that placing the prostaglandin suppository close to the cervix facilitates this effect. A large, well-designed study has proved that vaginal misoprostol is highly effective in achieving complete abortion; however, the study did not compare vaginal and oral dosing directly.210

Because approximately 50% of women expel the conceptus within 4 hours and an additional 11% do so during the fifth hour after the administration of misoprostol, women should be under observation at the clinic for up to 5 hours. If they do not expel the conceptus and there are no medical reasons for further observation or intervention, the woman is allowed to complete the abortion at home. She is scheduled for a follow-up examination within the next 14 days and instructed to notify the clinic if she experiences excessive uterine bleeding or other medical complication or if no bleeding occurs. Within 24 hours after receiving misoprostol, 75% of women expel the conceptus completely.

In the United States, a large study was conducted under the auspices of the Population Council.34 A total of 2015 women, 49 to 63 days pregnant, were enrolled. Of these, 92% achieved complete termination of pregnancy when the hormones were given on or before the 49th day of pregnancy. With more advanced gestation, the success rate decreased substantially. Complete abortion occurred in 83% of women pregnant 50 to 56 days and in 77% of women pregnant 57 to 63 days. Overall, for 85% of the 2015 patients, the treatment resulted in complete abortion. This percentage would probably be close to the success rate of the method in “real-life” situations, when the selection of suitable patients is less strict than under tightly controlled research conditions.

After dosing, the patients bleed for a median of 13 to 15 days. Although uterine bleeding is expected after pregnancy interruption, this bleeding was the main source of complaints and complications. Some of the complications required surgical intervention, intravenous administration of fluids, and, infrequently, blood transfusions. Other reported adverse events included uterine cramping, abdominal pain, nausea and vomiting, diarrhea, dizziness, and headaches. In some patients, endometritis was reported.

Mifepristone has been approved for general clinical use in the United States and is available under the name Mifeprex. Women seeking abortion should be counseled about the details of all options. Surgical intervention remains most common in the United States.
INTRAUTERINE DEVICES

Currently, there are three types of IUDs: nonmedicated, copper, and hormone-releasing intrauterine systems. Only the latter two are available in the United States.

The first modern IUD was designed by E. Gräfenberg, a German physician, in the 1920s. It was a pliable ring of coiled silver wire, 18 mm in diameter, inserted into the uterus after cervical dilation. This basic design has been modified over the years in various parts of the world, and stainless steel has replaced silver. With the advent of plastic materials with memory, new forms and shapes of IUDs were invented in order to increase efficacy, make the insertion easier, and facilitate the IUD's removal when needed.

Today, the WHO does not recommend the use of inert IUDs because copper IUDs and the hormone-releasing intrauterine systems are more effective and safer.

In the United States, IUDs have declined in popularity because of problems with the Dalkon Shield. This IUD was designed to cover a large area of the uterine cavity because it was speculated that the contraceptive efficacy of IUDs was directly related to the area of contact with the endometrium. However, the use of this device became associated with an increased risk of PID, and the Dalkon Shield was withdrawn from the market in 1974. It is believed that the reason for the increased incidence of PID was the multiframe tailstring, which facilitated proliferation of bacteria between the meshes of the string and ascension of infection into the uterine cavity. Current IUDs are equipped with a monofilament tailstring, and no substantive increased risk of PID has been associated with this appendage.

The first copper-bearing IUD was devised by the Chilean scientist Jaime Zipper, who studied the antifertility properties of copper. The first marketed copper IUD was the copper 7 IUD, based on Zipper's studies. It was distributed by the G.D. Searle company until 1986, when it was withdrawn from the market because of putative infertility of women who discontinued its use.

The modern copper IUDs are represented by Paragard T380. The IUD has a T-shaped polyethylene body with a copper wire wound around the vertical arm of the T; each of the transverse arms carries a copper sleeve. The exposed areas of copper are 380 mm², hence the name. This IUD is designed for 10-year protection against pregnancy. However, it is likely that the duration of efficacy is years longer. The pregnancy rate under controlled conditions is 0.7% and 0.3% for the first and the second year, respectively.

In Europe, a frameless copper IUD is being developed. The protective duration is 5 years. This IUD consists of a string with six copper beads covering a surface area of 330 mm². The string is anchored in the fundal myometrium. This kind of IUD is recommended for women who experience difficulties, principally uterine cramping, with other IUDs. However, users of this frameless IUD must be alert to the possibility of "silent" expulsion.

The first hormone-releasing intrauterine system was the Progestasert (1971). It is a T-shaped device with a drug reservoir in the vertical arm of the T containing progesterone in silicone oil. A vinyl membrane controls the rate of releaseapproximately 65 µg of progesterone per day. The pregnancy rate for Progestasert is about 2 per 100 woman-years. In addition, it must be replaced annually.

Another hormone-releasing IUD uses levonorgestrel (LNG). The LNG-20 or Mirena is also a T-shaped device with a sleeve on the vertical arm, which is a reservoir for a total of 52 mg of levonorgestrel. The hormone is released into the uterine cavity at a daily rate of 15 µg (calculated from in vitro release). The reported pregnancy rate for perfect use is 0.2% for the first year. It is intended for 5-year use.

Management Issues with Intrauterine Devices

Pregnancy and Continuation Rates

The modern copper and hormone-releasing IUDs are among the most effective means of reversible contraception. With long-term use the yearly pregnancy rate is less than 0.5%, and the cumulative pregnancy rate of 2.2% over a 12-year period rivals the 10-year pregnancy rate of tubal ligation (1.9%). The continuation rate of IUDs is the highest among all reversible methods of fertility control80% after the first year of use. The 5-year continuation rate is 40% among women using the copper IUD and 33% among women fitted with the levonorgestrel-releasing device. Medical therapy for Wilson's disease has been increasingly successful with the result that many affected women enter the reproductive age and need contraceptive counseling. Concerns have been expressed that copper ions released from the copper IUD may reach the general circulation and precipitate symptoms of the disease. In healthy women wearing copper IUDs,

![Figure 17-20 Types of intrauterine devices (IUDs). The copper T380 and the levonorgestrel-releasing IUD are clinically used in the United States. The frameless copper IUD is available in Europe.](image)

circulating concentrations of copper and ceruloplasmin have not been elevated even after 12 to 24 months of use. The amount of copper released from the IUD has been calculated as 14 to 29 µg/day, which constitutes only about 1% to 2% of the dietary intake of copper. In Wilson's disease, serum ceruloplasmin levels are decreased and free serum copper concentrations are increased, although total copper levels can be within normal limits or low. It seems unlikely that wearing a copper IUD could affect circulating copper levels in Wilson's disease. This condition is rare, 1 in 200,000; thus, it has been difficult to generate data on contraceptive practices of the affected women. However, because Wilson's disease is a serious condition and numerous contraceptive options exist, the lack of substantive safety data relegates copper-bearing IUDs to a bottom-tier choice for women with this condition.

Uterine Perforation

The rate of uterine perforations is low, 0.6 per 1000 insertions. With the progestagen-containing IUD, it is 1.1 per 1000 insertions.

Expulsion

The spontaneous expulsion rate for modern IUDs, including the copper T380 and levonorgestrel-releasing systems, is 5%.
to 6% during the first year and 1% to 2% per year during the subsequent years of use. Primiparas, secundiparas, and women 15 to 24 years of age have a higher rate of expulsions than women who are older or have had more than two pregnancies, or both. 159

**Changes in Uterine Bleeding**

The copper T380 is removed for uterine bleeding and cramping in 12% of users during the first year and 2% to 4% during subsequent years. The levonorgestrel-releasing system usually decreases uterine bleeding, and this phenomenon can be utilized therapeutically in patients with menorrhagia or adenomyosis and can prevent proliferative endometrial changes in patients undergoing long-term tamoxifen treatment. 156 On the other hand, patients using the levonorgestrel-releasing system complain of a higher incidence of amenorrhea, and removal of the device for this side effect is the main reason for the lower continuation rate compared with the copper T380 IUD. In both cases, proper counseling enhances continuation.

**Pelvic Inflammatory Disease**

The risk of IUD users for PID is inversely related to the time since insertion. A large WHO study concluded that the risk of PID is increased during the first 20 days after IUD insertion; beyond this time period, PID is an infrequent event. 166 167 This finding may be related to bacteriologic evidence showing bacterial contamination of the uterus at insertion of the IUD. Women scheduled for hysterectomy had an IUD inserted 24 hours before the surgery. Upon culture of the uterine contents, bacterial contamination was frequently found. It is even more significant that 30 days after IUD insertion, the uterine contents were sterile. 159

The development of PID also depends on the lifestyle of the women receiving the IUD. Women who are at minimal or no risk of exposure to sexually transmitted infections are less likely to experience PID than those who are. This is true with or without an IUD. As with other effective methods, an IUD user who is at risk for STDs should use a barrier method as well. In order to prevent PID, the use of prophylactic antibiotics is sometimes recommended. However, large clinical trials have shown that with copper IUDs the rate of salpingitis is 1 per 1000, regardless of whether prophylactic antibiotics were given. In general, in the United States, prophylaxis is not routinely recommended. 159

A direct comparison of levonorgestrel-releasing systems and copper IUDs demonstrated a low PID rate over a 7-year period with both IUDs and no significant difference between the two types. 169 A large multicenter European study showed that after 36 months, the cumulative discontinuation rates for PID were 0.8% and 2.0% for the levonorgestrel and copper IUDs, respectively (P < .02); after 60 months, the corresponding rates were 0.8% and 2.2%. 170

Tubal infertility is closely related to the problem of PID. Conflicting results have been reported; however, a thorough review of the question concluded that fair evidence indicates no important effect of IUD use on infertility. 159
Pregnancy and Intrauterine Devices

Intrauterine Pregnancy

Pregnancy with an IUD in situ frequently leads to spontaneous abortion and, rarely, sepsis. Physicians recommend removal of the IUD when pregnancy has been diagnosed. There is no evidence that the risk of malformations is increased in infants born to women with an IUD in situ.

Ectopic Pregnancy

Studies with the modern copper and levonorgestrel-releasing IUDs have demonstrated a low incidence of ectopic pregnancy. For users fitted with IUDs releasing copper from a 200-mm² surface, the estimated rate of ectopic pregnancy was four tenths that of nonusers of contraception; the rate of ectopic pregnancy with devices releasing copper from a 350-mm² surface was one tenth of that of nonusers of contraception. With the hormonal IUD releasing 20 µg of levonorgestrel daily, the ectopic pregnancy rate was also one tenth of the rate for nonusers. These studies dispel the worries of physicians and women about ectopic pregnancy with respect to the modern copper and levonorgestrel IUDs. It can be concluded that the modern IUDs reduce the risk of both uterine and ectopic pregnancy.
Selection of Candidates and Contraindications for Intrauterine Devices

Selection

As with any contraceptive method, contraceptive protection, acceptable adverse events, and continuation of use of IUDs depend on the proper selection of candidates. PID is best prevented by selecting candidates at low risk for PID and excluding the presence of PID during the preinsertion examination. Clinically manifest vaginal and cervical infection should be identified and treated, and bacteriologic examination of the cervix should be carried out when possible. The appropriate IUD candidate is a parous woman at low risk for sexually transmitted infection. IUDs are well suited for women who cannot use or do not desire hormonal contraception. It is a particularly important option for women considering surgical sterilization. A decade of easily reversible contraception may suit many women. Because the association between IUDs and PID is limited to the 3 weeks after insertion, it is prudent to observe the patient for symptoms of PID.

Contraindications

Insertion of an IUD is contraindicated when a pregnancy is suspected or confirmed and during an acute episode of PID. Caution should be exercised for women with a history of PID, undiagnosed uterine bleeding, enlarged or distorted uterus, confirmed or suspected uterine or cervical malignancy, untreated acute cervicitis or vaginitis, multiple sexual partners, and the presence of genital actinomycosis. Patients (and providers) sometimes have a sketchy memory; therefore, the presence of a previously inserted IUD must be excluded.
Noncontraceptive Benefits and Advantages and Shortcomings of Intrauterine Devices

The possible role of IUDs in carcinogenesis has been addressed in several population-based studies. Two well-conducted case-control studies have provided reassurance that IUDs do not increase the risk of endometrial cancer and suggested a possible protective effect, even in women older than 55 years, an age group with an increased risk of endometrial cancer. Two other case-control studies have shown a straightforward protective effect. The relative risk of developing an endometrial carcinoma was 0.51 and 0.61 in these two respective studies, with a 95% CI of 0.3 to 0.8. IUD use has no effect on the development of cervical carcinoma or breast tumors. An extended nonmedical benefit of IUDs is their low life-use cost. When used in the long term, they are the least expensive contraceptive agents.

Table 17-15 summarizes advantages and disadvantages of IUDs and shows that benefits outweigh shortcomings of intrauterine contraception.
Mechanism of Action

At least two studies have demonstrated that IUDs exert their contraceptive action before fertilization occurs.

An especially intriguing finding came from studies with volunteers wearing copper IUDs. Over several cycles, blood samples were examined for human chorionic gonadotropin (hCG) by extremely sensitive methods from day 10 of the cycle until the onset of menses. Concentrations of hCG were not elevated in any subject with an IUD. In the control group of women who wished to become pregnant, two showed an increase of hCG during the critical period of the luteal phase and this "chemical" pregnancy was later confirmed by obstetric examination. Further, failed attempts to retrieve fertilized ova support prevention of fertilization as the mechanism of action.

Several modes of contraceptive action of IUDs have been suggested. After insertion, IUDs incite a foreign body reaction in the uterine cavity, which includes rapid invasion of leukocytes and increased production of prostaglandins. The changed environment in the uterine cavity impairs the viability of the sperm and inhibits its motility. This reaction is particularly well expressed with the copper IUDs. The levonorgestrel IUDs inhibit endometrial growth and induce changes in the cervical mucus that make sperm penetration difficult. However, the contraceptive action of IUDs has not been clarified.
BARRIER METHODS

The Male Condom

The male condom is important not only as a contraceptive device but also as one that reduces the risk of both HIV and HPV infection, the latter being associated with condylomata acuminata and cervical cancer. Modern condoms were originally made of rubber, but since a liquid latex process was introduced in the 1930s, the bulk of the world's condom supply has been made of latex. The production process was gradually mechanized, which lowered the cost substantially. The condom is one of the least expensive contraceptive methods and is the best protection against the spread of sexually transmitted infection.

Condoms can also be made of polyurethane and silicone rubber as well as of lamb intestines. The "natural skin" condoms do not protect against the viral STDs.

The production standard requires that the thickness of a condom does not exceed 0.03 to 0.08 mm, and there are strict quality control safeguards against leakage and breakage. For example, FDA regulations require that randomly selected condoms from the production run be filled with 300 mL of water and inspected for leaks. In Britain, the condom is required to hold 3 L of water without breaking.

Latex can be weakened by oil-based lubricants and by various vaginal ointments or creams. Instructions for the use of vaginal antifungal preparations contain a warning that women should not rely on condoms for protection against pregnancy and STDs while using such preparations. Condoms made of polyurethane are not harmed by oil-based lubricants, but their breakage rate is significantly higher than that of latex condoms.

| TABLE 17-15 -- Advantages and Shortcomings of Intrauterine Devices |
|---------------------------------|------------------|
| **Advantages**                  | **Disadvantages** |
| Easiness of application, low costs | Transient increased risk of pelvic inflammatory disease |
| High protection against pregnancy | Expulsion |
| Highest continuation rate of all reversible methods | Uterine perforation |
| Immediately reversible | |
| Suitable for breast-feeding mothers | Candidates must be carefully examined for factors predisposing to the development of pelvic inflammatory disease |
| Protection against endometrial cancer | |
| No increased risk of cervical and breast cancer | Close follow-up during the first 34 wk after insertion is needed |
| Levonorgestrel-releasing IUDs control excessive uterine bleeding | Amenorrhea associated with levonorgestrel-releasing IUD causes discontinuation of use |

The surface of some types of condoms can be coated with an appropriate lubricant or spermicide, or both, to improve acceptability and efficacy.
The Female Condom

The female condom is much less frequently used than the male condom. The female condom is a polyurethane sheath 15 cm in length and 7 cm in diameter. A flexible ring is attached to its open end, and a loose removable ring inside the condom facilitates the insertion of the device into the vagina. The advantages of the female condom over the male condom are that the method is controlled by the woman and it is independent of the immediate sexual act. The female condom can be placed into the vagina several hours before intercourse takes place. \[179\]
The Diaphragm

Diaphragms are made of latex in the form of a thin dome mounted on a ring containing a flat or spiral spring. The sizes vary in diameter from 45 to 105 mm in steps of 2.5 to 5.0 mm. In determining the correct size of diaphragm, the distance between the posterior aspect of the symphysis pubis and the posterior fornix of the vagina is assessed. A correctly fitted diaphragm occupies this space with the dome covering the cervix. The diaphragm should always be applied with a spermicidal cream or jelly and should be left in place for a minimum of 6 hours after intercourse. After removal, the diaphragm should be washed and dried but no talcum powder should be applied.

Used consistently, the diaphragm has a low failure rate: 2% to 3%. This low failure rate is rarely achieved; in fact, rates of 20 to 30 pregnancies per 100 woman-years have been reported. The high failure rate results from poor motivation, inconsistency of use, poor fit of the diaphragm because the anatomy of the genitalia has changed (e.g., after delivery), incorrect insertion, slippage during intercourse, defects in the latex, and reluctance of users to apply the spermicide because it is "messy." A diaphragm should not be prescribed if there is an evident uterine prolapse, poor vaginal tonus, or a marked cystocele. After delivery, fitting of a diaphragm should be delayed at least a few weeks to allow rebuilding of muscular tone.

Clinicians speculate that diaphragms may reduce the risk of cervical gonorrhea, PID, and possibly cervical and endocervical malignancies. However, data on these subjects are not as supportive as those for the male condom. On the other hand, an increased incidence of urinary tract infection has been observed among diaphragm users. A few cases of toxic shock syndrome have been reported in cases in which the diaphragm was left in situ for 24 hours or longer.

In conclusion, a diaphragm with a spermicide is an inexpensive way to prevent pregnancy, and it is an excellent choice for women who do not wish to use one of the more effective methods, have infrequent intercourse, and are compliant. Such women are more inclined to use the method consistently, and consistency of use is the key to barrier success.
The Cervical Cap

In the United States, one type of cervical cap has been approved for general use. The Prentif cap is made of soft rubber, and a firm round rim is affixed to the open end. When successfully inserted, the rim fits around the cervix close to the vaginal fornices.

The efficacy is about the same as that of a diaphragm at best and can be ensured by filling the cup with a spermicide. For nulliparous women, the failure rate is 20% for typical use and 9% for perfect use. For parous women, the failure rate is higher, 40% for typical use and 26% for perfect use. The cap can stay in place for 48 hours irrespective of the frequency of intercourse during that time. The earliest time for removal of the cervical cap is 8 hours after intercourse. The major problems with the cap are difficulties with insertion and with proper fitting and the possibility of displacement. Also, some women find it difficult to remove.

[181]
NATURAL FAMILY PLANNING

Natural family planning restricts intercourse to the "safe period," that is, the period of "physiologic sterility" in each menstrual cycle. Assuming that ova are available for fertilization for only 24 hours and that spermatozoa remain highly viable for 48 hours, theoretically, there would be only 3 days in each cycle during which conception is possible. To identify the days of possible conception and decide which days are safe is difficult given the less than perfect regularity of the menstrual cycle. The calendar method is based on the expectation that ovulation takes place 12 to 16 days prior to the next menstruation. In this "rhythm" method, the woman records the length of 6 or preferably 12 cycles and determines the beginning of the fertile period by subtracting 18 days from the shortest period. Then she estimates the end of the fertile period by subtracting 11 days from the longest cycle.

Women practicing the cervical mucus method are instructed to observe the changes of the cervical mucus throughout the cycle and recognize the characteristics of the mucus during the fertile period. The symptothermal method combines the cervical mucus method with measurements of basal body temperature. With this method, abstinence begins on the day when mucus becomes clear, slippery, and stretchy (spinnbarkeit); this usually happens a few days before ovulation. Intercourse is resumed on the third or fourth day after the basal body temperature shifts to the luteal phase pattern.

There are various modifications of these three methods, and instruments have been designed to facilitate monitoring the safe period. The failure rate, when the methods are correctly practiced, can be as low as 2% to 3%. However, typical failure rates can be 20% or more.

The high typical failure rate for natural family planning results from several factors. The method depends on keeping meticulous records of menstrual cycles for 6 to 12 months in order to determine the safe period. This may be difficult for the busy modern woman. A less controllable reason is the unpredictable duration of the follicular phase of the menstrual cycle and consequently of the entire cycle. If the differences between the cycle lengths exceed 10 days, natural birth control should not be practiced.

In addition, in some individuals, emotional or physical stress can postpone the time of ovulation, creating the possibility that intercourse, based on previous calculations, may take place during the "unsafe" period. Although the efficacy of the rhythm method can be improved by daily measurements of basal body temperature, this too can be complicated because not all basal body temperature curves can pinpoint the occurrence of ovulation. Self-examination of cervical mucus can be somewhat awkward and the results even more difficult to interpret.

All in all, natural family planning requires alertness on the part of the woman and a high level of motivation, cooperation, and discipline in both partners. It also requires a clear understanding that pregnancy is a distinct possibility.
STERILIZATION

Female Sterilization

Female sterilization involves occluding or dividing the fallopian tubes. The approach of choice is laparoscopy as an outpatient procedure. Minilaparotomy or a full laparotomy is usually performed when complications are anticipated with laparoscopy, for example, when a patient has extensive adhesions after previous surgeries, after PID, and post partum.

The procedure is usually done using general anesthesia. Local anesthesia with sedation is also common. After proper insertion of the laparoscope into the abdominal cavity under visual control, the fallopian tube is grasped by an instrument approximately 1 to 2 cm from the isthmus, and electrocoagulation is performed to the extent of 2 to 3 cm. The tubes can also be occluded by insertion of clips or a silastic ring. A loop of the fallopian tube is pushed through a dilated ring, which is then allowed to contract, thus providing tight occlusion of the tube.

When sterilization is done through minilaparotomy or laparotomy, most surgeons prefer the simple Pomeroy procedure or partial tubectomy. A loop of the tube is formed by grasping the tube with a forceps and ligating the loop about 1 to 2 cm below the apex of the loop. The segment of the tube that forms the loop is then dissected, and the open ends of tube may be electrocoagulated. A vaginal approach to tubal ligation has also been developed.

Efficacy

All methods of tubal sterilization are highly effective in reducing the risk of pregnancy. The pregnancy rate after tubal sterilization is 0.1% during the first year after the procedure, although the rates vary by procedure. The long-term pregnancy rates are higher. In the United States, a Collaborative Review of Sterilization Group observed 10,685 women for 8 to 14 years after tubal sterilization and identified 143 failures, with pregnancies occurring later than 2 years after the procedure. The cumulative 10-year probability of pregnancy was 18.5 per 1000 procedures irrespective of which method was used. The most effective methods were partial postpartum salpingectomy and laparoscopic unipolar coagulation, with 7.5 pregnancies per 1000 procedures. The clip was associated with 36.5 pregnancies per 1000 procedures. An overall cumulative probability of ectopic pregnancies was calculated as 7.5 per 1000 procedures.

Women sterilized at young age had the highest risk of pregnancy, most likely because they were being exposed to the possibility of recanalization or fistulization of the tubes for a longer period of time.

Complications

The most serious complications are those associated with bowel or blood vessel damage. The post-tubal sterilization syndrome includes a number of symptoms that women frequently ascribe to the procedure, such as premenstrual tension, dysmenorrhea, and emotional and psychosexual problems. The attributable reality of this syndrome is controversial. The length of the menstrual cycle is not influenced by tubal sterilization.

Manifestations of regret are more common in younger women. A functional reanastomosis may be difficult to achieve; however, success rates up to 70% have been reported. As discussed previously, the long-duration IUD and other methods should be discussed as alternatives to surgical sterilization.

Chemical Sterilization

Cyanoacrylates and quinacrine, a resin-like substance, have been evaluated as tubal occlusive materials. Neither has yet succeeded. For example, quinacrine causes local inflammation, fibrosis, and occlusion of the intramural segment of the tube. The method had been used on a large scale in India, where it was banned because of the associated high rate of pregnancies and complications. Data from Vietnam indicate a pregnancy rate of 13% for women younger than 35 years of age and 7% for women older than 35 years, higher occurrence of ectopic pregnancy, and PID.
Male Surgical Sterilization

Vasectomy

This procedure is done mostly using local anesthesia. The vas is palpated through the skin of the upper scrotum and exposed through a small incision. The vas is ligated at two sites, and the part between the two ligations is then surgically severed or excised. The vas may also be occluded by diathermy or by a clip.

Efficacy

It is important to convey to the patient that vasectomy is not immediately effective. It may take several weeks before all the sperm that remains in the distal part of the vas is ejaculated. An alternative method of contraception must be used until azoospermia is achieved, usually after 12 weeks or 20 ejaculations. In about 2% of cases, vasectomy is not successful and must be redone if sperm is present in the seminal fluid for several months. Rarely, spontaneous late reanastomosis can occur up to 10 years after vasectomy. Pregnancy that occurs such a long time after the procedure is a sensitive issue because questions of paternity may arise, and it must be handled judiciously.

Complications

Complications of vasectomy are rare and consist mostly of local infection and hematoma. Antisperm antibodies develop in about 50% of men, but there is no evidence that this is associated with any pathologic condition. [185]
SURGICAL ABORTION

Surgical abortion should not be a primary method of fertility control, although in some countries it is practiced as such. It is estimated that, worldwide, 50 million abortions are performed annually, 20 million of these illegally.

In the United States, after legalization of therapeutic interruption of pregnancy, the number of abortions rose from 16.3 per 1000 women in 1973 to 29.3 per 1000 in 1980. Thereafter, the rate of therapeutic abortions declined steadily to 22.2 per 1000 women in 1987. In that year, 1.33 million abortions were performed. Statistics also indicated that the majority (52%) of those undergoing abortion were younger than 25 years and 29% were teenagers.

In the United States, therapeutic abortions are performed by suction curettage under local or light general anesthesia. The mortality associated with therapeutic abortion is exceedingly low at 1 in 530,000 if performed within the first 8 weeks of pregnancy. This rate increases to 1 per 17,000 at 16 to 20 weeks and 1 per 6000 at 21 weeks of gestation. The overall mortality rate associated with abortion is 0.8 per 100,000 legal abortions. The maternal mortality rate at childbirth is 10 per 100,000 births, and that associated with ectopic pregnancy is 50 per 100,000 cases. The risk of nonfatal complications is less than 1%. These complications include pelvic infection, hemorrhage requiring a blood transfusion, and unintended major surgery.

The majority of abortions, 57%, are performed during gestations of less than 6 to 8 weeks duration. Thirty percent of women have abortions between weeks 9 and 12, and 12% have abortions later than 15 weeks. Approximately 50% of women having an abortion later than 15 weeks indicate that the delay was caused by problems of finding abortion facilities and having the procedure done. Teenagers are more inclined to delay abortion than older women.
DISCUSSING CONTRACEPTIVE CHOICES WITH THE PATIENT

Before recommending a specific method of contraception, the health status of the woman and her personal circumstances must be thoroughly evaluated. Crucial in the contraceptive decision-making process is whether the woman wishes to preserve her fertility or has attained her reproductive goals and is ready for permanent cessation of childbearing. The contraceptive choice is further influenced by the age of the patient, her smoking habits, her weight, and the presence or a history of health problems such as cardiovascular and cerebrovascular diseases and diabetes mellitus.

The initial visit and the yearly follow-up examinations are an opportunity to conduct a complete physical examination, take Pap smears, check for the presence of vaginal and sexually transmitted infections, and perform breast examinations. Mammography may be recommended, depending on the findings, the age of the woman, and her family history of breast disease. It is also prudent to check the lipid profile and blood sugar. The physician should discourage smoking, particularly if the contraceptive choice is hormonal.

Sexually active women frequently need protection not only against pregnancy but also against STDs, particularly in regions with a high prevalence of HIV and AIDS, chlamydia, gonorrhea, and syphilis infections. Dual protection, such as oral contraception along with the use of a condom, may be recommended, depending on the individual circumstances of the woman. Hormonal contraception for teenagers may be started after menarche. At the other end of the spectrum, women who have completed their families may choose an IUD or surgical sterilization. The important watchword is individualization of choice among all appropriate methods.

Lactation provides protection against pregnancy only if the woman is exclusively breast-feeding, but even such women can experience breakthrough ovulations, particularly after the 10th postpartum week. As long as a woman is breast-feeding, combination hormonal contraceptives should be avoided because they suppress production of breast milk. If breast-feeding is interrupted by bottle feeding, the risk of breakthrough ovulations increases significantly. For lactating women, progestagen-only contraception is an excellent choice irrespective of whether they breast-feed exclusively or supplement breast-feeding by bottle feeding. Low-dose progestagen-only contraception does not affect the composition and quantity of breast milk and can be started immediately after delivery.

In nonlactating women, uterine bleeding followed by ovulation usually occurs after the sixth week of the puerperium but can happen earlier. Therefore, nonlactating women should not delay contraception for more than 2 to 3 weeks. Combination OCs have not been recommended before the third week after delivery because of concern that they could compound the naturally increased risk of postpartum thromboembolism. In nonlactating women, a progestagen-only OC or an effective nonhormonal contraceptive method, such as an IUD, can be started immediately post partum.

After an uncomplicated spontaneous or induced abortion, women can start OCs on the same day. Provided there are no contraindications, an IUD can be inserted at the conclusion of the procedure.

Women who have achieved their desired family size should be offered surgical sterilization or an appropriate reversible method, such as a copper IUD.

It is good clinical practice to schedule women for a follow-up examination within 3 months after starting contraception. At that time, the physician determines whether the patient practices contraception correctly and inquires about possible adverse effects, such as chest pain, shortness of breath, edema, headaches, blurred vision, and depression. If the findings are satisfactory, the patient typically moves to an annual visit schedule.

An important part of recommending a contraceptive method is counseling about expected adverse events and pointing out those that are transient and benign versus those that require medical attention. Factual and balanced information about adverse events frequently averts discontinuation of the contraceptive method by the patient but can also facilitate therapeutic intervention if required.
CONTRACEPTION FOR WOMEN WITH HEALTH PROBLEMS

Recommending an adequate contraceptive method to women with health problems is always a challenge. In women with compromised health, both pregnancy and contraception could bring about a deterioration of the underlying disease. The risks of contraception versus the risks of pregnancy must be carefully weighed. Extensive and sensitive counseling by health personnel is especially important for women with medical problems because erroneous information can further imperil their health. Support of the partner can also be a positive influence on the success of a contraceptive method.

Both physician and patient must collaborate fully in making a decision concerning which contraceptive method to use. The seriousness of the disease in question must first be evaluated. When there are reasons to believe that pregnancy would permanently worsen the patient's health or be fatal for her, sterilization of one of the partners should be discussed. When pregnancy would substantially impair the patient's health but would not be life-threatening, long-term methods with high efficacy and safety could be recommended, for example, copper or levonorgestel-containing IUDs or hormonal methods when not contraindicated. Certain diseases do not impose major safety risks with respect to pregnancy. Should that be the case, patients who express the desire to preserve their fertility may use a wide range of methods according to preference and clinical compatibility. Before prescribing, the physician should take a thorough inventory of the patient's medications because certain preparations decrease the effectiveness of hormonal contraception.

Contraceptive Choices in Individual Diseases

Cardiovascular Disease

For a patient with any heart disease, a copper IUD may be the method of first choice because of its high efficacy and lack of systemic effects. The insertion of the IUD should be covered by antibiotics if the patient is at risk for bacterial endocarditis. Patients who receive antiocoagulation therapy can be fitted with a levonorgestrel IUD; the progestagen released suppresses the growth of the endometrium. Combination OCs should not be recommended if the patient is currently suffering from or indicates a personal history of MI, congestive heart failure, stroke, uncontrolled hypertension, or thromboembolism. Women with varicose veins can use OCs. Uncomplicated valvular heart disease, including mitral valve prolapse, is not a contraindication to hormonal contraception. If valvular disease is complicated by pulmonary hypertension, atrial fibrillation, subacute bacterial endocarditis, and signs and symptoms of cardiac congestion, combination OCs must be avoided.

For those with a number of heart conditions, progestagen-only contraception can be used in the form of pills, injections, and subdermal implants. These conditions include a history of thombophlebitis, valvular heart disease (uncomplicated and complicated), and controlled hypertension with systolic blood pressure 140 mmHg/90 mmHg. A history of ischemic heart disease or stroke does not preclude the use of progestagen-only contraception, but it should be used only if other options are not available or acceptable to the patient and if the patient can be properly monitored. In the United States, the product label of progestagen-only contraceptives is as restrictive as that for combination OCs. However, international guidelines permit a wider range of indications for progestagen-only contraception.

Diabetes Mellitus

In diabetes mellitus, the frequency and severity of vascular complications increase with age and in pregnancy. Diabetic patients should be advised of these possibilities and counseled on early family planning decisions. Effective contraception is critical for the health of diabetic patients. The contraceptive method of first choice should be the copper IUD, which is suited for patients with uncomplicated as well as with complicated diabetes mellitus.

The American College of Obstetricians and Gynecologists developed guidelines recommending that use of combined OCs be limited to nonsmoking, otherwise healthy women with diabetes who are younger than 35 years and show no evidence of hypertension, nephropathy, retinopathy, or other vascular diseases.

Combined OCs do not impair the course of either type 1 or type 2 diabetes. The critical question of whether OCs increase the risk of early diabetic renal or retinal complications, or both, has been explored in a retrospective case-control study of two matched groups of patients with type 1, insulin-dependent diabetes mellitus. The study group patients took OCs for at least 1 year; the control group consisted of never-users of OCs. During the observation period, markers of diabetic renal damage and results of eye examinations were not significantly different in the two groups; the longitudinal hemoglobin A1c values were similar for both study and control subjects. The study indicates that the use of OCs by young women with insulin-dependent diabetes mellitus does not pose an additional risk for the development of early diabetic retinopathy or nephropathy. A prospective cohort study of 98,000 women nurses found no increased risk of type 2 diabetes when subjects were taking OCs. A history of gestational diabetes does not preclude the use of hormonal contraception. If the health status and desire of the patient indicate, sterilization should be considered.

Liver Disease

The interaction between the liver and contraceptive hormones is manifold. In experimental animals, conditions resembling peliosis hepatis can be induced by estrogens, and the possibility of liver adenomas in OC users has been considered. The incidence of these tumors is rare, 1 to 3 per 1 million women per year, and the risk attributable to oral contraception is difficult to assess. An increased risk for development of primary hepatocellular carcinoma in OC users has not been proved. In patients with impaired liver function, contraceptive steroids may not be efficiently metabolized, and such patients may have increased circulating levels of these hormones. The clinical impact of these biochemical changes is uncertain. Gallbladder disease has been linked to oral contraception, but epidemiologic evidence has not been consistent.

A copper IUD should be considered the method of first choice for patients with liver problems. As stated in the section "Contraindications," the use of combined hormonal contraception should be avoided in patients with active liver disease, severe cirrhosis, and liver tumors or even a history of liver tumor. Under close supervision, patients with compensated cirrhosis can use progestagen-only contraception. There are no restrictions for the use of any type of contraception for carriers of hepatitis viruses.

Human Immunodeficiency Virus Infection

HIV-positive women have shown less inclination to procreate than healthy women, irrespective of their socioeconomic status and background. Sexually active HIV-infected women who desire contraception should practice procedures that are highly effective in preventing an unintended pregnancy and at the same time protect them from acquiring another STD and prevent transmission of the HIV infection to the woman's sexual partner. These objectives are best achieved by the dual-protection technique, with which the HIV-infected woman uses an effective contraceptive method (OC, medicated IUD, possibly tubal sterilization) and she or her male partner uses a condom.

HIV-positive women who want to become pregnant or are already pregnant should be counseled on the availability of treatment and prevention of vertical transmission of the infection from mother to infant.

Patients with Compromised Mental Health

Patients with reduced or limited mental abilities have a right to a safe sexual life. This has to be balanced with the right of caregivers, for whom a pregnant mentally ill patient can cause substantial distress. Also, one has to take into account a patient's need for adequate parental care. For all these reasons, the choice of contraception is crucial. Before the choice is made, several questions must be answered. Is the patient competent enough to decide upon the right form of contraception? Is she able to use a method consistently and correctly? Are bleeding and hygiene manageable? In no case must the patient be forced to use a
Hormonal contraception is not contraindicated for patients with epilepsy; however, one has to bear in mind that antiepileptic drugs may decrease the effects of contraceptive hormones. Similarly, rifampin decreases the effects of hormonal contraceptives in patients with tuberculosis. In both conditions, the choice of appropriate contraception should be individualized. For example, some women who take antiepileptic medication do well with a relatively higher dose OC. On the other hand, drugs used for the treatment of malaria do not affect contraceptive hormones, and oral contraception is not contraindicated in patients with malaria.

A coincidence of migraine and stroke has been recognized in several epidemiologic studies. A large-scale prospective epidemiologic study in the United States followed up 12,220 subjects between the years 1971 and 1984 and was published in 1997. The results of the study strengthened previous evidence regarding a nonrandom association of migraine and severe nonspecific headaches with a significantly increased risk of stroke, particularly among young women. In women younger than 45 years, all cases of stroke occurred in OC users. However, the study was conducted during a period when use of low-dose OCs was not common. This probably also applies to a case-control study examining the risk of ischemic stroke in young women with migraine. In this study, stroke was strongly associated with migraine with aura (odds ratio 6.2) as well as without aura (odds ratio 3.0). OCs increased the risk of ischemic stroke to 13.9. Another study estimated the odds ratio in migrainoid women using COC as 16.9.

Contraindications to the use of combined hormonal contraception are migraines with aura, migraines without aura if they are unusually severe or last more than 72 hours, migraines treated with ergot preparations, and all types of migraines when other risk factors for stroke are present. These contraindications are included in the British practice guidelines. The American College of Obstetrics and Gynecology recommends methods other than OCs for women with migraines. This prudent approach to combined OCs is warranted because an array of reliable nonhormonal methods is available today. Progestagen-only oral contraception can be tried in patients with simple migraines.

Patients with concomitant diseases who want and need contraception require special care and possibly more frequent follow-up. With the present availability of a wide variety of contraceptives, no health-compromised patient needs to be without protection from pregnancy.
THE FUTURE OF FAMILY PLANNING: NEEDS AND RESEARCH

Family planning in the second part of the 20th century was a remarkable success; almost 60% of couples of reproductive age now use contraception, and the fertility rate worldwide decreased nearly to the replacement level (see Fig. 17-3). Is this a reason for satisfaction? The number of unwanted pregnancies, over 2.5 million in the United States alone and 50 million worldwide, sends us the message that there is ample room for improvement.

Human Immunodeficiency Virus-Acquired Immunodeficiency Syndrome

Today’s major medical problem is HIV infection and AIDS. With respect to reproductive health care, the research challenge is to develop a reliable microbicidal-spermicidal agent that will simultaneously protect against the two dangerous viral infections, HIV and HPV. Another research task is to design ways to prevent the transmission of the HIV infection from mother to fetus. Finally, prevention of pregnancy in women with HIV-AIDS requires education of infected women with regard to an effective method of fertility control and provision of suitable means including financial to do so.

The statistics of the HIV-AIDS epidemic are chilling: HIV-AIDS affects 16 million women and 1.4 million children younger than 15 years worldwide. The yearly death toll for women is 1.3 million and for children younger than 15 years is 500,000. To protect women from HIV-AIDS and to prevent transmission of the infection to infants is an urgent task.
Expanding Contraceptive Choices

There are several goals at the heart of expanding contraceptive choices that include tailoring contraception to the individual needs of the woman or the couple and designing new delivery systems for existing hormones. A meaningful contribution has been the successful completion of clinical trials with a contraceptive patch and the current research on vaginal rings releasing contraceptive hormones. These and other parenteral delivery systems for contraceptive steroids have been described in the appropriate section of this chapter.

With respect to the development of new nonhormonal methods of fertility control, most of the current efforts involve modifications of previous designs.
Departure from 1-Month to 3-Month Regimen

Traditionally, OC treatment was designed for 1-month use. However, a number of women perceive monthly withdrawal bleedings as a nuisance and would prefer to reduce the number of bleeding episodes and symptoms accompanying such episodes. For many years, physicians have prescribed continuous combined OCs to avoid withdrawal bleeding. Currently, a 3-month regimen of ethinylestradiol and levonorgestrel is being investigated in the United States. The effect on the endometrium and evolution of endometrial hyperproliferation at the end of the 3-month dosing period must be investigated and defined.
Contraception for Nursing Women

There is still a need for additional contraceptive options for breast-feeding women. Although lactation reduces fertility substantially, unpredictable breakthrough ovulations do occur, particularly when breast-feeding is not exclusive or is irregular. Pilot studies with breast-feeding women have shown that small amounts of estradiol delivered by transdermal patches significantly reduce pituitary gonadotropins more than in untreated breast-feeding women. The growth of ovarian follicles was also significantly suppressed in the treated women. Before this information can be translated into a practical contraceptive method for lactating women, many questions must be answered, including the long-term effects of low doses of estrogen on the mother, the transfer of estrogen into the breast milk, and possible effects on the infant. For now, barrier and progestagen-only methods remain most reasonable.
Immunocontraception

Since 1975, a WHO research program has devoted considerable effort to developing an anti-hCG antibody that would protect women from pregnancy for 6 to 12 months. Although the research has yielded an impressive amount of basic scientific data, human application had to be discontinued because of unexpected adverse events. New immunologic leads are being pursued. Antisperm vaccines, antiovum vaccines, and anti-GnRH vaccines are also being contemplated but have not crossed from the laboratory into the clinic.
Developing a Male Contraceptive

After years of trying various approaches, the only effective and well-tolerated methods of male contraception are vasectomy and the condom. Attempts have been made to develop systemic male contraception using the same principles as in female contraception, that is, blocking the pituitary gonadotropins in order to reduce sperm production in the testes to the point at which the man becomes infertile. Testosterone enanthate alone suppresses pituitary gonadotropins and interferes with spermatogenesis. Combination of this androgen with a GnRH agonist or a GnRH antagonist achieved an impressive reduction of spermatogenesis without affecting the sexual function of the men studied.

The addition of a progestagen to testosterone has been associated with more rapid and effective suppression of spermatogenesis than with testosterone alone. Therefore, considerable research efforts have been directed toward testing combinations of testosterone with contraceptive progestagens, including levonorgestrel, cyproterone acetate, MPA, norethindrone enanthate, and desogestrel. The major disadvantage of testosterone is its conversion to dihydrotestosterone and the consequent stimulation of the prostate gland. The negative effect of testosterone on lipid metabolism is also of concern. Although treatments with testosterone alone and in combination with either GnRH agonists and antagonists or with various progestagens have contributed substantially to our knowledge of spermatogenesis and its suppression, a practical method of male contraception has not been developed.

However, one approach is promising. The Population Council is conducting clinical testing with steroid hormones of the 19-nortestosterone series, notably with 7-methyl-19-nortestosterone. The 7-methyl group protects the compound from conversion to dihydrotestosterone, therefore, its effects on the prostate gland are limited. However, 7-methyl-19-nortestosterone is 10 times more potent in suppressing pituitary gonadotropins. Because of its lack of oral bioavailability, the compound has been formulated in a subdermal implant.

Human studies have progressed to initial clinical pharmacology testing. One-month dose-response studies have identified the daily amount of 7-methyl-19-nortestosterone that successfully suppresses both luteinizing hormone and follicle-stimulating hormone, and studies of the effect of 7-methyl-19-nortestosterone on spermatogenesis in humans are in progress.
The Future

In the search for a new method of male contraception, it might be necessary to depart from the traditional hormonal approach and look for potential leads in the molecular regulation of sperm biology. One approach would be to interfere with sperm capacitation in the epididymis. Maturation of the sperm surface composition is critical for fertilization of the egg, and any disruption can impede sperm progress through the cells and carbohydrate matrix of the cumulus oophorus and the glycoproteins of the zona pellucida. Decapacitated sperm may attach to the surface of the cumulus but fail to penetrate it.

During epididymal transit, the sperm undergoes maturation, a series of processes that have been only partially elucidated. Incomplete sperm maturation within the epididymis could explain the failure of sperm binding to the zona pellucida in unsuccessful human in vitro fertilizations. A group of newly discovered proteins (HE2 and HE2') that are synthesized in the epididymis and secreted into that tubule might interact with sperm and affect their maturation. If the role of these epididymis-specific proteins in sperm maturation is confirmed and defined, they could be a source of potential contraceptive targets. This approach is an attempt to apply the lessons of molecular cell biology in the clinic.

With respect to hormonal contraception for women, the trend has been to introduce new types of progestagens with minimal metabolic impact and a specific action on the endometrium. Is there really a need for new progestagens? Progestagens of the advanced gonane series, such as norgestimate and desogestrel, already have no negative effects on the metabolism of lipids, proteins, and insulin, and they may be the answer to the quest for a progestagen with minimal metabolic action.

With respect to estrogens, we are still using a compound synthesized in biochemically prehistoric times in 1938. Perhaps we should start looking for a selective estrogen receptor modulator that suppresses the pituitary gland with no or little effect on the endometrium, no effect on the blood clotting mechanisms, and no potential for thromboembolism. To date, no product is available that meets these criteria.

The development of new contraceptive modalities depends on close cooperation between three institutions: academia, government agencies, and the pharmaceutical industry. This is true for the development of any drug, but it is perhaps more important for the development of contraceptives because contraceptive research involves a number of public health issues and, in many instances, financial support for reproductive research depends on public opinion and attitudes.
References


References


Chapter 18 - Disorders of the Testes and the Male Reproductive Tract

James E. Griffin
Jean D. Wilson

The testes produce sperm and the hormones that regulate male sexual life; both functions are controlled by the hypothalamic-pituitary system. The pathways for hormone formation and the regulatory control of the testes are similar to those in the ovaries and the adrenal glands, and all steroid hormones work by similar mechanisms. However, the major steroid hormone of the testes, testosterone, has few direct actions; instead, it serves as a circulating prohormone (or precursor) for formation of 5-reduced androgens and estrogens, metabolites that mediate most androgen actions.

The regulation of testicular hormone production and the mechanisms of hormone action are similar at all stages of life, but the physiologic effects of these hormones differ at different times of life; for example, inducing formation of the male urogenital tract during embryogenesis and promoting sexual maturation at the time of puberty. As a result, abnormal testicular function causes different consequences depending on when it develops, from early gestation to old age.
DEVELOPMENT OF THE TESTES

Embryogenesis

Testicular differentiation is controlled by genes on the Y chromosome. The short arm of the Y is invariable in size, whereas in normal men the long arm can vary considerably in length. The short arm contains genes responsible for testicular development, and additional Y-encoded genes are essential for spermatogenesis. The short arm of the Y chromosome is composed of two distinct regions: the so-called pairing segment on the distal end of the short arm is homologous to a region on the end of the short arm of the X chromosome and is responsible for the pairing between the X and Y that is essential for complete segregation of the sex chromosomes during meiosis. Recombination can occur between the shared regions of the X and the Y. Genes and sequences in this region of the X chromosome fail to exhibit typical X linkage, and the pattern of inheritance of such genes is termed pseudoautosomal.

The region of the short arm of the Y chromosome between the pairing segment and the centromere encodes genes that do not recombine with the X chromosome, including the SRY (sex-determining region Y chromosome) gene responsible for testis determination. The SRY gene is Y-chromosome specific, conserved among mammals, and expressed principally in the testes, and female mice carrying an SRY transgene develop into phenotypic males with testes. SRY encodes a protein of the HMG (high-mobility group) box type that probably acts to influence chromatin structure, and other genes in the sex determination cascade, including the steroidogenic factor 1 gene (SF1), Wilms' tumorrelated gene 1 (WT1), the dosage-sensitive sex reversals adrenal hypoplasia congenital gene (DAX1), and the SRY-related genes HMG-box 3 (SOX3) and HMG-box 9 (SOX9) (Fig. 18-1). Mutations in WT1, SF1, SRY or any of the downstream genes under its control can impair testicular development (see Chapter 22).

Testes contain three principal cell types:

1. Germ cells, derived from primitive endodermal cells of the inner cell mass (initially identifiable in the yolk sac).
2. Supporting cells, derived from the coelomic epithelium of the gonadal ridge that differentiate into the Sertoli cells in the testis (or granulosa cells in the ovary).
3. Stromal (interstitial) cells, derived from the mesenchyme of the gonadal ridge that differentiate into Leydig cells.

The primordial germ cells express specific cell markers and are recognizable in the 4.5-day-old human blastocyst. Before day 23 of human gestation, these cells are located in the dorsal and caudal portions of the yolk sac entoderm (Fig. 18-2A). They then migrate by ameboid movement from the gut endoderm through the mesentery to reach the genital ridge (see Fig. 18-2B). The forces that control the migration are unknown, but the germ cells replicate during migration so that more cells reach the genital ridge than were present in the yolk sac. On reaching the genital ridge the germ cells, together with adhering epithelial cells, infiltrate the underlying mesenchyme. This process is identical in male and female embryos and culminates by 5 to 6 weeks of gestation in the formation of the genital blastema containing the three basic cell types. Primordial germ cells that fail to reach the genital ridge degenerate or differentiate into other cell types and may serve as the progenitors of extragonadal germ cell tumors in later life.

Sexual dimorphism of the gonad begins between 6 and 7 weeks of gestation with the development of seminiferous cords in the fetal testis. By contrast, histologic differentiation of the fetal ovary is not apparent until the sixth month of gestation, when granulosa cells organize around the dividing oocytes to form the primary ovarian follicle. The somatic cells of the gonad can undergo partial organization into ovary or testis as specified by the sex chromosomal even when the germ cells are prevented from migrating to the genital ridge, suggesting that some determinants for gonadal development are inherent in the cells of the genital ridge.
Testicular Descent

Histologic development of the testes is largely complete by the end of the third month of gestation, whereas testicular descent from the abdominal cavity to the scrotum occurs later. Between 10 and 15 weeks of human gestation, the testes remain anchored to the future inguinal canal by the caudal ligament of the testis, the gubernaculum, whereas the ovary moves cranially. Simultaneously, the cranial suspensory ligament that anchors the testes to the posterior abdominal wall regresses. The gonadal positions in the two sexes deviate further after 25 weeks of gestation, when the gubernaculum descends into the scrotum and begins to degenerate as it is hollowed out by a diverticulum of the peritoneum, the processus vaginalis.

The actual descent of the testes through the processus vaginalis into the scrotum occurs as intra-abdominal pressure increases as a consequence of the closure of the umbilical cord, descent being completed between 7 months of gestation and shortly after birth. Continued development of the abdominal musculature causes closure of the inguinal rings and obliteration of the processus vaginalis. Conditions that impair development of intra-abdominal pressure, such as congenital defects in the abdominal musculature, are associated with cryptorchidism.

The genetic and endocrine factors that control testicular descent are now understood in large part. The role of androgens has been established by two types of evidence: (1) dihydrotestosterone promotes testicular descent in rats, and (2) the position of the testes in 46,XY subjects with mutations that impair the androgen receptor correlates with the severity of impairment of receptor function. Thus, androgen action may involve the formation, enlargement, or degeneration of the processus vaginalis or the gubernaculum. Androgens may also enhance the release of calcitonin generelated peptide, from the genitofemoral nerve, which in turn may promote descent.

Antimüllerian hormone (AMH) (also called müllerian-inhibiting substance and müllerian duct inhibitor) may also play a role in testicular descent, possibly in the contraction of the gubernaculum. The testes in some men with persistent müllerian duct syndrome are located high in the retroperitoneal space as a result of impaired formation or action of AMH. The INSL3 gene (also designated Ley I-L and relaxin-like factor, RLF), a member of the insulin-like superfamily, controls development of the gubernaculum, and mice with targeted disruption of INSL3 have bilateral cryptorchidism with freely moving intra-abdominal testes. Diethylstilbestrol-induced cryptorchidism in mice is also associated with failure of gubernaculum development and impairment of expression of INSL3. A similar phenotype, namely long gubernacular cords and intra-abdominal testes, results from targeted disruption of the homeobox gene HOXA-10.
STRUCTURAL ORGANIZATION OF THE TESTES

The testes contain a network of tubules for the production and transport of sperm to the excretory-ejaculatory ducts and a system of interstitial or Leydig cells that synthesize androgens. Tight junctions between the Sertoli cells separate the spermatogonia from the primary spermatocytes and form a diffusion barrier that divides the testis into two functional compartments: basal and adluminal. The barrier between these two compartments has limited permeability to macromolecules, analogous to the blood-brain barrier and other epithelial barriers. The basal compartment consists of the Leydig cells, the boundary tissue of the tubule including peritubular myoid cells, and the outer layers of the spermatogenic tubules that contain the spermatogonia. The adluminal compartment consists of the inner two thirds of the tubules, including primary spermatocytes and cells in more advanced stages of spermatogenesis.

The structure and function of the Sertoli cell are closely linked (see Fig. 18-4). The base of the cell is adjacent to the outer basement membrane of the spermatogenic tubule, whereas the inner portion consists of an arborized cytoplasm containing large gaps or lacunae, analogous to the branches of a tree. The mechanism by which the spermatogonia pass through the tight junctional complexes between the Sertoli cells as spermatogenesis commences is not known, but the arborized cytoplasm of the Sertoli cell encompasses the differentiating spermatocytes and spermatids so that spermatogenesis takes place within the Sertoli cell cytoplasm network. Sertoli cells synthesize hormones such as AMH, inhibin, activin, and prodynorphin as well as factors essential for spermatogenesis such as transferrin.

The lipid droplets responsible for the foamy appearance of Leydig cell cytoplasm are composed largely of esterified cholesterol, derived in part from circulating lipoproteins and in part from locally synthesized cholesterol. The esterified cholesterol serves as a reservoir of substrate for testosterone synthesis. After hydrolysis of cholesterol ester, free cholesterol moves to mitochondria under the control of the steroidogenic acute regulatory (StAR) protein, where the initial reaction in testosterone biosynthesis takes place, namely side-chain cleavage of cholesterol to pregnenolone. Pregnenolone in turn is converted to testosterone in the endoplasmic reticulum. The amount of testosterone stored in the Leydig cell is small because newly synthesized testosterone diffuses promptly into the testicular venous blood.
PHYSIOLOGY OF TESTICULAR FUNCTION

Hypothalamic-Pituitary-Testicular Axis

Hypothalamic Hormones

The hypothalamus is connected to the pituitary gland both by a portal vascular system and by neural pathways (Fig. 18-5) (Figure Not Available). The portal vascular system provides a mechanism for the delivery of releasing hormones from the brain to the pituitary gland, the major system by which the brain controls anterior pituitary function. Reverse flow through this hypophyseal-portal circulation may also allow pituitary hormones to reach the brain by a more direct path than through the general circulation. The preoptic area and the medial basal region of the hypothalamus (particularly the arcuate nucleus) contain important centers for control of gonadotropin secretion. Peptidergic neurons in this region secrete gonadotropin-releasing hormone (GnRH), also called luteinizing hormone-releasing hormone (LHRH), in a pulsatile fashion. Neurons from other regions of the brain terminate in this area and influence both the frequency and the amplitude of GnRH secretory pulses through catecholamine-related, dopamine-related, and endorphin-related mechanisms.

GnRH is a decapeptide that is widely distributed in the central nervous system (CNS) and other tissues. However, a physiologic role has not been established for the hormone in sites other than the pituitary. The metabolic clearance rate of GnRH averages about 800 L/min body surface area per day. Increased excretion in urine of immunoreactive GnRH metabolites coincides with pubertal development in boys, and in adult men GnRH levels in urine correlate with those of luteinizing hormone (LH) and follicle-stimulating hormone (FSH).

Pituitary Hormones

LH and FSH are the primary pituitary hormones that regulate testosterone function. These hormones were named on the basis of their ovarian effects before their roles in testicular function were recognized. LH and FSH are secreted by the same basophilic cells in the pituitary. Like thyrotropin (or thyroid-stimulating hormone) and human chorionic gonadotropin (hCG), LH and FSH are glycoproteins composed of two polypeptide chains designated α and β. The subunits of the four hormones are identical; the individual immunologic and functional characteristics of the hormones are determined by unique subunits. Both subunits are required for full biologic activity.

The structures of the subunits of LH and hCG are similar except that the carboxy end of hCG contains an additional 30 amino acids and additional carbohydrate residues. The disappearance of exogenous LH from blood is described by two linear exponentials with half-times of 40 and 120 minutes, and the metabolic clearance rate is coupled primarily to G₁,4,5-triphosphate systems in the Leydig cell. Frequent pulses favor LH secretion. In some species, GnRH effects can be demonstrated in testes, but GnRH does not appear to have a direct effect on the human Leydig cell.

Mechanism of Action of GnRH-Releasing Hormone and Gonadotropins

GnRH interacts with high-affinity cell-surface receptors coupled to G proteins on the plasma membrane of pituitary gonadotrophs. The acute administration of GnRH stimulates the release of both LH and FSH by a mechanism involving calcium or phosphoinositides as second messengers. GnRH probably also acts long-term to enhance gonadotropin synthesis. The amounts of LH and FSH released in response to GnRH depend on age and hormonal status. In monkeys, the gonadotroph response to GnRH reaches a peak in the first few months of life and then declines and remains low until the onset of puberty, when it again increases to attain an adult response. Before puberty, the secretion of FSH in response to GnRH is greater than that of LH. Slow-frequency GnRH pulses favor FSH secretion, whereas frequent pulses favor LH secretion. In some species, GnRH effects can be demonstrated in testes, but GnRH does not appear to have a direct effect on the human Leydig cell.

The LH receptor on the plasma membrane of Leydig cells is a member of the superfamily of G protein-coupled, seven-transmembrane domain receptors. The binding of LH to the receptor activates signal transduction by both the adenylyl cyclase and phosphoinositide pathways (see Chapter 5). The intracellular loops of the receptor form contact sites for interaction with G proteins. In the testsis, receptor activation is coupled primarily to G_s proteins, leading to stimulation of adenylyl cyclase and formation of cAMP, which binds to the regulatory subunit of protein kinase and causes dissociation of the regulatory subunit and activation of the catalytic subunit. The activated protein kinase operates through unidentified steps to stimulate the synthesis of the enzymes of testosterone biosynthesis. The signal is terminated by endocytosis and degradation of the LH receptor complex.

In the intact testis and in cultured Leydig cells, the number of LH receptors decreases after administration of LH or hCG. The loss in receptor number is dose-dependent, reaches a nadir 24 hours after LH administration, is associated with a decrease in LH receptor messenger ribonucleic acid (mRNA), and returns to control levels within several days. This down-regulation of receptor number is associated with decreased responsiveness (desensitization) to subsequent LH administration. Desensitization cannot be solely the result of the decrease in receptor number and appears to result in part from inhibition of some postreceptor event because cAMP is ineffective in reversing desensitization. Whatever the mechanism, the diminished response of the Leydig cell to LH after administration of LH is a critical component of the regulation of testosterone production.

The primary site of action of FSH is the basal aspect of the plasma membrane of Sertoli cells, where the hormone binds to the FSH receptor, also a member of the G protein-coupled, seven-transmembrane domain receptor family. The second messenger is cAMP, which is also linked to the activation of protein kinase and stimulation of the synthesis of proteins such as androgen-binding protein and the aromatase that converts testosterone to estradiol. The precise role of FSH in the control of spermatogenesis remains uncertain and may vary among species (see later).

FSH plays an indirect role in androgen biosynthesis by inducing maturation of Leydig cells during development, possibly a consequence of the release of a paracrine factor by Sertoli cells; but FSH does not play a major role in the control of Leydig cell function in adults. Like LH and other peptide hormones, FSH regulates the number of its own receptors, but the physiologic significance of this phenomenon is unclear.

Regulation of Secretion of GnRH-Releasing Hormone and Gonadotropins

Episodic secretion of GnRH into the hypophyseal-portal system causes episodic secretion of both immunoreactive and bioactive LH. In adult men, LH secretory pulses occur at a frequency of 8 to 14 per 24 hours and vary in magnitude. Pulsatile secretion of FSH is temporally coupled to that of LH but is lower in amplitude. LH secretion is under negative-feedback control by gonadal steroids at the level of the hypothalamus and the pituitary gland (see Fig. 18-5) (Figure Not Available). Both testosterone and estradiol can effect this inhibition. Testosterone can be converted to estradiol in the brain and pituitary gland, but the two hormones are thought
One major effect of androgen in the CNS is to slow the hypothalamic pulse generator and consequently decrease the frequency of LH pulsatile release.

Endogenous opiates have a role in the negative-feedback actions of androgen and estrogen on pulsatile LH secretion in men. Furthermore, in monkeys with hypothalamic lesions that abolished endogenous GnRH release and in whom normal pulsatile secretion of LH was mimicked by chronic intermittent intravenous GnRH administration, bilateral orchidectomy caused minor elevations of plasma LH levels, whereas castration of monkeys with an intact hypothalamus given similar pulses of GnRH was followed by a marked rise in LH levels, indicating that LH is controlled by the negative feedback of gonadal steroids at the hypothalamic level. Acute infusions of estradiol also lowered LH levels associated with an increased frequency and a decreased amplitude of the LH pulses.

The fact that dihydrotestosterone, which cannot be converted to estrogen, exerts a negative-feedback control on LH secretion indicates that testosterone does not require aromatization to inhibit LH secretion. Testosterone also appears to have a negative-feedback action on LH secretion at the pituitary level because administration of exogenous testosterone to GnRH-deficient men who were given pulsatile LHRH infusions caused a decrease in mean plasma LH levels and in LH pulse amplitude. Hyperprolactinemia suppressed LH secretion, probably by inhibiting the pulsatile secretion of GnRH.

The negative-feedback control of FSH secretion involves peptide and steroid hormones from the testes. Serum FSH concentrations increase in proportion to the loss of germinal elements in the testis, whereas LH levels change little. Inhibin, a peptide secreted by Sertoli cells that inhibits pituitary FSH secretion, is a heterodimer consisting of a 20-kd subunit and a 15-kd subunit. The subunit occurs in two forms, so that there are two inhibins, inhibin A and inhibin B, each 31 kd in size. Inhibin B is thought to be the physiologically important hormone in men. FSH stimulates inhibin production, and both FSH and androgen are involved in normal inhibin production. Dimers of the subunit of inhibin (termed activin) stimulate FSH release, and the pituitary protein follistatin binds and inactivates activin.

Testosterone and estradiol also influence FSH secretion. In castrated rats treated with subphysiologic amounts of testosterone but physiologic amounts of estradiol, plasma FSH levels increase to the castration range but LH concentrations are normal. Indeed, alterations in the ratio of testosterone to estradiol can alter plasma FSH levels. In addition, varying the pattern of GnRH administration to hypogonadotropic men so that the same total dose was administered but with less frequent pulses caused selective increases in FSH levels.

In the rhesus monkey with a hypophysemotropic clamp, administration of inhibin maintained FSH levels in the precastration range when episodic gonadotropin secretion was maintained by intermittent GnRH infusion. Thus, the suppression of FSH secretion by inhibin does not require the action of testosterone and must take place at the level of the pituitary gland.
Androgen Physiology

Testosterone Synthesis and Secretion

The pathways of testosterone synthesis are illustrated in Figure 18-6 (Figure Not Available) and Figure 18-7. As stated earlier, the precursor steroid cholesterol can either be synthesized de novo or derived from the plasma pool by receptor-mediated endocytosis of low-density lipoprotein (LDL), and both sources are important in the human Leydig cell. 26

Figure 18-6 (Figure Not Available) Pathway of testosterone formation in the testis and the conversion of testosterone to active metabolites in peripheral tissues. SIAR, steroidogenic acute regulatory protein. (Revised from Griffin JE, Wilson JD. Disorders of the testes. In BraunwaldE, Fauci AS, Kasper DL, et al [eds]. Harrison’s Principles of Internal Medicine, 15th ed. New York, McGraw-Hill, 2001, pp 21432154.)

The conversion of cholesterol to testosterone involves several enzymatic reactions. The cholesterol side chain is cleaved in two steps to reduce the size from 27 to 19 carbons, and the A ring of the steroid is oxidized to the 4,3-keto configuration. The initial reaction in the process involves the transfer of cholesterol by the SIAR protein to the inner mitochondrial membrane, 26 where it undergoes side-chain cleavage by CYP11A1 to form pregnenolone. The subsequent conversion of pregnenolone to testosterone involves both random and ordered enzymatic reactions.

For the second side-chain cleavage to take place, 17-hydroxylation and cleavage of the 17,20 bond through CYP17 must occur before the reduction of the 17-ketone by 17-hydroxysteroid dehydrogenase III (17-HSD-III). In contrast, oxidation of the A ring by 3-HSD-II can take place at any stage in the process. Thus, the point in the pathway at which A ring oxidation occurs depends on the amounts and affinities of the enzymes for the various substrates and their compartmentalization within the cell. The predominant pathway in the human testis appears to be the 5 pathway (see Fig. 18-7, left side). A ring oxidation being the terminal reaction in the sequence. 26 Although

![Pathways of testosterone synthesis in human testis.](image)

**TABLE 18-1** -- Plasma Concentration of Spermatic and Peripheral Venous Steroids.

<table>
<thead>
<tr>
<th>Steroid</th>
<th>Concentration in Spermatic Vein</th>
<th>Concentration in Peripheral Vein</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ng/mL</td>
<td>ng/mL</td>
</tr>
<tr>
<td>Testosterone</td>
<td>3402000</td>
<td>100600</td>
</tr>
<tr>
<td></td>
<td>8.735</td>
<td>2.510</td>
</tr>
<tr>
<td>Dihydrotestosterone</td>
<td>228</td>
<td>0.680</td>
</tr>
<tr>
<td></td>
<td>0.31.6</td>
<td>0.10.45</td>
</tr>
<tr>
<td>Androsterone</td>
<td>1.438</td>
<td>0.411</td>
</tr>
<tr>
<td></td>
<td>0.51.4</td>
<td>0.150.4</td>
</tr>
<tr>
<td>Androstenedione</td>
<td>3.842</td>
<td>1.112</td>
</tr>
<tr>
<td></td>
<td>1.43.8</td>
<td>0.41.1</td>
</tr>
<tr>
<td>17-Hydroxyprogesterone</td>
<td>3.300</td>
<td>1.1100</td>
</tr>
<tr>
<td></td>
<td>1.23.3</td>
<td>0.41.1</td>
</tr>
<tr>
<td>Progesterone</td>
<td>3.535</td>
<td>1.111</td>
</tr>
<tr>
<td></td>
<td>0.31.9</td>
<td>0.10.6</td>
</tr>
<tr>
<td>Pregnenolone</td>
<td>3.535</td>
<td>1.112</td>
</tr>
<tr>
<td></td>
<td>0.93.0</td>
<td>0.31.0</td>
</tr>
</tbody>
</table>


**Blood samples from the spermatic vein were collected, with the use of local anesthesia, from patients with carcinoma of the prostate, hernia, varicocele, or hydrocele.**

17-hydroxylase and 17,20-lyase were originally regarded as separate enzymes, both activities are the function of a single cytochrome P450 enzyme CYP17. 26

The rate-limiting reaction in testosterone synthesis under most circumstances is the transport of cholesterol to the inner mitochondrial membrane. As mentioned earlier, the acute administration of LH stimulates testosterone synthesis by enhancing the delivery of cholesterol to the mitochondria through SIAR protein 26 for side-chain cleavage by CYP11A1. 26 In the steady state, LH also stimulates testosterone synthesis by enhancing the formation of CYP11A1 and other enzymes in the pathway.

Testosterone is the major product, but dihydrotestosterone, androsterone, 5-androstane-3,17-diol, androstenedione, 17-hydroxyprogesterone, progesterone, and pregnenolone are also formed by the tests in small amounts 26 (Table 18-1). The role of 5-reduced androgens formed in the tests has not been established, the major sites of steroid 5-reduction being extraglandular. The functions in the male of pregnenolone, progesterone, and 17-hydroxyprogesterone are not known.

Androstenedione serves as a substrate for extraglandular estrogen formation (see later).

The concentrations of testosterone in testicular lymph and testicular venous blood are similar, but the flow of testicular lymph is small so that the major route for steroid secretion into the circulation is through venous blood. The mechanism of transport of testosterone and other steroids from the sites of production to blood and lymph is not completely understood. Because only about 25 ng of testosterone is stored in the normal testes, the total testicular content turns over more than 200 times to provide the average of 6 mg secreted into plasma each day by normal men. 26

Gonadotropin Regulation of Testosterone Secretion

Rat Leydig cells have excess or "spare" LH receptors; that is, the physiologic response is maximal when only a fraction of the receptors are occupied by LH. These spare receptors neverthless coupled to cAMP generation so that maximal testosterone biosynthesis occurs with doses of hCG or LH that result in only 10% of maximal cAMP production. 26 Human Leydig cells contain fewer LH receptors than rat Leydig cells, but the fact that the maximal rates of testosterone biosynthesis
per Leydig cell are similar in both species suggests that the major difference is in the number of spare receptors rather than the number of receptors coupled to androgen synthesis.

The decreased response of target cells to LH or hCG after an initial exposure was described earlier and involves both reduced LH receptor number (down-regulation) and some effects distal to receptor binding. The possibility that hCG desensitization is mediated by a local increase in estrogen production by the testes is suggested by the findings that the testicular estradiol level is elevated within 30 minutes after hCG administration and that estrogen antagonists can block hCG-induced desensitization. In mouse Leydig cells, testosterone itself appears to act directly to control the activity of its biosynthetic pathway.

The long-term in vivo response to hCG is complex. Administration of hCG causes an elevated plasma testosterone level within 2 hours in both rats and humans. The plasma level of testosterone reaches a plateau or declines after the initial rise and then starts to increase again by 48 hours, and plasma estradiol levels increase and reach a peak 24 hours after hCG injection, a time corresponding to the nadir between the two testosterone responses.

The concept that the mechanism of this temporary desensitization involves inhibition of the 17,20-lyase reaction is supported by the fact that 17-hydroxyprogesterone increases 24 hours after hCG administration and then declines at 48 hours while testosterone levels are rising. Administration of high doses of hCG to men for long periods results in steady-state elevations of plasma estradiol, 17-hydroxyprogesterone, and testosterone. The increase in total plasma testosterone levels after hCG administration is greater in the morning than in the afternoon, in keeping with the normal circadian rhythm of serum testosterone levels. The higher testosterone levels in the morning in normal men do not correlate with changes in mean LH values but may be the consequence of changes in pulse frequency of LH secretion.

In rats, prolactin receptors are present on Leydig cells, and prolactin potentiates the effect of LH on testosterone synthesis, possibly by enhancing lipoprotein transport into the cells and increasing the availability of cholesterol for steroidogenesis. A role for prolactin has not been shown in human Leydig cell function, and the prolactin receptor is not expressed in human Leydig cells.

A variety of paracrine control mechanisms have been described in the testis, including:

1. Testicular peptides (inhibin, activin).
2. Growth factors (transforming growth factors and , fibroblast growth factor, insulin-like growth factor I [IGF-I]).
3. Immune-derived cytokines (tumor necrosis factor-).
4. Vasoactive peptides (endothelin, angiotensin II, and atrial natriuretic peptide [ANP]).

To date, however, no integrated view of the relevance of these various agents for Leydig cell function or for treatment of testicular defects has emerged.

Testosterone Transport in Plasma

In plasma, testosterone is largely bound to proteins, mainly albumin and sex hormone-binding globulin (SHBG, also called testosterone-binding globulin). SHBG, a -globulin composed of nonidentical subunits, has a molecular mass of about 95 k, contains about 30% carbohydrate, and has one androgen binding site per molecule. SHBG shares sequence homology with the coagulation factor protein S, laminin, and the growth arrest-specific gene (GASG). A variant allele of the SHBG gene is present in 13% of normal persons.

In the blood of normal men, about 2% of testosterone is free (unbound), 44% is bound to SHBG, and 54% is bound to albumin and other proteins. The fraction of testosterone bound to SHBG in serum is proportional to the SHBG level. Albumin has about a 1000-fold lower affinity for testosterone than SHBG, but the concentration of albumin is so much higher that the binding capacities are similar. Furthermore, protein-bound testosterone can dissociate within the capillary bed so that the active fraction is actually larger than the free fraction as estimated by equilibrium dialysis. In fact, nearly all albumin-bound testosterone is available for tissue uptake in vivo so that the bioavailable testosterone in men is about half the total (equal to the free plus the albumin-bound fraction). SHBG-steroid complexes can also bind to receptor sites on the cell surface and increase intracellular AMP levels by some mechanism not involving the intracellular androgen receptor.

Estradiol binds to SHBG with a lower affinity than testosterone, a consequence of which is that SHBG-bound estradiol is taken up by tissues. The multiple isoforms of SHBG arise from differential post-translational processing of the carbohydrate moiety.

The level of SHBG is increased by estrogens, as in normal pregnancy, and is decreased by testosterone administration. The plasma level in normal men is one third to one half that in women, and the level is higher than normal in hypogonadal men. SHBG levels are decreased in hypothyroidism and increased with thyroid hormone excess, possibly because of changes in estrogen formation.

The effects of changes in SHBG levels differ, depending on the physiologic state. In normal men, alterations in SHBG levels have little effect on androgen physiology in the steady state, for example, an increase in plasma SHBG is followed by temporary decreases in free (active) plasma testosterone and an increased rate of LH secretion followed by an increase in testosterone synthesis until the normal level of free hormone is reconstituted. As a consequence, in the steady state both increases and decreases in SHBG levels are compatible with normal rates of androgen synthesis and degradation and normal free hormone concentrations in plasma.

In contrast, changes in the plasma SHBG levels have profound consequences when the level of free hormone is not tightly regulated. Such is the case in two circumstances. First, with disorders of the hypothalamic-pituitary-testicular axis, the ability to regulate the free levels of the hormone is limited and replacement therapy may have to be adjusted when SHBG levels change. Second, and more important, even in men with an intact hypothalamic-pituitary-testicular axis, not all plasma hormones are under such tight regulation as testosterone.

The level of plasma estradiol in men is probably determined by the amount of androgen available as substrate for estrogen formation and by the amount of aromatase activity in extraglandular sites and thus is not regulated directly by the usual feedback mechanisms. Because SHBG binds estradiol less avidly than testosterone or dihydrotestosterone, increases in SHBG cause decreased hepatic clearance of testosterone but have little effect on the hepatic clearance of estradiol. Thus, in normal men, changes in SHBG levels can alter the ratio of androgens to estrogens even when androgen levels themselves are not significantly altered.

Extraadrenal Metabolism of Androgens

As noted earlier, testosterone serves as a circulating precursor or prohormone for the formation of two types of active metabolites, which in turn mediate many androgen actions (see Fig. 18-6) (Figure Not Available). On the one hand, testosterone can undergo irreversible 5-reduction to steroids such as dihydrotestosterone, which are responsible for many aspects of male sexual development and virilization. Dihydrotestosterone in turn can be further metabolized to 17-ketosteroids and polar derivatives excreted in urine. Alternatively, circulating androgens containing a 1,3-keto configuration can be converted to estrogens in the extraglandular tissues of both sexes. Estrogens in some instances act in concert with androgens to influence physiologic processes and in others exert effects that are independent of androgens. Thus, the physiologic actions of testosterone are the result of the combined effects of testosterone itself plus those of estrogen and androgen metabolites of testosterone.

In normal men, small amounts of estradiol (15% to 25% of the total daily production) and dihydrotestosterone are secreted directly by the testis, and small amounts of both can be formed indirectly from adrenal androgen through the sequence androstenedione estrone estradiol or androstenedione androstenedione dihydrotestosterone. The formation of estradiol in normal young men is illustrated diagrammatically in Figure 18-8 (Figure Not Available). Of the average daily
estradiol production of 45 pg, 17 µg is derived from the aromatization of circulating testosterone. 22 µg is formed extraglandularly from the weak estrogen estrone, and 6 µg is secreted directly by the testes. In some instances, these metabolites exert local actions in the tissues in which they are formed; in others, the 5-reduced and estrogenic metabolites can reenter the plasma and act as circulating hormones.

The factors that regulate estradiol and dihydrotestosterone formation are poorly understood. Circulating dihydrotestosterone is formed principally in androgen target tissues. Aromatization takes place in many tissues, the most significant of which is probably adipose tissue; the overall rate of extraglandular aromatization increases with age and body size.

In the past, androgen metabolism was characterized in considerable detail before it was recognized that some testosterone derivatives are themselves active hormones rather than inactive metabolites. For example, for many years it had been known that testosterone is converted in the body to a variety of 5-reduced and 5-metabolites. The finding that 5-dihydrotestosterone is the principal intracellular androgen in target tissues such as the rat prostate and and that dihydrotestosterone is more potent than testosterone in bioassay systems indicated that dihydrotestosterone mediates some androgen actions. The importance of dihydrotestosterone in normal androgen physiology has been confirmed both by physiologic studies and by studies of human mutations that impair dihydrotestosterone formation (see Chapter 22).

Two enzyme activities are responsible for 5-reductase activity in human tissues, one of which is defective in human steroid 5-reductase deficiency. These activities are due to two separate isoenzymes; steroid 5-reductase 1 is encoded by a gene on the short arm of chromosome 5, and the gene for steroid 5-reductase 2 is on the short arm of chromosome 2. The two enzymes have predicted molecular masses of 28 to 29 kd, share similar intron-exon organizations, and contain about 40% hydrophobic amino acids. Each is believed to be an intracellular enzyme. In the lipil biliary, 5-Reductase 2 is defective in subjects with 5-reductase deficiency, and administration to pregnant rats of finasteride, a potent inhibitor of 5-reductase 2, produced a phenocopy of 5-reductase deficiency in male embryos.

5-Reductase activity is under complex regulatory control. In the rat prostate gland, the enzyme activity is up-regulated by androgens, whereas thyroid hormones appear to be the major regulators in liver. In cultured human skin fibroblasts, the enhancement of enzyme activity by androgens appears to be mediated by IGF-1. In intact subjects, the production rate of dihydrotestosterone appears to be determined predominantly by the amount of testosterone available to serve as substrate for the reductases.

The distribution and ontogeny of the two isoenzymes are distinct; 5-reductase 2 is expressed in the male urogenital tract early in embryogenesis, whereas 5-reductase 1 is expressed in skin and liver predominantly after the time of puberty. Within the skin of the scalp, 5-reductase 1 is expressed in sebaceous glands, whereas isoenzyme 2 is the predominant enzyme in the hair follicle. Other steroids with 4,3-keto configurations, including progesterone and aldosterone, also undergo 5-reduction by the enzymes, but a role has not been established for other 5-reduced steroids in human physiology.

Estrogen formation involves hydroxylation, oxidation, and removal of the carbon at position 19 of the steroid molecule and aromatization of the A ring of the steroid. Estrogen formation in testes and in extraglandular tissues is catalyzed by the same enzyme that operates in placenta and ovary. Expression of aromatase in individual human tissues is determined by tissue-specific promoters that give rise to transcripts with different N-terminal noncoding sequences. Testosterone and androstenedione are substrates for aromatase, whereas 5-reduced steroids such as dihydrotestosterone cannot serve as estrogen precursors because reduction of the A ring precludes aromatization. Estrogen formation in the tests is regulated by gonadotropins (see earlier). The rate of the overall aromatase activity in nongonadal tissue is not influenced by castration or adrenalectomy but is enhanced with increasing body weight and with age.

The metabolism of testosterone is illustrated in Figure 18-9. The metabolites are excreted primarily in the urine (>90%), approximately half of the daily turnover being recovered as urinary 17-ketosteroids and the other half as a series of polar compounds, including hydroxylated metabolites and conjugates. Various excretory metabolites are thought to be largely inactive.

Androgen Action

Current concepts of androgen action in target cells are summarized in Figure 18-10. Major androgen functions include regulation of gonadotropin secretion by the hypothalamic-pituitary system, initiation and maintenance of spermatogenesis, formation of the male phenotype during sexual differentiation, promotion of sexual maturation at puberty, and control of sexual drive and potential in men (see Chapter 24).

Testosterone is believed to enter cells by passive diffusion. Inside cells that express 5-reductase, testosterone can be converted to dihydrotestosterone. Testosterone and dihydrotestosterone bind to the same high-affinity androgen receptor, and the hormone-receptor complexes attach to hormone response elements in deoxyribonucleic acid (DNA) to initiate biologic reactions. (See Chapter 5 for a discussion of the ancillary proteins involved in the regulation of transcription by the androgen-receptor complex.) Few of the presumed large number of acceptor sites in DNA have been defined, and the factors that determine the specificity of hormone response are poorly understood. However, interactions between the hormone-receptor complex and the hormone response elements either increase or decrease gene transcription.

The model of androgen action shown in Figure 18-10 is based on studies of androgen metabolism in animals and humans of various ages and on studies of single-gene mutations that impair androgen action. The testosterone-receptor complex regulates gonadotropin secretion and virilization of the Wolffian ducts during male sexual differentiation and is probably responsible for sexual dimorphism of muscle development. The dihydrotestosterone-receptor complex controls external virilization during embryogenesis and the development of most male secondary sexual characteristics during puberty, including androgen-mediated hair growth and loss.

The question of which hormone is involved in spermatogenesis is unresolved. On the basis of studies of androgen metabolism in rodent testis, it is generally believed that testosterone is the active hormone for this function; however, dihydrotestosterone is formed in the spermatogenic tubule and sperm production is impaired in subjects with 5-reductase 2 deficiency, raising the possibility that dihydrotestosterone may play a role in human spermatogenesis.

Androgen receptors are present in highest concentration in androgen target tissues such as the accessory organs of male reproduction and some areas of the brain. Tissues such as skeletal muscle, heart and vascular smooth muscle, placenta have small amounts of receptor, and in the tests androgen
receptors are present in both Sertoli cells and Leydig cells. Whether the presence of androgen receptors identifies a tissue as androgen-responsive is not clear. The amount of receptor present in a tissue may be affected by androgen or estrogen, by age, and by single-gene mutations and genetic polymorphisms.

A single androgen receptor binds both testosterone and dihydrotestosterone, and the receptor is encoded by a gene on the long arm of the X chromosome. The complementary DNA encoding the human androgen receptor predicts a protein of 917 amino acids and a molecular mass of about 99 kd. In its amino acid sequence, the receptor shares a high degree of homology with the progesterone, glucocorticoid, and mineralocorticoid receptors in the hormone and DNA binding domains. The N-terminus is only weakly homologous to comparable regions of other steroid hormone receptors and contains glutamine, proline, and glycine homopolymeric sequences. The length of the glutamine repeat is polymorphic in normal populations, and more than 90% of normal women are heterozygous at this locus. Very long glutamine repeat sequences are associated with a neurologic disease termed spinobulbar muscular atrophy (Kennedy's syndrome). The active form of the androgen receptor is a homodimer that forms as a result of interacting sites in the N-terminal and C-terminal domains of the protein; the homodimer in turn interacts with other proteins to form an active transcription regulatory complex.

If a single receptor mediates the action of both testosterone and dihydrotestosterone, why is dihydrotestosterone formation important for normal androgen action? A partial answer is that the affinity of the human androgen receptor for testosterone is less than it is for dihydrotestosterone, a difference that results in more rapid dissociation for the testosterone-receptor complex. Testosterone-receptor complexes are also less stable and transform to the DNA-binding state less efficiently.

Dihydrotestosterone formation may serve fundamentally to amplify the androgen signal rather than to allow interaction of the hormone-receptor complex with specific DNA target sequences.

Many issues in androgen action remain unresolved. For example, most acceptor sites for binding androgen-receptor complexes in nuclei, the so-called hormone response elements (HREs), are palindromes in DNA sequences located 5' to genes under control of the hormone and recognize specific sequences in the DNA binding domains of the receptors. However, the characterized glucocorticoid and androgen regulatory elements respond in vitro to either hormone-receptor complex, and other factors determine specificity of action for these hormones in vivo, including the site of location of the HRE in relation to the coding sequence under hormonal control, differences in binding affinities of the hormone-receptor complexes to the HREs, and participation of additional transcription regulatory proteins including suppressor proteins in the active transcription complex.

It is noteworthy that in most target tissues (the human and dog prostate glands being notable exceptions) androgen action is limited; for example, under the control of androgen, the penis increases approximately 10-fold in size during male puberty, but when some maximal level is achieved growth ceases regardless of the androgen status of the individual. This cessation of growth corresponds temporally with a decrease in the level of the androgen receptor in the penis, but it is not clear whether the decrease in androgen receptor level is the cause or the consequence of growth cessation.
Estrogen Physiology

As discussed earlier, estrogens may be either secreted directly by the testes or formed in peripheral tissues from 19-carbon precursors. The role of estrogen as an independent hormone in male physiology is incompletely understood but includes closure of the epiphyses, acceleration of the pubertal growth spurt, accrual and maintenance of bone density, influence on gonadotropin secretion, a role in male sexual drive or potencia or both, and control of epididymal function. Estrogen excess in men causes gynecomastia (see later).

These various actions appear to be mediated predominantly through the estrogen receptor. In addition, estrogen interacts with androgen action by mechanisms that are not fully understood. In the prostate, estrogens may enhance androgen action by increasing the number of androgen receptors, whereas under physiologic conditions estrogens do not cause growth of the male breast because androgens appear to act as weak antiestrogens and prevent the binding of estrogen to the estrogen receptor.
Spermatogenesis and Fertilization

Spermatogenic Cycle

Spermatogenesis involves three processes: (1) multiplication of the germ cells, (2) reduction of the number of chromosomes from the diploid to the haploid state (meiosis), and (3) formation of a superstructure that allows sperm motility as well as generation of energy to drive motility. This superstructure also protects the chromosomes against environmental damage and provides sperm with the capacity to penetrate the ovum.\(^{140}\)

After migration of the germ cells to the genital ridge is completed during the second month of gestation (see Fig. 18-2), the total number of spermatogonia is approximately \(3 \times 10^8\) per gonad, and by puberty this number increases to about \(6 \times 10^8\) per testis. As a result of the hormonal events accompanying puberty, a profound cellular proliferation ensues.\(^{141}\) The net result is the production of approximately 1 billion sperm each day from the completion of puberty to extreme old age, a total of more than 1 trillion sperm during the usual male reproductive life span.

Although the process of spermatogenesis is similar among species, there are major histologic differences, including differences between man and other primates. As illustrated in Figure 18-12, each spermatogonium undergoing differentiation after puberty gives rise to 16 primary spermatocytes, each of which then enters meiosis and gives rise to four spermatids and subsequently to spermatozoa. Thus, 64 spermatogonia can develop from each spermatogonium. In the steady state at least 1.5 million spermatogonia begin this cycle each day, and because nearly half of potential sperm production is lost during meiosis, the actual number of spermatogonia that commence meiosis may be closer to 3 million/day.\(^{142}\) The commitment of spermatogonia to differentiation does not occur randomly; instead, because clumps or groups of adjacent cells share a similar if not identical degree of histologic development, contiguous groups of spermatogonia undertake differentiation simultaneously.

During spermatogenesis, cytokinesis is frequently incomplete and cytoplasmic bridges form between differentiating spermatocytes and spermatids. This interconnection facilitates coordinated development of groups of germ cells, the so-called wave or synchrony of histologically distinct stages. Clermont\(^{143}\) identified six typical cellular associations in human seminiferous tubules; thus, one or two generations of spermatids at given steps of spermatogenesis are always associated with one or two generations of spermatocytes and with specific groups of spermatogonia. The succession of these six stages in any one area of tubular epithelium constitutes the cycle of the seminiferous epithelium.\(^{144}\)

The ultrastructural changes during spermatogenesis involve a reorganization of cytoplasm and development of the flagellum (Fig. 18-13). The chromatin becomes progressively more dense, and the nucleus comes to occupy an eccentric position at the cranial pole of the spermatid adjacent to

![Figure 18-12 Cell divisions during spermatogenesis. The overall number of cell divisions is much higher than that during oogenesis.](image)

![Figure 18-13 Schematic diagram illustrating conversion of spermatocyte to spermatid to spermatozoon.](image)

the acrosomal cap. The latter is probably formed from the Golgi apparatus and is believed to be essential for the penetration by the sperm of the zona pellucida of the ovum. The core of the sperm tail develops from a centriole near the Golgi apparatus and consists of nine outer fibers and two inner fibers. Mitochondria form a helix around the cilia from the neck to the annulus of the tail. The terminal region of the tail consists of the axial filament surrounded by the cell membrane; most of the cytoplasm is shed as spermatids are released into the lumen of the tubule.

Spermatogenesis takes approximately 70 days from the beginning of the differentiation of the spermatocyte to the formation of motile sperm.\(^{145}\) The transport of the sperm through the epididymis to the ejaculatory duct requires an additional 12 to 21 days\(^{146}\) and is a journey that involves peristaltic movement, bulk fluid drag, and intrinsic sperm motility. Sperm leaving the testes are relatively immature and have a poor capacity to fertilize. Sperm maturation during passage through the epididymis involves development of the capacity for sustained motility, modification of the structural state of the nuclear chromatin and the tail organelles, and loss of remnant cytoplasm (the cytoplasmic droplet).\(^{147}\) Acrosome acquisition by the sperm of the capacity to fertilize (capacitation) is poorly understood and may be completed in the female genital tract.

Energy to drive motility is derived from the hydrolysis of adenosine triphosphate generated in mitochondria in the middle piece of the tail (see Fig. 18-13). The axial structure of the tail contains a central pair of microtubules surrounded by nine doublet tubules and nine dense fibers; the doublets are attached to the central tubules by a series of radial spokes, to each other by dynein arms, and to the axonemal membrane by so-called Y links. Motility is believed to involve a sliding action of the microtubules, analogous to the interaction of actin and myosin in muscles. The dynein arms contain a powerful adenosine triphosphatase, and sliding is generated by interaction of the dynein arms and is restricted by the radial spokes.\(^{148}\) Mutations that influence the doublet arms, the spokes, or the spoke heads can lead to the immotile cilia syndromes (see later).\(^{149}\)

Control of Spermatogenesis

Spermatogenesis does not occur in the hypophysectomized state, and restoration of spermatogenesis after hypophysectomy and its initiation at puberty require LH and FSH. FSH acts directly on the spermatogenic tubule, whereas LH enhances spermatogenesis indirectly by increasing testosterone formation in Leydig cells.\(^{150}\) FSH and testosterone act in the testis by the same general mechanisms as peptide and steroid hormones in other tissues (see Chapter 4 and Chapter 5). For example, FSH binds to receptors on the surface of Sertoli cells and spermatogonia and stimulates adenylate cyclase, resulting in increased intracellular cAMP levels, activation of protein kinases, and phosphorylation of a variety of proteins.\(^{151}\)

FSH stimulates spermatogenesis at several levels.\(^{152}\) For example, it stimulates mitosis of Sertoli cells, increasing their number during puberty, and promotes maturation and development of tight junctions between Sertoli cells. FSH actions in Sertoli cells include increased production of androgen-binding protein, transferrin, inhibin, aromatase, and plasminogen activators and enhanced uptake of glucose and enhanced conversion of glucose to lactate. FSH receptor mRNA levels vary during different stages of the spermatogenic cycle, suggesting that the levels are regulated and may play a role in spermatogenesis. However, the fact that spermatogenesis and fertility can occur in the presence of loss-of-function mutations that impair FSH and its receptor suggests that the principal role of FSH in spermatogenesis is a quantitative one.\(^{153}\)

Androgen receptors are present in Sertoli cells, Leydig cells, and peritubular myoid cells, and androgen receptor levels in Sertoli cells vary with the stage of spermatogenesis and appear to be under the control of androgen.\(^{154}\) Genetic evidence has provided insight into the role of androgen in spermatogenesis; mouse
The hormonal requirements for the initiation of spermatogenesis in maturing animals differ from those for maintenance in adults or for reinitiation after hypophysectomy. After hypophysectomy in the adult male, spermatogenesis can be restored by treatment with FSH (human menopausal gonadotropin [hMG]) plus hCG. After spermatogenesis is restored, it can usually be maintained by hCG treatment alone. The latter phenomenon, together with the finding that in otherwise normal subjects with suppressed FSH activity spermatogenesis can be restored by LH alone, suggests that FSH is essential for initiation but not maintenance of spermatogenesis. However, FSH may be necessary for quantitatively normal sperm production in men, and a hypophysectomized man with an activating mutation of the FSH receptor had sustained spermatogenesis.

Fertilization

Fertilization normally takes place within the fallopian tube, and spermatozoa usually require a period in the female genital tract before they can fertilize. This functional change, termed capacitation, is believed to consist of at least two components: (1) enhancement of the rate of flagellar beat with acceleration of sperm movement and (2) development of the capacity to undergo an acrosome reaction and consequently allow the plasma membrane of the sperm to fuse with the ovum.

Whether capacitation is an absolute requirement in the human or serves only to enhance fertilizing capabilities is not known. Because fertilization can take place in vitro when sperm and eggs are combined with no preincubation, the minimal time required for some spermatozoa to undergo capacitation must be short.

Capacitation involves a change in the intracellular concentration or metabolism of calcium or cAMP. The acrosome reaction may also involve calcium. Neither the fallopian tube nor the egg itself appears to be essential for the acrosome reaction, which begins as a fusion between the acrosomal membrane and the overlying plasmalemma and is followed by calcium influx into the sperm. Subsequently, the acrosome fragments and disappears. The acrosome is derived from lysosomes, and its disintegration causes release of hydrolytic enzymes and proteases.

The fact that the acrosome reaction is followed within a few hours by a loss of sperm motility means that variability in the timing of capacitation in a sperm population relative to the moment of insemination increases the chance of successful fertilization. Ordinarily, about a fifth of motile spermatozoa recovered from the oviduct at variable times after insemination have undergone the reaction. The net effect of the enhanced motility and the acrosome reaction is that sperm acquire the capacity to penetrate the formidable vestments of the ovum.

One consequence of the sequential acceleration of motility and initiation of the acrosome reaction is that sperm transport to the site of fertilization in the fallopian tube is a culling process. Only a small number of the millions of sperm that are ejaculated reach the site of fertilization. The features that distinguish spermatozoa that reach the ampulla and fertilize the egg are not known, but these sperm are presumed to exhibit the fastest motility and the most delayed initiation of the acrosome reaction.

Understanding of the mechanism of sperm penetration is based largely on studies of fertilization of human eggs in vitro, a situation that may not be identical to the phenomenon in intact humans. Ovulated eggs are surrounded by layers of cumulus cells embedded in a matrix of hyaluronic acid. The mechanism by which sperm penetration through the cumulus is not known. Possibly, hyaluronidase is released by the degenerating acrosome, and the mechanical agitation of the flagellum may disperse the cumulus cells. Under in vitro conditions, prior disposal of the cumulus with hyaluronidase is necessary to allow penetration of the zona pellucida and hence to permit fertilization by the sperm.
Phases of Normal Testicular Function

The phases of normal testicular function can be delineated in terms of the plasma testosterone concentration (Fig. 18-14). In the male embryo, the production of testosterone by the testes begins to rise at the end of the second month of gestation and shortly thereafter reaches a maximal value that is maintained until late in gestation and then decreases. At the time of birth, the plasma testosterone level is only slightly higher in males than in females. Shortly afterward, the plasma testosterone level again commences to rise in the male infant and remains elevated for approximately 3 months, falling to low levels by 1 year. The plasma level then remains low (but higher in boys than in girls) until the onset of puberty, when it again increases in boys and reaches adult levels by about age 17.

Plasma levels remain more or less constant in the adult until middle age and then gradually decline during the later decades of life. Sperm production takes place after puberty. The physiologic events during these various periods differ, as do the consequences of testicular derangements at different stages of life.

Embryonic Male Sexual Differentiation

The process of sexual differentiation is described in Chapter 22. In brief, the embryos of both sexes develop in an identical fashion until the seventh week of gestation. Thereafter, the anatomic development and the physiologic development diverge, with formation of the male or female phenotypes.

As formulated by Jost, normal sexual development in the mammalian embryo depends on three sequential processes. The first involves the establishment of genetic sex, which is defined by the sex chromosome constitution established at the time of conception. The heterogametic sex (XY) in mammals is male, whereas the homogametic sex (XX) is female. In the second phase, the sex chromosomal determine whether the indifferent gonad differentiates into a testis in the male or an ovary in the female. The third step involves the translation of gonadal sex into phenotypic sex and is the direct consequence of the type of gonad formed; that is, testicular secretions determine that the urogenital tract and external genitalia will be male in character.

The internal genitalia in the two sexes are derived from the wolffian and müllerian ducts that exist side by side in early embryos of both sexes. The wolffian ducts serve as the excretory ducts of the mesonephric kidney and are physically attached to the indifferent gonad, whereas the müllerian duct has no continuity with the gonad. In the male, the wolffian ducts give rise to the epididymis, vasa deferentia, seminal vesicles, and ejaculatory ducts and the müllerian ducts disappear. In the female, the fallopian tubes, uterus, and upper vagina are derived from the müllerian ducts and the wolffian ducts disappear.

Thus, masculinization of the fetus requires the action of testicular hormones, whereas the female phenotype develops in the absence of gonadal secretions. Under ordinary circumstances, chromosomal sex, gonadal sex, and phenotypic sex are concordant; that is, chromosomal sex determines gonadal sex and gonadal sex in turn determines phenotypic sex, without deviation from the chromosomal program.

Control over the formation of the male phenotype is vested in the action of three hormones. Two of the three, AMH and testosterone, are secretory products of the fetal testis. AMH, a glycoprotein hormone of the embryonic testis, acts ipsilaterally in the male embryo to suppress the müllerian ducts and consequently prevents development of the uterus and fallopian tubes. Testosterone converts the wolffian ducts into the epididymides, vasa deferentia, and seminal vesicles and is also the precursor for the third fetal hormone, dihydrotestosterone. The latter hormone, which is formed within the urogenital sinus and lower urogenital tract from circulating testosterone, acts in the urogenital sinus to induce formation of the male urethra and prostate and in the genital tubercle, swelling, and folds to cause the midline fusion, elongation, and enlargement that eventuate in the male external genitalia.

Thus, androgens function during fetal life to induce the formation of the accessory organs of male reproduction. Testosterone and dihydrotestosterone act through the same receptor mechanism during embryogenesis and in the adult (see Fig. 18-14). The formation of the male phenotype is largely completed by 12 to 15 weeks of gestation, but at the time of completion of the male urethra, the external genitalia in the two sexes do not differ in size. Descent of the testes and differential growth of the external genitalia in the male take place largely during the second half of gestation.

The control of testosterone formation by the embryonic testis is incompletely understood. By the 13th week of human gestation, testosterone secretion appears to be regulated by LH from the fetal pituitary gland or by placental hCG in the fetal circulation, or by both. The decrease in testosterone synthesis late in gestation correlates both with a decline in the number of LH-hCG receptors in the testis and with a decrease in the level of hCG and LH in the fetal circulation. Castration of the male rhesus monkey during late gestation results in a further decrease in plasma testosterone elevation of plasma gonadotropins. Anencephaly and other forms of congenital hypopituitarism cause the syndrome of microphallus. Taken together, these findings indicate that testosterone production during the second half of gestation is regulated by LH or hCG and that LH production itself is under negative-feedback control by testosterone.

The mechanism by which testosterone production is controlled between gestational weeks 8 and 12, when male phenotypic development takes place in the human embryo, is not clear. In the rabbit embryo, testosterone production during the analogous phase of male development appears to be independent of gonadotropin, but for technical reasons this phase of embryonic development in human gestation has not been adequately examined. The fact that most male infants with anencephaly, congenital hypopituitarism, or both have normal male urothras suggests either that androgen synthesis during early gestation is independent of gonadotropins or that chorionic gonadotropin, which apparently is not present in the rabbit, acts as a fail-safe mechanism to guarantee normal male development in the absence of LH from the fetal pituitary gland. However, the fact that loss-of-function mutations of the LH receptor gene cause impairment of the virilization of the male external genitalia (but not of the wolffian ducts) indicates that LH or hCG plays a role in early androgen synthesis.

In addition to their role in male phenotypic development, androgens secreted during fetal or neonatal life (or both) exert at least two types of effects on the CNS in some species: regulation of the hypothalamic-pituitary system and control of diverse sexually dimorphic behavior patterns. Androgens are presumed to act in brain through the same receptor as in the urogenital tract and other androgen target tissues. Social imprinting also plays a critical role in sex-specific behavior in some species. Male sexual development, apart from spermatogenesis, is remarkably complete during embryogenesis. For example, male infants have periodic erections during the later phases of gestation, which indicates that the complex neurogenic pathways that regulate this process have developed by that time.

The extent to which androgen action in the human CNS influences human sexual behavior has not been established. There is no evidence for permanent imprinting by fetal androgens on the hypothalamic control of gonadotropin production in the human, and it is not established whether gonadal hormones have any direct effect on gender identity or gender behavior apart from their role in anatomic development of the sexual phenotype. Nevertheless, both androgens and cultural factors probably play important roles in the development of characteristic male behavior. Therefore, in making clinical decisions about sex assignment in subjects with ambiguous genitalia, it is important to undertake a thorough diagnostic evaluation and appropriate therapeutic intervention as early as possible, preferably in the newborn nursery, to ensure that the
The neonatal surge in testosterone secretion is the consequence of a rise in plasma gonadotropin levels, but neither the cause of the increase in gonadotropin levels nor the precise function of the temporary increase in testosterone secretion is understood. Serum inhibin levels increase in boys in the first year of life in parallel with testosterone levels. In some species, the neonatal testosterone surge is believed to be responsible for two aspects of male development: (1) permanent virilization of the hypothalamus so that it secretes LH tonically rather than cyclically as in the female and (2) the priming of androgen target tissues for subsequent androgen-mediated growth and maturation in later life.

Blockade of the neonatal activation of the pituitary-testicular axis in male monkeys resulted in subnormal increases in LH and testosterone and impairment of testicular enlargement at the time of puberty. However, in humans there is no evidence that neonatal deprivation or excess of androgen has any permanent effect on hypothalamic-pituitary function. Whether neonatal androgen plays a specific role in the male gender identity or gender role behavior is likewise uncertain. Indirect evidence suggests that neonatal androgen influences the subsequent androgen-mediated growth of the male urogenital tract. Boys who are born with micropenis caused by deficient androgen biosynthesis may have subnormal androgen-mediated growth of the external genitalia if androgen replacement therapy is not started until the time of normal male puberty. However, their response may be normal if androgen is administered temporarily during infancy. Such observations are consistent with the view that late fetal or neonatal androgen primes the male urogenital tract by promoting early growth and potentiating maturational effects of the hormone at puberty.

In the prepubertal years, the plasma levels of gonadotropins and gonadal steroids are low. Maturation of adrenal androgen secretion, termed adrenarche, results in enhanced secretion of dehydroepiandrosterone (DHEA), dehydroepiandrosterone sulfate (DHEAS), and androstenedione in boys as early as age 6 or 7, several years before maturation of the hypothalamic-pituitary-gonadal axis. The secretion of these androgens is probably under the control of corticotropin (also called adrenocorticotropic hormone [ACTH] or adrenocorticotropic) and appears to be independent of activation of the pituitary-gonadal axis. In part, the prepubertal growth spurt and the early development of axillary and pubic hair are mediated by these adrenal androgens, which are believed to bind to the androgen receptor only after conversion to testosterone or dihydrotestosterone in target tissues.

Before the onset of puberty, the low levels of plasma gonadotropins are under feedback control by the small amounts of androgen secreted by the testes, as evidenced by the fact that castration at this time results in a rise in plasma gonadotropins to levels similar to those of the postpubertal castrate. Gonadotropins in children, as in adults, are secreted in a pulsatile fashion, with the pulses occurring at 2- to 3-hour intervals. These facts suggest that the negative-feedback control of gonadotropin secretion is exquisitely sensitive to plasma testosterone levels before puberty and subsequently changes during puberty development.

The factors that determine the onset of puberty are poorly understood and may reside in the hypothalamic-pituitary system, in the testis itself, in the adrenal gland, or at some undefined level (see Chapter 24). The sequence of pubertal maturation has, however, been well characterized. Its onset is heralded by sleep-associated pulses of LH secretion and, to a lesser extent, by increases in the episodic secretion of FSH. Later in puberty, the increased plasma gonadotropin levels become sustained throughout the day, as do the resulting increases in plasma testosterone and dihydrotestosterone levels. The rise in gonadotropin secretion is believed to be the consequence of both an increase in GnRH secretion and an increase in sensitivity of the pituitary to GnRH. Plasma levels of bioactive LH increase even more than those of the immunoreactive hormone.

The pubertal changes in gonadotropin and steroid hormone levels in plasma are compatible with the concept that with maturation, the hypothalamic-pituitary system becomes less sensitive to feedback inhibition by circulating androgens, which results in a higher mean plasma androgen concentration. The maturational change in the hypothalamic-pituitary system appears to be triggered by the attainment of a critical body mass or percent body fat, possibly mediated by an increase in plasma leptin levels.

The pubertal changes in the testes are illustrated in Figure 18-16. In prepubertal testes, the interstitial cells consist of an undifferentiated mesenchyme with immature tubules. After puberty, the cytoplasm of the functioning Leydig cells develops a characteristic foamy appearance and the various stages of spermatogenesis can be delineated within the tubule. Initiation of spermatogenesis early in puberty is associated with a rise in serum inhibin B levels. As indicated by the appearance of sperm in centrifuged urine samples from boys entering puberty, sperm production is an early pubertal event and can occur when testicular growth has just begun.

The anatomic and functional changes of puberty are largely the consequence of the action of testicular androgens. It is probable that all androgenic actions are mediated by dihydrotestosterone or testosterone and that other naturally occurring 19-carbon steroids act as androgens only if they are converted to testosterone or dihydrotestosterone within extraglandular tissues. The physiologic effects of androgens have been classified as either androgenic (maturational of the male urogenital tract and spermatogenesis) or anabolic (promotion of growth in muscle and other somatic tissues). These various effects are mediated by the same androgen receptor.
Androgen is responsible for many effects at puberty. Rugal folds appear in scrotal skin. The testes, penis, and scrotum enlarge, and the penis and scrotum become pigmented. The prostate, seminal vesicles, and epididymis increase in size. Growth of the various accessory organs of reproduction accounts for about a fourth of androgen-mediated nitrogen retention in puberty. One consequence of this growth and maturation process is the transformation of the cuboidal epithelia of the secretory tissues of the urogenital tract into columnar epithelia. The characteristic hair growth of male puberty involves development of the mustache and beard; regression of the scalp line; appearance of truncal, extremity.

<table>
<thead>
<tr>
<th>Pubic Hair Stage</th>
<th>Genital Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 1: Prepubertal. Vellus over the pubes is no further developed than that over the abdominal wall, i.e., no pubic hair.</td>
<td>Stage 1: Preadolescent. Testes, scrotum, and penis are about the same size and proportion as those in early childhood.</td>
</tr>
<tr>
<td>Stage 2: There is sparse growth of long, slightly pigmented, downy hair, straight pigmented, downy hair, straight or only slightly curled, appearing chiefly at base of penis.</td>
<td>Stage 2: Scrotum and testes have enlarged, and there is a change in the texture of scrotal skin and some reddening of scrotal skin.</td>
</tr>
<tr>
<td>Stage 3: Hair is considerably darker, coarser, and more curled and spreads sparsely over junction of pubes.</td>
<td>Stage 3: Growth of the penis has occurred, at first mainly in length but with some increase in breadth. There has been further growth of the testes and the scrotum.</td>
</tr>
<tr>
<td>Stage 4: Hair is now adult in type, but the area covered by it is smaller than that in most adults. There is no spread to the medial surface of the thighs.</td>
<td>Stage 4: The penis is further enlarged in length and breadth, with development of glans. The testes and the scrotum are further enlarged. There is also further darkening of scrotal skin.</td>
</tr>
<tr>
<td>Stage 5: Hair is adult in quantity and type, distributed as an inverse triangle. There is spread to the medial surface of the thighs but not up the linea alba or elsewhere above the base of the inverse triangle.</td>
<td>Stage 5: Genitalia are adult in size and shape. No further enlargement takes place after stage 5 is reached.</td>
</tr>
</tbody>
</table>


Adulthood

Adulthood.

TABLE 18-2 — Stages of Puberty

TABLE 18-2

- Pubic Hair Stage
- Genital Stage

Stage 1: Prepubertal. Vellus over the pubes is no further developed than that over the abdominal wall, i.e., no pubic hair.

Stage 2: Scrotum and testes have enlarged, and there is a change in the texture of scrotal skin and some reddening of scrotal skin.

Stage 3: Growth of the penis has occurred, at first mainly in length but with some increase in breadth. There has been further growth of the testes and the scrotum.

Stage 4: The penis is further enlarged in length and breadth, with development of glans. The testes and the scrotum are further enlarged. There is also further darkening of scrotal skin.

Stage 5: Hair is adult in quantity and type, distributed as an inverse triangle. There is spread to the medial surface of the thighs but not up the linea alba or elsewhere above the base of the inverse triangle.

Adulthood.

On average, reproductive capacity is attained between the ages of 16 and 19 years. As indicated in Figure 18-17, most anatomic changes are also completed by this time. However, androgen-mediated growth of body hair is usually not maximal until the middle to late 20s.

The various physiologic actions of androgen during puberty and adulthood can be separated into two general types: permanent and concurrent. Permanent effects encompass anatomic actions that are irreversible and do not regress if androgen production ceases, such as the effects on the larynx. Concurrent effects are those that require a continuing male level of the hormones, such as the enhancement of erythropoietin production and hemoglobin levels. Other physiologic effects of androgen have both permanent and concurrent components; for example, beard growth slows but rarely stops in men who are castrated postpubertally.

Many features of castration have been described in anecdotal

form only, but several aspects have been studied in some detail.

First, postpubertal castration results in a negative nitrogen balance. The source and exact magnitude of the nitrogen loss have not been established, but probable
sites of loss include the secretory tissues of the male urogenital tract and, to some extent, other androgen target tissues such as muscle. Androgen replacement in castrated men restores nitrogen balance and the secretory capacity of the epididymis, seminal vesicles, and prostate.

Second, castration is followed by a progressive decline in male sexual drive so that only rare castrated subjects can have intercourse after a few years. In such individuals, physiologic androgen replacement results in a rapid and predictable restoration of male sexual activity.

Third, two thirds of men subjected to surgical or medical castration experience hot flushes that may persist for 5 years or longer. Long-term consequences of castration in men include osteoporosis, gynecomastia, profound shrinkage (apparent disappearance) of the prostate, and enlargement of the pituitary gland (assumed to be due to hyperplasia of LH-secreting cells).

At the completion of puberty, the plasma testosterone levels have attained the adult male level of 10 to 35 nmol/L (3 to 10 ng/mL), sperm production has reached a steady level, and plasma concentrations of gonadotropins are in the adult range. Thus, the mature feedback regulatory system shown in Figure 18-5 (Figure Not Available) is established and is maintained for approximately 40 years.

Even under the best circumstances, the system can be perturbed, usually temporarily, by a variety of influences at the level of both the testis and the hypothalamic-pituitary system. One of the most important of these influences is scrotal temperature; spermatogenesis is exquisitely sensitive to alterations in temperature, and temporary increases in systemic or local temperature (as in a hot bath) can be followed by temporary decreases in sperm production. Spermatogenesis can also be influenced by diet, drugs, environmental agents, and a variety of psychological stresses. Testosterone production is more stable than spermatogenesis but can also be impeded by drugs (see later).

Old Age

The term male climacteric implies an analogy to the complete cessation of ovarian function in women at the time of menopause. Men do not experience a relatively rapid total cessation of Leydig cell or seminiferous tubule function with old age, nor do men with intact testes experience the hot flashes characteristic of menopause. Male sexual function does decrease with age, but this decline does not appear to coincide with hormonal changes. Nevertheless, in older men carefully screened to exclude major health problems or medication use, serum testosterone levels do decline with age, although most older men still have serum testosterone levels within the range of normal for young men.

The normal circadian rhythm in serum testosterone levels is lost with age, and levels of bioavailable testosterone, namely serum testosterone not bound to SHBG, and Leydig cell reserve, as assessed by response to hCG, are decreased in older men. In a cross-sectional analysis of the Massachusetts Male Aging Study, testosterone parameters began to change around age 40 years; free testosterone levels decreased about 1.2%/year and SHBG levels increased about 1.2%/year, so the total testosterone levels do not reflect the true level of bioavailable testosterone in older men. A longitudinal follow-up of these subjects also showed a fall in total and free testosterone levels of about 1.5%/year. In another study, serum levels of bioavailable estradiol and testosterone both declined with age so that most men older than 65 had low levels of bioavailable testosterone as compared with young men.

When controlled for ejaculatory frequency, sperm density does not change with age, but older men have lower ejaculate volumes, decreased sperm motility, and an increased percentage of abnormal sperm. Total daily sperm production, as assessed histologically, also declines with age. However, most older men have values for these parameters that are within the normal range for young fertile men, and decreased sperm production in older men does not correlate with Leydig cell number. Levels of serum inhibin B decrease after age 35.

Serum LH and FSH levels increase in elderly men in keeping with some decrease in bioavailable testosterone and sperm production. Studies of gonadotropin secretion indicate a decreased LH pulse frequency but no change in pulse amplitude with age. Levels of bioactive and immunoreactive LH are increased to a similar degree in elderly men. These observations suggest that the hypothalamic-pituitary responsiveness to the decreased Leydig cell function is appropriate. In one study, the bioactive pituitary LH reserve was decreased in elderly men, as indicated by the response to GnRH and tamoxifen; in another study, the response to clomiphene citrate was normal. Basal levels of bioactive FSH are similar in young and elderly men, but total levels of FSH are higher in elderly men. Administration of clomiphene citrate increases bioactive and immunoreactive FSH levels to a similar degree.

The modest increase in serum gonadotropin levels with age is less than would be predicted for the decline in bioavailable testosterone levels, possibly a consequence of enhanced negative feedback of androgens. Veldhuis and colleagues have suggested that many of the endocrine changes with age are due to impaired synchrony of neuroendocrine function. The close coupling between sleep stage and nocturnal penile tumescence and episodic LH and testosterone secretion becomes blurred in the elderly. Pulsatile administration of GnRH to older men returns LH pulse frequency and amplitude and plasma LH levels to normal, but plasma testosterone levels do not increase equivalently. These findings suggest that defects may occur in the Leydig cell and in the CNS with aging.

The role of these changes in other aspects of aging is yet to be determined. In healthy men aged 73 to 94 years, levels of total and bioavailable testosterone correlate positively with muscle strength and total body bone mineral density (BMD) and negatively with fat mass. Levels of plasma estrone and estradiol correlate positively with BMD, and the relation between testosterone levels and BMD is independent of estradiol levels. Estrogen appears to play a major role in preventing bone resorption, whereas both estrogen and testosterone regulate bone formation. In other studies, the decline in bioavailable testosterone in elderly men correlated with depressed mood and with changes in cognitive function. The possible relation between testosterone levels and body composition provides some rationale for testosterone replacement in healthy older men with low testosterone levels (see later).
ASSESSMENT OF TESTICULAR FUNCTION

Leydig Cell Function

History and Physical Examination

The assessment of androgen status should include an inquiry about the following:

1. Presence of developmental abnormalities at birth (e.g., hypospadias, microphallus, cryptorchidism) (see later).
2. Timing and extent of sexual maturation at puberty.
3. Rate of beard growth.
4. Current libido, sexual function, muscle strength, and energy.

Inadequate androgen production or action during embryogenesis impairs development of the normal male phenotype, and Leydig cell failure before puberty impairs sexual maturation and causes eunuchoidism, namely an infantile amount and distribution of body hair, poor development of skeletal muscles, and failure of closure of the epiphyses so that the arm span is more than 5 cm greater than height and the lower body segment (heel to pubis) is more than 5 cm longer than the upper body segment (pubis to crown).

Detection of Leydig cell failure that begins after puberty requires a high index of suspicion and appropriate laboratory assessment. One reason for difficulty in detecting this condition is that decreased sexual function in adult men is more common than Leydig cell failure. Erectile dysfunction with preservation of normal libido is usually not due to testosterone deficiency but rather to the adverse effects of systemic disease or of medications prescribed to treat such disease. Likewise, decreased libido may not be recognized by individual patients and may be revealed only by discussion with sexual partners. A second reason is that some manifestations of testosterone deficiency are nonspecific, including depression, loss of drive in the workplace, decreased stamina, increased irritability, hostility, and nervousness.

Another reason that androgen deficiency is missed is that some functions that require androgens for initiation continue unabated with Leydig cell failure and some functions that eventually regress do so slowly. The frequency of shaving may not decrease for many months or years because of slow decline in the rate of beard growth once established. Finally, some consequences of testosterone deficiency such as increased abdominal obesity and decreased muscle mass may not be recognized as out of the range of normal variation.

Assessment of Plasma Luteinizing Hormone, Androgens, and Sex Hormone Binding Globulin

Because the LH level must be interpreted in light of plasma testosterone, it is usually appropriate to measure both hormones by using a pool formed by combining equal quantities of blood obtained from three or four samples at 15- to 20-minute intervals. In this way, only a single pooled sample of plasma is submitted to the laboratory and the averaging of values is accomplished before the assay. Dual-site immunometric assays using an appropriate international standard or reference preparation have largely replaced competitive radioimmunoassays. The usual normal range of plasma LH in adult men is 1.3 to 13 IU/L.

Plasma testosterone is also measured by immunoassay. Like LH, testosterone is secreted in a pulsatile fashion. Because the diurnal variation of plasma testosterone is significant in many men, it is desirable to measure testosterone in the morning along with plasma LH in pooled samples as just described. The normal range in adult men is 10 to 35 nmol/L (3 to 10 ng/mL). The plasma testosterone level is higher in prepubertal boys than girls, the normal range in boys being 0.2 to 0.7 nmol/L (0.05 to 0.2 ng/mL). The start of puberty is marked by a rise in plasma testosterone at night as a consequence of sleep-related nocturnal gonadotropic surges, and daytime levels of plasma testosterone gradually increase with the stages of puberty. In young men, the plasma testosterone is higher in the morning than in the late afternoon. In healthy middle-aged and elderly men, morning testosterone levels tend to remain stable over long periods.

Plasma dihydrotestosterone is also measured by immunoassay. In normal young men, the plasma concentration averages about 10% of the testosterone value.

Testicular function cannot be assessed by measurement of urinary 17-ketosteroids, which are composed of metabolites of testicular and adrenal androgens.

Measurement of the SHBG level is sometimes useful for interpretation of levels of total plasma testosterone. Binding capacity of SHBG can be assessed by the use of radioactive androgen, and the protein is measured by immunoassay.

In most situations, measurement of total testosterone provides an adequate assessment of Leydig cell function. However, in men with altered levels of SHBG (men older than 55 years of age and men with human immunodeficiency virus [HIV], abnormal liver function, or marked obesity), it is appropriate to assess the free or bioavailable testosterone level. Measurement of free testosterone levels by equilibrium dialysis is rarely done, and the analogue free testosterone assay is unreliable because it is influenced by the SHBG level. It is more practical to measure bioavailable (non-SHBG-bound) testosterone; this assay is performed by precipitating SHBG and measuring testosterone in the supernatant.

Alternatively, free or bioavailable testosterone can be estimated from the levels of total testosterone and SHBG, using the affinity constants of binding of testosterone to albumin and SHBG. In most circumstances, these calculated values correspond closely to directly measured levels.

Dynamic Tests of the Hypothalamic-Pituitary-Leydig Cell Axis

To assess Leydig cell function before puberty, it is common to measure the response of plasma testosterone to gonadotropin stimulation as an index of Leydig cell reserve. Normal prepubertal boys respond to 3 to 5 days of injection of hCG at 1000 to 2000 IU/day with an increase of plasma testosterone to about 7 nmol/L (2 ng/mL); the magnitude of the response increases with the initiation of puberty and peaks in early puberty.

In some circumstances, the response of plasma LH to GnRH is measured to assess the functional integrity of the hypothalamic-pituitary-Leydig cell axis. Before puberty, the responses of LH and FSH are similar, and with puberty the LH response to acute administration of GnRH increases whereas the FSH response remains the same. The amount of LH released after acute administration of GnRH is believed to reflect the amount of hormone stored in the pituitary. Administration of 100 µg of GnRH to normal men causes LH levels to increase fourfold to fivefold with a peak level at 30 minutes. However, the range of response is broad, and the peak LH after a single dose of GnRH usually correlates with the basal level.

In men with primary testicular failure, measurement of basal LH is usually sufficient and assessment of the GnRH response adds little information. Men with pituitary or hypothalamic disease may have either a normal or an abnormal LH response to an acute dose of GnRH, and consequently a normal response is of no value in determining the presence of secondary disease or in distinguishing hypothalamic from pituitary disease. A subnormal response to an acute dose indicates that an abnormality exists but provides no evidence concerning the site of the abnormality.

Measurement of long-term GnRH responsiveness is useful for evaluation of men with secondary hypogonadism and a subnormal LH response to an acute dose of GnRH. If daily...
infusions of GnRH for 1 week lead to the development of a normal LH response to an acute dose of GnRH, a hypothalamic cause of the hypogonadism is likely. More often, a hypothalamic or pituitary defect is diagnosed with imaging studies.
Seminal Tube Function

History and Physical Examination

Leydig cell dysfunction usually results in defective spermatogenesis, and men with the clinical features of Leydig cell dysfunction are usually infertile. In contrast, men with primary disorders of the seminiferous tubules can present with infertility as the sole clinical manifestation.

Examination of the testes is an essential portion of the physical examination. The seminiferous tubules account for about 60% of testicular volume. The prepubertal testis is about 2 cm in length (and 2 mL in volume, as assessed by the Prader orchidometer) and increases in size with puberty to reach the adult size by about age 16. When damage to the seminiferous tubules occurs before puberty the testes are small and firm, whereas postpubertal damage characteristically results in small, soft testes.

The normal adult testis is at least 4 cm in length with a volume of 15 to 25 mL. Considerable damage can occur before overall size shrinks below the lower limits of normal. Testis size varies among ethnic groups. Asian men have smaller testes than European men independent of differences in overall body size. Testicular size can be assessed with a ruler or with a Prader orchidometer. Because of the frequent occurrence of varicocele among fertile men and its possible causal role in infertility, the testes should be carefully palpated while the patient is standing.

Seminal Fluid Examination

Routine evaluation of the seminal fluid assesses parameters that do not necessarily reflect the functional capacity of the sperm. Seminal fluid should be obtained by masturbation into a clean glass or plastic container. Collection in a condom or after coitus interruptus may be incomplete and is not recommended. The volume of the normal ejaculate is 2 to 6 mL. The seminal fluid coagulates immediately after ejaculation and then liquefies within 15 to 30 minutes. The specimen should be analyzed within an hour.

Motility is estimated by examining a drop of undiluted seminal fluid and recording the percentage of motile forms. The quality of motility can be graded 1 to 3. Spermatozoa with grade 3 motility move rapidly across the field, grade 2 spermatozoa move aimlessly, and grade 1 spermatozoa have a beating tail but do not move. Normally, 60% or more of sperm should be motile, with an average quality of motility of grade 2.5 or more.

After it liquefies, the semen sample should be mixed well and diluted 1:20 either in water or with dilute sodium bicarbonate containing 1% phenol to immobilize the spermatozoa. A drop of this specimen is placed on a standard blood-counting chamber, and the spermatozoa are counted within five blocks containing 16 squares each. This number multiplied by 10^6 represents the count per milliliter. Sperm density can also be estimated by using an electronic particle counter. The normal value is usually considered to be greater than 20 million/mL, with total sperm per ejaculate greater than 60 million.

Random sperm counts are complicated by the effects of factors such as hot baths, acute febrile illness, and medications. The net result is that it is difficult to define the minimally adequate ejaculate. When 24 to 36 hours of sexual rest is specified and ejaculates are examined at 2-week intervals, average semen quality and sperm output are lower than previously considered normal for fertile men. Three ejaculates are usually required to determine sperm number and cytologic features, and six estimates or more may be necessary for valid assessment if the initial ejaculates are of equivocal quality.

Seminal fluid cytology can provide a useful index of fertility. The seminal fluid smear is prepared similarly to a blood smear but with special stains. Normal spermatozoa have symmetrically oval heads, middle pieces that are larger at the proximal ends and inserted symmetrically into the heads, and tails 7 to 15 times longer than the heads. Some abnormal spermatozoa are present in all semen. The best correlations between histologic abnormalities and infertility occur when a single anomaly (e.g., lack of the acrosome) is found in a large percentage of sperm. More than one abnormality may be present. Although there is no clear definition of the minimal structural features compatible with fertility, 60% or more of the spermatozoa should have normal morphology. Evaluation of sperm structure by electron microscopy, if available, is useful for identifying specific defects in immotile sperm (see later).

Testicular Biopsy

In men with azoospermia, testicular biopsy may be helpful if intracytoplasmic sperm injection (ICSI) is contemplated for treatment of infertility. It is most appropriate to perform such biopsies at the time ICSI is scheduled. Fine-needle aspiration can provide biopsy material and sperm for ICSI. In most men with infertility associated with oligospermia, testicular biopsy is of little value. The diagnosis of Klinefelter's syndrome related to chromosomal mosaicism limited to the testes can be established only by tissue culture and karyotypic analysis of the biopsy material. The histologic features of several testicular disorders are illustrated in Figure 18-16.

Plasma Follicle-Stimulating Hormone and Inhibin B

Levels of plasma FSH usually correlate inversely with spermatogenesis; that is, elevations of FSH occur in men with intact hypothalamic-pituitary axes when there is severe damage to the germinal epithelium. An inverse relationship also exists between levels of plasma inhibin B and FSH in men, including semen donors, infertile men, and men with elevated FSH levels. FSH is measured by immunoradiometric assay using the appropriate international standard or reference preparation. As with LH assays, dual-site immunometric assays are now commonly used and the usual range in normal men is 0.9 to 15 IU/L. Oligospermia caused by primary testicular defects is usually associated with elevated FSH levels.

Chromosomal Analysis

Examination of buccal mucosa cells for the presence of chromatin clumps on the nuclear membrane (the Barr body) provides evidence for the number of X chromosomes. In general, there is one Barr body for every X chromosome in excess of one. The Barr body, which represents the second X chromosome in XX individuals, is identifiable in 20% or more of the nuclei of cells in normal females and in less than 2% of cells of normal males. If buccal mucosa cells are stained with quinacrine or its mustard derivative and examined by fluorescence microscopy, the Y chromosome can be identified. This method provides a rapid and accurate means of determining the sex chromosome complement under some circumstances such as suspected male pseudohernphroditism.

Analysis of the chromosomal karyotype, the most accurate means of determining the chromosome complement, involves culture of peripheral blood leukocytes or of tissue fibroblasts in medium containing an agent such as phytohemagglutinin that induces the cells to divide. A spindle poison such as colchicine, which arrests mitosis at metaphase, is added; the cells are harvested and stained; and the number and histologic characteristics of the chromosomes are assessed in several cells. This technique is valuable for establishing the exact chromosome complement, the presence of mosaicism, the presence of structural chromosome alterations, and the sex chromosome composition. The study of multiple tissues may be necessary to establish chromosome mosaicism. In a given tissue, 20 cells must be examined to exclude with 95% confidence a mosaicism of 15% or greater.
Estrogenic Function

History and Physical Examination

Gynecomastia (enlargement of the male breast), the most consistent feature of feminizing states in men, is the consequence of proliferation of glandular tissue. The physician should seek the presence of gynecomastia by examining the patient while he is in the sitting position; the fingers are used to grasp the glandular tissue. Palpation with the flat part of the hand while the patient is supine may result in failure to detect early or minimal breast enlargement. In obese men, it is important to try to detect the edge of the rim of glandular tissue that separates it from the adipose tissue of the chest wall. Ultrasonography or mammography may be useful in separating true gynecomastia from lipomastia.

Plasma Estrogens

As discussed earlier, most estradiol and estrone in normal men are formed by extraglandular aromatization of circulating androgens. As assessed by immunoassay, plasma estradiol is usually less than 180 pmol/L (50 pg/mL) in normal men and plasma estrone is somewhat higher but usually less than 300 pmol/L (80 pg/mL). A recombinant cell bioassay has been developed to measure estradiol at low levels.

[244]
ABNORMALITIES OF ANDROGEN METABOLISM AND TESTICULAR FUNCTION

Abnormalities of testicular function have different consequences, depending on the phase of sexual life in which they are first manifested. Although there are problems inherent in all classifications and although some assignments are arbitrary, such a categorization of testicular diseases has a sound physiologic rationale. For example, although Klinefelter's syndrome is a disorder of chromosomal sex, it is usually diagnosed in individuals when manifestations become apparent after the time of expected puberty. Although such limitations must be recognized, disorders of the testes can be classified as abnormalities of fetal development, puberty, adult life, and senescence.

Fetal Life

Abnormalities of Male Sexual Differentiation

Disturbances in sexual differentiation can arise from a variety of mechanisms:

1. Environmental insult, as in the ingestion of a virilizing drug during pregnancy.
2. Nonfamilial aberrations of the sex chromosomes, as in 45,X/46,XY chromosomal mosaicism.
3. Developmental birth defects of multifactorial origin, as in most cases of hypospadias.
4. Hereditary disorders resulting from single-gene mutations, as in the testicular feminization syndrome.

The disorders of sexual differentiation and their management are described in Chapter 22, but because 46,XX men and men with Klinefelter's syndrome ordinarily present with problems of undervirilization or infertility, they are also discussed in this chapter.

Cryptorchidism

Descent of the testes is essential to normal function because spermatogenesis requires the lower temperature that is present in the scrotum. Failure can occur at any site in the normal pathway of descent, from high in the abdomen to the scrotum itself. The implications and sequelae of cryptorchidism differ, depending on the site at which descent ceases.

A large portion of the literature in this field is difficult to interpret because of imprecise definitions. Cryptorchidism can be defined as a testis that is not 4 cm or more below the public tubercle in an infant of normal size and subclassified, depending on the location of the maldescended testis, as follows:

1. The intra-abdominal testis (10%) cannot be felt. It is usually located just above the internal inguinal ring. Infants with bilateral intra-abdominal testes can be distinguished from female pseudhermaphrodites by assessment of the chromosomal karyotype and from boys with bilateral anorchia by demonstrating that the plasma testosterone level increases after administration of HCG. Unilateral intra-abdominal testes must be separated from the syndrome of mixed gonadal dysgenesis, in which a testis is present on one side and an intra-abdominal streak gonad is present on the other.
2. The canalicular testis (20%) has traversed the internal inguinal ring and is present in the inguinal canal; it may move intermittently between the canal and the upper scrotum. Such testes are small or would not be able to pass the external inguinal ring. When the testis is in the canal, the aponoeurosis of the external oblique muscle forms a firm barrier that the testis can rarely be palpated.
3. The high scrotal testis (40%) is farther along the pathway of descent but does not reach the bottom of the scrotum. It is characteristically smaller than its normal partner and has a limited range of motion so that it can retract into the groin but not past the internal ring. Retraction may make accurate diagnosis and classification difficult.
4. The obstructed testis (30%) is a fourth category in which failure of descent appears to be due to a physical barrier formed by a cord of fascia between the inguinal pouch and the inlet of the scrotum.

Another category is the ectopic testis. On rare occasions, the testis may deviate from its normal pathway of descent and become ectopic in location. The five most frequent sites of ectopia are the perineum, the femoral canal, the superficial inguinal pouch, the suprapubic area, and the opposite scrotum. Testicular ectopia is believed to be caused by an abnormality of the gubernaculum. In most situations, the higher the location of the testis or the more extreme the ectopia, the more difficult surgical repair becomes.

About 3% of full-term male infants have at least one cryptorchid testis at birth. Completion of descent usually occurs during the first few weeks after birth so that the incidence of cryptorchidism at 6 to 9 months and in adult men is about

0.7% to 0.8%. Accurate diagnosis and classification may require careful, repeated observations by a single observer to be certain that a normally (or partially) descended testis has not retracted into the groin. The concept that spontaneous descent can occur after a few months of age is a misconception that arose because of the failure to recognize that many normal testes are retractile in young boys and that elicitation of the cremasteric reflex can cause at least partial retraction of fully descended testes in three fourths of boys. The incidence of retraction declines with age, and it rarely occurs after midpuberty.

It is important to appreciate that a testis in the superficial inguinal pouch may be a temporarily retracted normal testis, a temporarily retracted high scrotal testis, a transiently palpable canalicular testis, or an obstructed testis and that differentiation among these possibilities is not always simple.

Pathogenesis

The cause of testicular maldescent is not well understood. The cryptorchid testis functions poorly in regard to both androgen secretion and spermatogenesis, but it is not always clear whether the testis functions poorly because of maldescent or fails to descend completely because it was abnormal to begin with. Maldescent of the testis occurs with increased frequency in many congenital defects, including virtually all disorders that impair virilization or prevent development of normal intra-abdominal pressure. Although maldescent is common in individuals with severe impairment of the androgen receptor, mutations of the androgen receptor gene are rare. Likewise, defects of the gubernaculum are common in cryptorchid testes, but mutations of the INSL 5 gene have not been detected in men with idiopathic cryptorchidism.

In some instances, a clear relation exists between maldescent and malfunction of the testis. For example, in the obstructed testis, in which a physical barrier prevents descent, and in syndromes in which intra-abdominal pressure is inadequate because the abdominal muscles are absent or incomplete, such as the prune-belly syndrome, inadequate testicular function in later life is the consequence of impeded descent. Conversely, in all series testes appear to have been abnormal from the first, and it is reasonable to assume that the defect in these instances plays a causal role in the maldescent.

As many as half of boys with a unilateral nonpalpable testis and a contralateral descended testis have an absent (vanishing) testis rather than a cryptorchid testis. The diagnosis of vanishing testis is suggested when compensatory testicular hypertrophy causes the contralateral descended testis to be larger than the mean for the age. Cryptorchid testes can sometimes be identified by magnetic resonance imaging or ultrasonography but definitive diagnosis may require laparoscopy.

Sequelae

About 10% of testicular tumors arise in an undescented testis, whereas cryptorchidism is present in fewer than 1% of adult men. Thus, malignancy is more likely to develop in an undescented testis than in a fully descended one. The greatest risk of malignancy is associated with intra-abdominal testes. Such malignancies commonly involve the germ cells, most commonly seminomas or embryonal cell carcinomas. Surgical correction of cryptorchidism does not remove this risk because
malignancy may develop in a previously cryptorchid testis many years after orchiopexy. Moreover, the contralateral normally descended scrotal testis is the site of development of malignancy in approximately a fifth of tumors associated with unilateral cryptorchidism.

The frequency of malignancy in cryptorchid testes should not be exaggerated because the chance of tumor development in any individual with cryptorchidism is low, but lifelong follow-up is required. Each cryptorchid testis should be surgically placed in a site that allows ready examination, and if this is not possible the cryptorchid testis should be removed. Periodic examination of the testes should be mandatory in the routine care of men with a history of cryptorchidism. In unilateral cryptorchidism, regardless of whether surgical correction has been undertaken, overall androgen production and levels are generally normal, presumably because malfunction of one testis can be compensated by the other testis.

Although as many as 60% of men in whom bilateral cryptorchidism was corrected in childhood can father children, cryptorchidism is associated with defective spermatogenesis. Mean sperm density is lower in adult men after surgical repair of cryptorchid testes in childhood, and spermatogenesis can also be decreased in the normally descended testis in men in whom one testis is cryptorchid. Basal FSH levels and FSH responsiveness to GnRH are higher on average in such men. Considered together, these types of evidence support the concept that testicular malfunction, as evidenced by impaired spermatogenesis, is a major factor in maldescent.

Management

Cryptorchidism should be treated by surgical or medical means, or both, and although correction may not prevent all sequelae, it is generally believed that correction should be undertaken before age 5 years. In the case of intra-abdominal testes (unilateral or bilateral), the issue is to exclude vanishing testes as the cause and bring intra-abdominal testes into the scrotum, where they can be monitored by physical examination. Measurement of levels of serum inhibin B and AMH may be useful in distinguishing bilateral intra-abdominal testes from vanishing testes. The testes that cannot be brought into the scrotum should be removed. Likewise, obstructed testes must be treated surgically.

The unresolved issue concerns the role of medical therapy in boys with canicular or high scrotal testes; in relatively large randomized trials, treatment of such boys with hCG, GnRH, or both agents in seriatiim was said to cause descent of the testes into a normal position in about a fifth of cases. However, the stimulation of apoptosis of germ cells by hCG and GnRH may have long-term deleterious consequences. Apoptosis is increased 1 month after hCG treatment but returns to normal subsequently. In one study, however, the level of apoptosis in the original testicular biopsy correlated many years later negatively with testis volume and positively with serum FSH levels, whereas sperm density was not affected. Treatment of cryptorchid boys with GnRH or hCG before orchiopexy caused a similar decrease in germ cells per tubule. Whether any hormonal treatment of cryptorchidism affects subsequent fertility is not known.
Neonatal Life

It is not clear whether abnormality in the neonatal surge in testosterone secretion results in pathologic consequences in humans. As mentioned earlier, however, temporary inhibition of the pituitary-testicular axis in the neonatal primate is associated with impaired testicular function at puberty.
Puberty

The central issue in dealing with disorders of puberty in both sexes is separating subjects with true absence or precocity of pubertal development from those at the extreme limits of normal variation. Normal puberty in the male is variable in onset, duration, and sequence of events. The spectrum of normal puberty and the disorders of puberty are discussed in detail in Chapter 24. (Feminizing states in prepubertal and pubertal boys can result from either absolute or relative increases in estrogen levels, as discussed later.) However, partial impairments of puberty may not be recognized until adulthood, as in Klinefelter's syndrome (see Chapter 22), in some androgen resistance syndromes in men (see Chapter 22), and in isolated gonadotropin deficiency (see Chapter 24).
Adult abnormalities of testicular function can be due to hypothalamic-pituitary defects, testicular disorders, or abnormalities in sperm transport (Table 18-3). Most such abnormalities are associated with underandrogenization and infertility, but some exhibit isolated infertility. Even partial defects in Leydig cell function can cause infertility because spermatogenesis requires androgen action. Therefore, although the evaluation of infertility differs from that of the underandrogenized man, it is essential to exclude the presence of subtle Leydig cell dysfunction in every man with infertility.

Certain factors or conditions (e.g., hyperprolactinemia, radiation, cyclophosphamide administration, environmental toxins, autoimmunity, paraplegia, androgen resistance) can cause either isolated infertility or a combined defect in testicular function (see Table 18-3). In addition to the manifestations of androgen deficiency described earlier in relation to assessment of Leydig cell function, testosterone deficiency can cause osteoporosis and mild elevations of serum total and LDL cholesterol and triglyceride levels.

Infertility with Impaired Virilization

Disorders of the hypothalamus and pituitary gland can impair secretion of gonadotropins and cause secondary decreases in androgen production and spermatogenesis, either as an isolated defect or as part of more complex pituitary insufficiency (see Chapter 8). Thus, destructive lesions of the hypothalamus and pituitary gland such as infarction, pituitary macroadenomas, metastatic or suprasellar tumors, infections, granulomatous processes, or radiation injury can cause hypopituitarism and lead to a secondary testicular defect. Likewise, isolated, functional impairment of hypothalamic GnRH secretion can occur with fasting or critical illness.

Congenital isolated gonadotropin deficiency occurs in both sporadic and familial forms. The incidence of the disorder is not established, but in most centers it is second only to Klinefelter’s syndrome as a cause of hypogonadism in men. The disorder was originally described by Kallmann as a familial syndrome associated with anosmia and can be manifest in childhood as microphallus or cryptorchidism. Male urethral development is usually complete. Because most penile growth occurs during the latter two thirds of gestation, the presence of microphallus in this disorder has been interpreted as evidence for a role of pituitary gonadotropin in regulating testosterone production during the later portion of gestation. Growth in childhood is normal although bone age is usually retarded.

Most affected individuals are identified because of failure to undergo puberty. Some individuals, particularly familial cases, have associated defects such as cleft lip or palate, or both; hearing defect; colorblindness; and eye movement abnormalities. Less severely affected individuals have only partial defects in the production of FSH or LH. A variant in which Leydig cell function is impaired despite testes of normal or near-normal size was originally known as the fertile eunuch

<table>
<thead>
<tr>
<th>TABLE 18-3 – Adult Abnormalities of Testicular Function</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Infertility with Undervirilization</strong></td>
</tr>
<tr>
<td><strong>Hypothalamic-Pituitary</strong></td>
</tr>
<tr>
<td>Fasting, critical illness</td>
</tr>
<tr>
<td>Isolated gonadotropin deficiency</td>
</tr>
<tr>
<td>Adrenal hypoplasia congenita</td>
</tr>
<tr>
<td>GnRH receptor mutations</td>
</tr>
<tr>
<td>LH/FSH mutations</td>
</tr>
<tr>
<td>Cushing’s syndrome</td>
</tr>
<tr>
<td>Hyperprolactinemia</td>
</tr>
<tr>
<td>Hemochromatosis</td>
</tr>
<tr>
<td>Panhypopituitarism</td>
</tr>
<tr>
<td><strong>Testicular</strong></td>
</tr>
<tr>
<td>Developmental and structural defects</td>
</tr>
<tr>
<td>LH receptor mutations</td>
</tr>
<tr>
<td>Klinefelter's syndrome</td>
</tr>
<tr>
<td>XX male syndrome</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Acquired defects</td>
</tr>
<tr>
<td>Viral orchitis</td>
</tr>
<tr>
<td>Trauma</td>
</tr>
<tr>
<td>Radiation</td>
</tr>
<tr>
<td>Drugs (e.g., spironolactone, alcohol, ketoconazole, cyclophosphamide)</td>
</tr>
<tr>
<td>Environmental toxins</td>
</tr>
<tr>
<td>Autoimmunity</td>
</tr>
<tr>
<td>Granulomatous disease</td>
</tr>
<tr>
<td>Defects associated with systemic disease</td>
</tr>
<tr>
<td>Liver disease</td>
</tr>
<tr>
<td>Renal failure</td>
</tr>
<tr>
<td>Sickle cell disease</td>
</tr>
<tr>
<td>Immunologic disease (HIV, rheumatoid arthritis)</td>
</tr>
</tbody>
</table>
Infertile man


Gonadotropin therapy to induce fertility. Administration of hCG over the long term also causes serum testosterone levels to increase to normal adult male levels.

In the infant or the young child with microphallus, administration of testosterone for limited periods (3 months) may cause enlargement of the penis to the normal size.

Three forms of therapy have been used for hypogonadotropic hypogonadism:

1. Androgen replacement to virilize.
2. Gonadotropin therapy to induce fertility.

In the infant or the young child with microphallus, administration of testosterone for limited periods (3 months) may cause enlargement of the penis to the normal size. The more severe the deficiency, the longer GnRH must be administered to correct gonadotropin secretion. Isolated gonadotropin deficiency can be inherited as an autosomal dominant, autosomal recessive, or X-linked trait. The X-linked form with anosmia is the best characterized. The neurons that secrete GnRH originate in the olfactory placode of the fetus and migrate into the brain with the olfactory, terminalis, and vomeronasal nerves. A defect in the KAL gene in the X-linked disorder impairs the migration of these nerves and thus causes GnRH deficiency, the olfactory disturbance, and hypoplasia of the olfactory bulbs. The genetic locus has been assigned to Xp22.3, and the gene encodes a protein, anosmin, that has homology to neural cell adhesion molecules. In a study of 104 individuals with isolated gonadotropin deficiency, Oliveira and colleagues identified KAL mutations in 3 of 21 familial and in 4 of 39 sporadic cases but in no instance in the absence of anosmia and concluded that the majority of mutations do not involve KAL.

In the presence of olfactory disturbances, other midline defects, a positive family history, or a combination of these factors, the diagnosis of the KAL disorder is not difficult to establish either in an infant with microphallus or in an under-virilized adult. In men with anosmia or hyposmia, defects of the rhinencephalon may be demonstrated by magnetic resonance imaging. In older individuals without midline abnormalities or anosmia and with uninformative family histories, the diagnosis can be made (after the presence of a pituitary tumor is excluded) by documenting a normal acute response to GnRH administration after a week of GnRH treatment. This approach is rarely followed in practice. In the middle teen years, separation of individuals with hypogonadotrophic hypogonadism from those with delayed puberty may require prolonged observation (see Chapter 24).

Additional mutations that cause gonadotropin deficiency encompass a variety of different mechanisms. One involves a mutation on the X chromosome that causes the X-linked form of adrenal hypoplasia congenita in which adrenarcheal insufficiency is associated with hypogonadotropic hypogonadism and which is due to mutations or deletions of the DAX1 gene (see earlier), which encodes a member of the nuclear hormone receptor family of transcription regulatory factors that are expressed in the hypothalamus, pituitary gland, adrenal glands, and gonads. The fact that the response of LH to GnRH is variable suggests underlying defects in both the hypothalamic and pituitary gland.

One individual with adrenal hypoplasia congenita virilized in response to hCG, but 3 years of combination therapy with hCG and HMG did not induce spermatogenesis. The adrenarcheal insufficiency is manifested during infancy, and the hypogonadotropic hypogonadism is recognized at the time of expected puberty, although rarely both may be incomplete or late in onset.

Mutations in the GnRH receptor can cause an autosomal recessive form of congenital isolated gonadotropin deficiency unassociated with anosmia. The phenotype is variable, even within a given family, with some patients having partial responses to GnRH and partial hypogonadism. One family exhibited a decrease in the amplitude of LH pulses.

A functional form of hypogonadotropic hypogonadism has been described in a family with a mutation in the LH gene. The proband did not undergo spontaneous puberty and had low plasma testosterone and elevated immunoreactive LH levels, but testosterone levels increased after administration of exogenous LH and hCG. A missense mutation in the LH gene impaired binding of LH to the LH receptor. Heterozygous male carriers for the mutation may have low testosterone levels and impaired fertility.

Similarly, a partial deletion of the coding sequence of the FSH gene caused a truncation of the molecule and hypogonadism in one man. Serum FSH was undetectable, and LH levels were elevated despite the low testosterone level, suggesting that the absence of FSH impaired Leydig cell function.

Hypogonadotropic hypogonadism also occurs in the Prader-Willi syndrome (obesity, short stature, mental retardation, and hypotonia) caused by partial deletions or uniparental disomy of chromosome 15.

Three forms of therapy have been used for hypogonadotropic hypogonadism:

1. Androgen replacement to virilize.
2. Gonadotropin therapy to induce fertility.

In the infant or the young child with microphallus, administration of testosterone for limited periods (3 months) may cause enlargement of the penis to the normal range without affecting linear growth or causing other significant virilization. In the older child or the adult, long-acting testosterone esters are administered parenterally, as with other forms of hypogonadism (see later). As in other forms of androgen deficiency, the closer to the time of onset of normal puberty that replacement therapy is begun, the more effective the promotion of normal virilization.

Administration of hCG over the long term also causes serum testosterone levels to increase to normal adult male levels. In men with severe (prepubertal) hypogonadotropic hypogonadism, however, the induction of fertility usually requires the administration of FSH, in the form of human menopausal gonadotropin, in addition to hCG. The response to gonadotropin therapy is not influenced by prior testosterone therapy but is a function of the initial testis size, men with testes less than 4 mL in volume responding less favorably. Once a normal sperm count is achieved, it may be maintained by use of hCG or, occasionally, by testosterone esters. In rare cases of partial defects in gonadotropin secretion, spermatogenesis can be promoted by testosterone therapy alone. The long-term administration of GnRH in a pulsatile manner to men with hypogonadotropic hypogonadism results in normal plasma testosterone levels, normal pulsatile secretion of LH, normal mean levels of plasma LH and FSH, and, in most, mature sperm in the ejaculate.

Acquired isolated gonadotropin deficiency can be caused by pathologic states that impair the hypothalamus or pituitary secondarily. For example, elevated plasma cortisol levels, as in Cushing's syndrome, can depress LH secretion independently of a space-occupying lesion of the pituitary. Likewise, chronic administration of exogenous glucocorticoids can lower testosterone levels by inhibiting GnRH secretion.
by impairing GnRH release. Administration of low doses of bromocriptine to men with microadenomas caused an initial increase in plasma LH level and a subsequent increase in serum testosterone level. The preferred treatment is bromocriptine or cabergoline (see Chapter 8).

Hemochromatosis causes iron deposition in the pituitary gland and testes, and about half of affected men have hypogonadism, usually accompanied by testicular atrophy. Abnormal testicular function in this disorder may result in part from the associated liver disease, but most testicular dysfunction is due to hypogonadotropic hypogonadism. The pituitary nature of the hypogonadism was recognized because of the lack of response of LH to GnRH administration and the normal response of plasma testosterone to hCG. A primary testicular abnormality may also occur. Acquired transfusional iron overload can cause similar abnormalities of the pituitary-testicular axis. In both states, reduction of iron stores may result in recovery of gonadotropin secretion.

Hypothalamic or pituitary injury can occur after head trauma even in the absence of fracture, and the most common manifestation is deficiency of gonadotropins and human growth hormone, although multiple deficiencies can be present. Clinical evidence of hormone deficiency may be apparent immediately after injury or not until years later. In conditions in which testosterone levels are decreased despite normal LH levels, the mechanism is less clear. Men with massive obesity have decreased SHBG levels and decreased levels of total and bioavailable testosterone that return toward normal with weight loss. For example, in men with a body mass index greater than 40, free testosterone levels, LH, and LH pulse amplitude are decreased, implying malfunction of the hypothalamic-pituitary system. Obesity may be the cause of decreased testosterone levels in the pickwickian syndrome. Men with temporal lobe seizures may also have hormonal findings consistent with hypogonadotropic hypogonadism. Finally, acquired hypogonadotropic hypogonadism may be idiopathic.

Abnormalities of testicular function in the adult can be grouped into developmental and structural defects of the testes, acquired testicular defects, abnormalities associated with systemic or neurologic diseases, and androgen resistance.

**TABLE 18-4 — Characteristics of Patients with Classic Versus Mosaic Klinefelter's Syndrome.**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>47,XXY (%)</th>
<th>47,XY/47,XXY (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abnormal testicular histologic features</td>
<td>100</td>
<td>94</td>
</tr>
<tr>
<td>Decreased length of testis</td>
<td>99</td>
<td>73</td>
</tr>
<tr>
<td>Azospermia</td>
<td>93</td>
<td>50</td>
</tr>
<tr>
<td>Decreased testosterone level</td>
<td>79</td>
<td>33</td>
</tr>
<tr>
<td>Decreased facial hair</td>
<td>77</td>
<td>64</td>
</tr>
<tr>
<td>Increased gonadotropin level</td>
<td>75</td>
<td>33</td>
</tr>
<tr>
<td>Decreased sexual function</td>
<td>68</td>
<td>56</td>
</tr>
<tr>
<td>Gynecomastia</td>
<td>55</td>
<td>33</td>
</tr>
<tr>
<td>Decreased axillary hair</td>
<td>49</td>
<td>46</td>
</tr>
<tr>
<td>Decreased length of penis</td>
<td>41</td>
<td>21</td>
</tr>
</tbody>
</table>


The preferred treatment is bromocriptine or cabergoline (see Chapter 8). The more severe defects in LH receptor function result in an autosomal recessive form of male pseudopseudohernaphroditism in which 46,XY individuals have a female phenotype, a blind-ending vagina, and inguinal testes with absence of Leydig cells (Leydig cell hypoplasia) (see also Chapter 22). Less severe impairments of LH receptor function are associated with variable defects, including a male phenotype and microphallus. Testosterone levels are low in the presence of elevated gonadotropins. In most subjects, the phenotypic defects correlate with both the basal level of the LH receptor and the response of plasma testosterone to hCG, but in one individual with a male phenotype associated with a deletion of exon 10 of the LH receptor gene, serum testosterone levels increased after treatment with hCG.

Klinefelter's Syndrome.

The most common developmental defect of the testis is Klinefelter's syndrome (see also Chapter 22). The disorder is characterized by small, firm testes; various degrees of impaired sexual maturation; azospermia; gynecomastia; and elevated gonadotropin levels. The underlying abnormality is the presence of an extra X chromosome, the usual chromosomal karyotype being either 47,XXX (classic form) or 46,XY/47,XXX (mosaic form). The incidence is approximately 1 in 500 males.

Prepubertal boys with Klinefelter's syndrome have small testes with a decreased number of spermatogonia but are endocrinologically normal; the diagnosis at this age may be made on the basis of the cognitive features (see later). The diagnosis is usually made after the time of expected puberty because of gynecomastia or underandrogenization and later by infertility (Table 18-4). Damage to the seminiferous tubules and azoospermia are consistent features of the 47,XXX variety. The small, firm testes usually less than 2 cm in length and less than 4 mL in volume. Histologic changes in the testes include hyalinization of the tubules, absence of spermatogenesis, and an apparent increase in the number of Leydig cells (see Fig. 18-17 C).

Mean body height is increased because of a longer lower body segment; the presence of this feature before puberty suggests that it is not secondary to androgen deficiency but is probably related to the underlying chromosomal abnormality. Gynecomastia occurs in about 85% of affected individuals, develops during adolescence, is usually bilateral and painless, and may become disfiguring.

Klinefelter's syndrome can cause learning disabilities and poor impulse control. These tendencies may explain the increased frequency of the disorder among men in mental and penal institutions. Indeed, although many men with the disorder have above average or superior intelligence, poor school performance is common, with decreased verbal scores and a higher incidence of dyslexia and attention-deficit disorder.

The risk of breast cancer is increased, presumably because of the presence of gynecomastia, and there is an increased prevalence of extragonadal germ cell tumors in the mediastinum and brain. Autoimmune disorders appear to be more common, perhaps related to the altered levels of gonadal hormones.
Varicose veins, venous stasis ulcers, and thromboembolic disease are also more common. Levels of plasminogen activator inhibitor 1 were reported to be elevated in subjects with Klinefelter's syndrome with leg ulcers but not in those without leg ulcers. Forty percent of men with Klinefelter's syndrome have taurodontism, an abnormality of the dental pulp that predisposes to early tooth decay. Most have a male psychosexual orientation and function sexually as men. The syndrome may go undiagnosed in the majority of affected men, even as adults.

46,XY/47,XXY mosaicism is the cause of about one fourth of cases of Klinefelter's syndrome, as estimated by chromosomal karyotypes of peripheral blood leukocytes. The true prevalence may be underestimated because chromosomal mosaicism can be present in the testes in individuals in whom the chromosomal karyotype of peripheral leukocytes is normal. As summarized in Table 18-4, the manifestations of the mosaic form are usually less severe, and the testes may be normal in size. The endocrine abnormalities are also less severe, and gynecomastia and azoospermia are less common, with occasional men being fertile. Additional karyotypic variations of Klinefelter's syndrome have been described (see Chapter 22).

The 47,XXY form of Klinefelter's syndrome is due to mectic nondisjunction of the chromosomes during gametogenesis. About 40% of the responsible meiotic nondisjunction occur in the father and 60% occur in the mother. Advanced maternal age is a predisposing factor in the latter cases. In contrast, mitotic nondisjunction after fertilization of the zygote causes the mosaic form and can arise in either a 46,XY zygote or a 47,XXY zygote.

Characteristic endocrine changes include elevation of plasma FSH and LH levels. FSH shows the best discrimination, and little overlap occurs with normal individuals, a consequence of the consistent damage to the seminiferous tubules. In the late teens, the plasma testosterone level may be normal. By the middle 20s, the plasma testosterone level averages half the normal value, but the range is broad and overlaps the normal range. Mean plasma estradiol levels are elevated, and SHBG levels are about twice normal. The net result is a variable degree of feminization and insufficient androgenization.

The feminization, including development of gynecomastia, is thought to depend on the ratio of circulating estrogen to androgen (see later). Before puberty, plasma gonadotropin levels and the response to GnRH are normal, but by the time of expected puberty plasma gonadotropins and the response to GnRH are elevated. Older men with untreated Klinefelter's syndrome may have an enlarged or an abnormal sella turcica, presumably secondary to impairment of gonadal steroid feedback and gonadotrope hyperplasia.

Optimally, affected boys should be identified in childhood to allow treatment of the testosterone deficiency early before it has an effect on BMD and to prevent adverse psychological effects of incomplete sexual maturation. The commencement of testosterone replacement before age 20 results in a normal BMD, whereas institution of therapy in older individuals does not enhance BMD. Androgen replacement has no effect on fertility but has the same general benefits as in other forms of male hypogonadism. Plasma LH levels usually return to normal with therapy, sometimes after many months. Men with rare spermatozoa in the ejaculate or with spermatids or more advanced stages of spermatogenesis on biopsy may achieve fertility with in vitro fertilization using ICSI (see later). However, as many of 15% of the sperm produced by men with the disorder contain a 24,XY chromosome composition, which may result in an increased incidence of the disorder in offspring.

XX Male Syndrome.

The XX male syndrome, a variant of Klinefelter's syndrome, occurs in about 1 in 20,000 male births. The tests are small and firm, generally less than 2 cm long; gynecomastia is usual; the penis is normal to small in size; and azoospermia and hyalinization of the seminiferous tubules are present. Affected individuals have male psychosexual identification and absence of female internal genitalia. The mean plasma testosterone level is low, and levels of plasma estradiol and gonadotropins are high. The phenotype differs from that of Klinefelter's syndrome in that the average height is less than that of normal men, the incidence of cognitive impairment is not increased, and hypospadias or ambiguous genitalia may be present.

Four theories were proposed to explain male development in the absence of a Y chromosome: (1) mosaicism in some tissues for a Y-chromosomecontaining cell line, (2) a gain-of-function mutation for some autosomal gene, (3) deletion or inactivation of some gene or genes that normally suppress testicular development, and (4) interchange of a portion of the Y chromosome with the X chromosome. Evidence has now been obtained for the presence of mechanisms 1, 2, and 4 in men with this disorder. Y-chromosome sequences are detectable in approximately 80% of 46,XX men and are usually located in the distal region of the X chromosome. Thus, the etiology in most XX males is analogous to that in the sxd mouse, in which a fragment of the X chromosome has been translocated to the X chromosome.

Mosaic involving an intact Y chromosome was present in 1% of the cells in one individual.

The other third of 46,XX men are Y-negative and lack sequences for SRY. The Y-negative group is more likely to have ambiguity of the external genitalia, whereas the Y-positive group has the Klinefelter phenotype. The translated region of the Y can be quite small and involve only the SRY gene itself. The SRY-negative variant is sometimes familial and may occur in families with true hermaphroditism, suggesting that these disorders are due to variable manifestations of the same genetic defect. Such a defect could either be autosomal or X-linked. One 46,XX male had a duplication of the 509X gene (a downstream gene involved in SRY-mediated testicular differentiation), indicating an autosomal mechanism.

In one SRY-positive man with genital ambiguity, the presence of a duplication of the 509X gene, a downstream target of SRY in testicular differentiation, indicated autosomal inheritance. Most SRY-positive 46,XX males do not have genital ambiguity, and in one SRY-positive man with genital ambiguity more than 90% of the of the Y fragment containing SRY was located on the inactive X chromosome in blood lymphocytes. In parallel studies of a 46,XX man with no ambiguity, the Y fragment was located predominately in the active X chromosome. These findings document the complexity and heterogeneity of the disorder. The management is similar to that for Klinefelter's syndrome.

Acquired Defects

Mumps.

The most common cause of acquired testicular failure in the adult is viral orchitis. Mumps virus is most frequent, but echovirus, lymphocytic choriomeningitis virus, and group B arboriviruses also cause orchitis. The disorder is due to infection of the tissue by the virus rather than to indirect effects of infection. Orchitis is common in mumps, occurring in as many as a fourth of adult men with the disease. In about two thirds of cases it is unilateral. It usually develops 4 to 8 days after the onset of parotitis but occasionally precedes it. After the acute inflammatory phase, the tests gradually decreases in size, although swelling can persist for months. The tests may return to normal size and function or undergo atrophy. The atrophy results from both the direct effects of the virus and ischemia caused by pressure and edema within the taut tunica albuginea. The histologic features of the atrophic tests include progressive tubular sclerosis and hyalinization. The degree of atrophy is not necessarily proportional to the severity of the orchitis. It is usually apparent within 1 to 6 months after the orchitis subsides, but the full extent of damage may not be evident for many years. Atrophy occurs in approximately a third of men with orchitis and is bilateral in about a tenth. The hormonal changes associated with gynecomastia related to mumps orchitis include normal estrogen and decreased testosterone production.

The frequency with which mumps results in infertility is not known. Almost 50% of men with unilateral mumps orchitis have sperm densities of less than 10 million/mL in the first 3 months, but the sperm count returns to normal within 1 to 2 years in about 75%. In contrast, semen parameters return to normal in less than a third of men with bilateral orchitis.

The initial treatment is bed rest and scrotal support. If pain is severe, administration of prednisone can reduce swelling and pain. Glucocorticoid therapy does not appear to have a beneficial effect on the return of the sperm count to normal. In one study, treatment with interferon shortened the duration of symptoms and...
Effects on plasma LH levels.

Both spermatogenesis and testosterone production are sensitive to radiation; impaired secretion of testosterone appears to result from decreased testicular blood flow. The incidence of radiation-associated damage to Leydig cells is directly related to the dosage and inversely related to age at treatment. Most prepubertal boys have normal plasma testosterone levels and normal pubertal maturation after receiving 12 Gy of radiation to the testes, but the presence of an increased LH level in some suggests that compensatory changes are involved in the achievement of normal testosterone levels. In most prepubertal boys, radiation doses greater than 20 Gy cause permanent testosterone deficiency. In contrast, radiation doses above 30 Gy cause testosterone deficiency in only half of adolescent boys and young adults. (Also see “Infertility with Normal Virilization.”)

Drugs.

Drugs can cause underandrogenization and infertility in several ways: direct inhibition of testosterone synthesis, blockade of the peripheral actions of androgen, and enhancement of estrogen levels. In addition, agents such as propranolol and guanethidine can impair erectile function in men whose hypothalamic-pituitary-testicular axis is normal.

Two drugs that in high doses block testosterone synthesis are spironolactone and cyproterone, both of which interfere with the late reactions in testosterone biosynthesis. Spironolactone appears to impair CYP17 activity. Plasma testosterone levels do not change appreciably, however, during usual therapeutic regimens.

The antifungal agent ketoconazole blocks testosterone synthesis, also by inhibiting CYP17 activity. The decrease in testosterone after a single dose of ketoconazole is transient, with the nadir occurring within 4 to 8 hours and testosterone returning to baseline within 24 hours as ketoconazole levels fall. However, with doses of ketoconazole greater than 400 mg/day, depression of plasma testosterone levels may be sustained. Impairment of libido is common in men with epilepsy, partly as a consequence of medication.

Enzyme-inducing antiepileptic drugs such as phenytoin and carbamazepine lower bioavailable testosterone, raise plasma SHBG and LH levels, and decrease the metabolic clearance of testosterone. The effect is more pronounced with multiple-drug regimens. Valproic acid does not appear to have as severe an adverse effect in this regard.

Independent of its effects on the liver, ethanol ingestion reduces testosterone levels acutely and chronically. The result of inhibition of testosterone synthesis. The inhibition of steroidogenesis appears to occur at the 3-β-HSD reaction as the result of a decrease in the concentration or availability of the pyridine nucleotide cofactors for the reaction, an effect probably mediated by the ethanol metabolite acetaldehyde. The fact that ethanol lowers testosterone levels without causing appropriate elevations of plasma LH suggests that hypothalamic-pituitary function is also impaired. Ethanol can also impair spermatogenesis.

Antineoplastic and chemotherapeutic agents, especially cyclophosphamide, commonly induce infertility (see later). Combination chemotherapy for acute leukemia, Hodgkin's disease, and other malignancies may also impair Leydig cell function. This toxic effect on the Leydig cell seems to be produced primarily by alkylating agents. Treatment with alkylating agents during the prepubertal years does not interfere with testicular function in later life, but elevated LH levels develop in many men after treatment, implying the presence of subclinical Leydig cell dysfunction. Treatment of adult men with alkylating agents does not alter LH or testosterone levels. High-dose interleukin-2 therapy for metastatic cancer causes a transient reduction in serum testosterone levels.

Plasma testosterone levels may be low in men ingesting large amounts of marijuana, heroin, or methadone. Plasma LH is usually normal, suggesting combined hypothalamic-pituitary and testicular defects. Hyperprolactinemia may contribute to the lowering of testosterone levels.

Elevated plasma estradiol and decreased plasma testosterone levels may occur in men taking digitalis preparations, the mechanism being unclear.

Drugs can interfere with gonadotropin production either as the result of a direct inhibition (as in medroxyprogesterone acetate administration) or as a secondary consequence of enhanced prolactin secretion (as with phenothiazine therapy). Medroxyprogesterone may also impair testosterone secretion at the testicular level.

Several drugs inhibit androgen action by competition at the receptor level. Although spironolactone can inhibit testosterone synthesis, in the usual dosage regimens it acts primarily by antagonizing the binding of androgen to the androgen receptor, which leads to gynecomastia and impotence.

Cyproterone also acts as an androgen antagonist. The most commonly administered androgen antagonist is cimetidine. Gynecomastia can occur in men who are treated with the drug, and decreased sperm density and elevated basal testosterone levels are accompanied by impairment of the LH response to GnRH.

Ranitidine appears to be a less potent antiandrogen. Omeprazole can also cause gynecomastia and impotence.

Environmental Toxins.

Prolonged exposure to lead results in direct testicular toxicity and an impaired pituitary response of plasma LH.

Autoimmune Disorders.

Testicular failure can occur as part of a generalized autoimmune disorder in which multiple primary endocrine deficiencies coexist and circulating antibodies to the basement membrane of the testes are present (see Chapter 37).

Granulomatous Disease.

The testes can also be involved in granulomatous disease. Testicular atrophy occurs in 10% to 20% of men with lepromatous leprosy as the result of invasion of the tissue and in some instances of paratesticular structures as well) by the bacilli. The result is a decreased plasma testosterone level and elevated plasma LH and FSH levels. Destruction of the testis is less common with other systemic granulomatous diseases.

Abnormalities of the hypothalamic-pituitary-testicular axis occur in a number of systemic diseases. Given the chronic ill health and generalized wasting that may coexist, it is often difficult to distinguish specific effects of the underlying condition (e.g., renal failure from those of malnutrition. The inflammatory cytokines interleukin-1, tumor necrosis factor, and interleukin-6 lower testosterone and have variable effects on plasma LH levels. In in vitro preparations, these agents inhibited several steroidogenic enzymes and SIAR protein.
Renal Failure.

About 50% of men undergoing dialysis for renal failure experience decreased libido and impotence associated with impairments in both spermatogenesis and testosterone biosynthesis. The defect in spermatogenesis varies from partial to total destruction of the germ cells. The decrease in plasma testosterone and increase in plasma LH and FSH levels indicate a defect at the testicular level. Plasma testosterone production rates are decreased, and the response of plasma testosterone to hCG is subnormal. After dialysis, plasma testosterone levels and testosterone production rates improve but usually not to the normal range.

Hyperplactinemia occurs in 25% of men who undergo long-term dialysis. In contrast, successful renal transplantation results in a return of testosterone and prolactin levels to normal and a slight decrease in LH and FSH levels. Most men experience improved sexual function after transplantation, and half have sperm densities of more than 10 million/ml.

Hepatic Diseases.

Cirrhosis of the liver can impair testicular function apart from the direct toxic effects of ethanol. Gynecomastia and testicular atrophy occur in half of men with cirrhosis, and 75% of men with hepatic cirrhosis are impotent. Decreased spermatogenesis and peritubular fibrosis are present in about 50% of patients. Plasma estradiol levels are usually elevated, and plasma testosterone levels are decreased. The net result is a ratio in serum of unbound estradiol to unbound testosterone that is about 10 times normal. Levels of SHBG are about twice normal. The metabolic clearance and production rates of testosterone are decreased, and estradiol production is increased.

Extraluminal conversion of androgens, primarily adrenal androgens, to estradiol and estrone is increased about three-fold, presumably because of decreased hepatic extraction of androgens. Basal levels of LH and FSH range from normal to moderately elevated. In men with low testosterone levels pulsatile LH secretion is impaired, implying a defect at the hypothalamic-pituitary level, whereas dynamic tests of the pituitary-testicular axis and hCG responsiveness tests point to a testicular defect.

Gonadotropin levels decrease as liver function fails. The reason for abnormal testicular and hypothalamic-pituitary function is uncertain. Elevated estrogen levels can cause both defects. Testosterone therapy has been tried. Although estradiol levels increase (in correlation with the severity of the cirrhosis) after androgen therapy, there may be normal or increased testosterone excretion, the estrogen/testosterone ratio may become normal, and gynecomastia may regress. Men with alcoholic cirrhosis may experience spontaneous recovery of sexual function with abstinence from alcohol despite persistent liver abnormalities. Men with alcoholic cirrhosis and testicular atrophy, however, are less likely to experience improvement in sexual function with abstinence from alcohol. Sexual function and testosterone levels are also decreased in men with other forms of liver disease. The hormonal abnormalities in the pituitary-testicular axis can be reversed by liver transplantation.

Sickle Cell Anemia.

Sexual maturation can be impaired in boys with sickle cell anemia. Furthermore, in adult men with sickle cell anemia, secondary sexual characteristics are subnormal in most and testicular atrophy occurs in about a third. Testicular biopsy in two men revealed maturation arrest of spermatogenesis. The defect may be either testicular or hypothalamic.

Chronic Illness.

Abnormal Leydig cell function, frequently accompanied by decreased sperm counts, occurs in many systemic diseases including protein-calorie malnutrition, advanced Hodgkin's disease and cancer before chemotherapy, cystic fibrosis, chronic pulmonary disease, and amyloidosis. Such disorders usually cause a low plasma testosterone level and either a normal or slightly increased plasma LH level, suggesting combined hypothalamic-pituitary and testicular defects. The low plasma testosterone level is not the result of plasma factors that inhibit its binding to SHBG and hence is not analogous to that in the euthyroid sick syndrome (see Chapter 10). Indeed, because mean plasma SHBG levels are elevated, bioavailable testosterone may be lower than total testosterone. These changes in testosterone and LH may be nonspecific effects of illness because similar changes occur after myocardial infarction, and severe burns.

Thyrotoxicosis.

The changes in the hypothalamic-pituitary-testicular axis in thyrotoxicosis may be secondary to increased estrogen levels and include decreased sperm count and semen volume, increased plasma total testosterone level, and normal levels of unbound testosterone. The testosterone response to hCG is blunted, and basal LH levels are increased.

Immune Disorders.

Immune disease may cause testicular dysfunction, and primary or secondary hypogonadism is common in men with acquired immunodeficiency syndrome (AIDS), about half of whom have low testosterone levels with normal or appropriately increased plasma LH. Although the disease may involve the testes directly, the hormonal changes suggest a nonspecific response to systemic illness and perhaps the toxic effects of immune cytokines. Many HIV-positive men who progress to AIDS have a transient state of increased LH and normal testosterone levels before testosterone levels become low. Serum SHBG levels are increased in HIV-positive men independent of CD4 count but bioavailable testosterone levels decrease progressively as CD4 counts decline. Levels of bioavailable testosterone decline in men with AIDS before wasting occurs.

The current practice is to provide androgen replacement in men with AIDS when testosterone deficiency becomes manifest. Testosterone replacement in such men results in an increase in lean body mass, improved quality of life, and increased muscle strength. Men with rheumatoid arthritis may have low serum testosterone, particularly during disease flares and while they are receiving glucocorticoids. In contrast, testosterone levels are usually normal in men with long-standing, stable rheumatoid arthritis.

Neurologic Disease.

Neurologic disease can cause testicular abnormalities. Men with myotonic dystrophy usually have small testes, low plasma testosterone levels, and elevated plasma LH and FSH levels. Spinobulbar muscular atrophy (Kennedy's syndrome) is a form of adult-onset degenerative motor neuropathy associated with gynecomastia, testicular atrophy, a hormonal profile suggestive of androgen resistance (see later), and an expansion of the homopolymeric glutamine repeat region in the amino-terminal end of the androgen receptor gene. Although the effects are variable, spinal cord lesions that cause quadriplegia or paraplegia initially cause diminished plasma testosterone levels that usually return toward normal, but semen parameters may be permanently abnormal. Some paralyzed men retain the capacity to have erections and ejaculate, depending on the extent of involvement of the lumbosacral spinal cord.

Men with trisomy 21 (Down's syndrome) have impairment of both germinal and Leydig cell function and elevation of FSH and LH levels.

Androgen Resistance.

Partial androgen resistance can cause underandrogenization and infertility in men with normal external genitalia. In such men, androgen resistance is manifested by increased testosterone production, elevated plasma LH levels in some, and abnormal androgen receptors in cultured genital skin fibroblasts. The presence of elevated testosterone or LH levels, or both, is not a reliable predictor of which men have a receptor defect. Testicular biopsy reveals maturation arrests or germinal
cell aplasia similar to that shown in Figure 18-16D and E. In some families, affected men have had gynecomastia and undervirilization but, in some, preserved fertility.\(^{112}\) Point mutations have been identified in the androgen receptor in men with the isolated infertility and undervirilized, fertile male phenotypes (see Chapter 22).\(^{113}\)

Infertility with Normal Virilization

Isolated infertility with normal Leydig cell function is caused by a separate group of disorders. Isolated infertility can be due to defects in the hypothalamic-pituitary system, the testis, or the sperm transport system (see Table 18-3).\(^{114}\)

Hypothalamic-Pituitary Disorders

Isolated FSH deficiency has been reported in men in whom virilization and plasma LH and plasma testosterone levels were normal but plasma FSH levels were persistently low.\(^{115}\) Plasma FSH levels in such men may increase or remain undetectable after GnRH administration. One man with no FSH response to GnRH had a point mutation in the FSH gene.\(^{116}\) In some men with chronic untreated or undertreated congenital adrenal hyperplasia related to CYP21 deficiency, suppression of gonadotropin secretion by adrenal androgens causes infertility.\(^{117}\) This diagnosis is suggested by the presence of small testes, normal to elevated levels of testosterone, and suppressed levels of gonadotropins and is confirmed by finding elevated plasma levels of 17-hydroxyprogesterone and androstenedione (see Chapter 14 and Chapter 22).

When androgens are administered in pharmacologic doses to normal men, gonadotropins are suppressed and about 50% of the men have azoospermia (see Chapter 17). Although men who present with isolated infertility are unlikely to be receiving testosterone replacement therapy, the use of androgens by weight-lifters and body-builders is common. Self-prescribed regimens may include parenteral testosterone esters and a variety of oral and parenteral substituted androgens, often termed anabolic steroids (see later). Supraphysiologic androgen administration can cause reversible azoospermia in normal men.\(^{118}\)

Testicular Disorders

Developmental and Structural Defects

Germinal Cell Defects.

Also known as the Sertoli cell only syndrome, germinal cell aplasia is a poorly understood defect of the testis. The disorder encompasses histologic features that can have several causes, one of which may be a single-gene defect. Other men with typical histologic and clinical features have a history of viral orchitis, cryptorchidism,\(^{119}\) alcoholism,\(^{120}\) or androgen resistance.\(^{121}\) Testicular biopsy reveals complete absence of germinal elements (see Fig. 18-16F). The clinical features include azoospermia, normal virilization, absence of gynecomastia, normal to small testes, and normal chromosomal complement. Plasma testosterone and LH values are usually normal, and plasma FSH values are high.

The concept of germinal cell aplasia became even more complex with the recognition that one or more Y chromosome determinants other than the SRY gene are essential for spermatogenesis.\(^{122}\) Some men have a deletion of the long arm of the Y chromosome that includes an azoospermia factor (AZF) that maps to Yq11.23-12.\(^{123}\) As many as 18% of men with azoospermia (occasionally severe oligospermia) have chromosome microdeletions in this region.\(^{124}\) Testicular histology varies from germinal cell aplasia to maturation arrest, and the plasma FSH values are high.

Candidate genes for AZF have been identified by positional cloning; the first is a family of genes termed Y-located RNA recognition motif (YRMR) genes\(^{125}\) that encode RNA-binding proteins. The YRMR genes are expressed in germ cells, but the fact that multiple genes exist in this family makes it hard to assess their function.

The second AZF candidate, termed DAZ (deleted in azoospermia), also encodes a testis-specific RNA recognition motif.\(^{126}\) In one study, microdeletions in the AZF region were present in a third of men with idiopathic azoospermia and a fourth of men with oligospermia of unknown origin.\(^{127}\) These microdeletions include the DAZ and YRMR regions and additional sequences.\(^{128}\) Microdeletions were present in 7% of 46,XY men with known causes of infertility.\(^{129}\)

A mutation in the FSH receptor gene in several Finnish families is associated with variable defects in spermatogenesis and infertility.\(^{130}\)

Histologic findings in men with azoospermia include hypospermatogenesis or spermatogenic arrest (see Fig. 18-16F). Familial male infertility with hypospermatogenesis or maturation arrest can be inherited as X-linked\(^{131}\) or autosomal recessive traits.\(^{132}\) In most men, however, the family history is uninformative\(^{133}\) and the cause of infertility is unknown. In both familial and sporadic cases, meiosis is defective.

Cryptorchidism.

Unilateral cryptorchidism, even when corrected before puberty, is associated with abnormal semen in many individuals (see earlier). This finding suggests that the testicular abnormality is bilateral even in unilateral cryptorchidism.

Varicocele.

Varicocele is believed to be the most common treatable cause of male infertility, possibly of causative importance in a third of infertile men.\(^{134}\) Varicocele is caused by retrograde flow of blood into the internal spermatic vein and results in a progressive, often palpable, dilation of the pampiniform plexus of veins. It is thought to be due to incompetence of the valve between the internal spermatic vein and the renal vein and is more common (85%) on the left. The incidence of varicocele is about 10% to 15% in the general population and 20% to 40% in men with infertility. The findings on semen analysis are usually nonspecific, but sperm density is often decreased with medium or large varicocoeles.\(^{135}\)

Most men with varicocele are fertile and have no detectable abnormality of the hypothalamic-pituitary-testicular axis. The leading theory concerning the adverse effect is that varicocele leads to an increased scrotal (and testicular) temperature, and the elevated scrotal surface temperatures in men with both unilateral and bilateral varicoceles can improve after surgical repair.\(^{136}\) The tests on the side of the varicocele may be small.\(^{137}\) Men with varicocele can also have unrelated causes of infertility.\(^{138}\) Of men with varicocele and sperm counts less than 5 million/mL, about a fifth have microdeletions in the Y chromosome.

On average, semen quality improves after surgical repair of varicoceles, but the effect on fertility is inconsistent in that impregnation rates after varicocele repair are probably less than 50%. In one large study of almost 100 men there was an association between subsequent fertility and preoperative sperm density, and the men with preoperative sperm densities greater than 10 million/mL had a 70% impregnation rate after repair.\(^{139}\)

Immotile Cilia Syndrome.

The immotile cilia syndrome is a hereditary disorder characterized by defective motility of the cilia in the airways and elsewhere in the body and either immotile or poorly motile sperm.\(^{140}\) The disorder is usually inherited as an autosomal recessive trait. In the airways, defective cilia cause chronic sinusitis and bronchiectasis and the immotile sperm cannot fertilize. Kartagener's syndrome is a subclass of the syndrome and is associated with situs inversus.

The structural abnormalities that impair motility of cilia can be defined by electron microscopy and include missing or abnormally short dynein arms, short spokes with no central sheath, missing central microtubules, and displacement of one of the microtubule doublets. Cilia from epithelia and sperm from the same individual usually exhibit the same defects, but some mutations can apparently result in immotile sperm without impairment of cilia in the lung.\(^{141}\) In evaluating sperm for structural
abnormalities, the physician should take care to examine a number of axonemes and to confirm the structural defect because axonemal structure can vary in normal respiratory cilia and sperm. The infertility should be treatable, at least theoretically, by empirical methods (see later).

Acquired Defects

Mycoplasma Infection.

A role for Mycoplasma (Ureaplasma urealyticum) in infertility is inferred because infections are common in women whose infertility is associated with a “male factor,” suggesting that genital tract mycoplasma infection may cause male infertility.  Other evidence suggests that the severity of mycoplasma infection in men does not correlate with any specific alteration in sperm density or morphology.  Other evidence suggests that the presence of mycoplasma in the lower urogenital tract may represent silent colonization rather than infection.

Radiation.

Radiation can cause isolated infertility, with damage sometimes being demonstrable after only 0.15 Gy (15 rad). Doses higher than 1 Gy (100 rad) can cause extreme oligospermia or azoospermia, and higher doses decrease sperm counts and damage spermatids. A return to baseline sperm density takes 9 to 18 months after doses of 1 Gy (100 rad) or less, 30 months for doses of 2 to 3 Gy (200 to 300 rad), and 5 years or more for doses of 4 to 6 Gy (400 to 600 rad).  Fracionated radiation may have a more profound effect on the testes than single doses.

Permanent infertility can occur after radiation for malignant lymphoma of the abdomen despite shielding of the testes.  Administration of radioactive iodine to men for thyroid cancer can also impair spermatogenesis and elevate plasma FSH levels; recovery occurs in about 2 years.  Prior suppression of testicular function by administration of testosterone or GnRH, or both, does not protect the testes from radiotherapy (or cytotoxic drugs).

Drugs.

The principal drugs that cause isolated infertility are alkylating agents such as cyclophosphamide.  Spermatocytes and spermatogonia may disappear completely, causing the picture of germinal aplasia with only Sertoli cells lining the tubular lumen. Serum FSH levels can increase fivefold and serve as a marker for germ cell loss; levels of inhibin B decline.  Serum LH and testosterone levels usually remain within normal limits. Cessation of cyclophosphamide therapy is followed by return of spermatogenesis within 3 years in about half of azoospermic men.  Vinblastine, doxorubicin, procarbazine, and cisplatin are also toxic to the germinal epithelium.

Combination regimens such as mechlorethamine, vincristine, procarbazine, and prednisone (MOPP) have an even more profound impact on spermatogenesis.  The combination of doxorubicin, bleomycin, vinblastine, and dacarbazine is less toxic than MOPP.  Combination chemotherapy that includes cyclophosphamide or procarbazine causes postpubertal azoospermia in about half of prepubertal boys, whereas etoposide in combination chemotherapy causes less toxicity.

Chemotherapy-induced azoospermia after treatment with vinblastine, bleomycin, and cisplatin for testicular cancer is usually reversible within 2 years of stopping treatment,  but cisplatin is thought to be the primary mediator of the toxicity.

Sulfasalazine and methotrexate can also cause oligospermia and infertility.

Environmental Toxins.

Because of the potential toxicity of physical and chemical agents, the occupational and recreational history should be carefully evaluated in all men with infertility.  Known environmental toxins include chemicals such as the nematocide dibromochloropropane and related compounds, ethylene glycol, cadmium, lead, and organic chloride compounds.  In a large meta-analysis of studies of normal men, sperm density was said to have declined from 113 million/mL in 1940 to 66 million/mL in 1990.  This report has subsequently been confirmed and refuted (reviewed in reference ). Environmental toxins that might act as estrogens or antiandrogens have been proposed as the cause. Cigarette smoking may also contribute to decreasing sperm density.

Autoimmunity.

Although autoimmunity may cause combined underandrogenization and infertility, the usual manifestation is isolated infertility, and antibodies to the basement membrane of the seminiferous tubules or to the sperm themselves may be responsible. Antisperm antibodies of the immunoglobulin A (IgA) class may prevent the penetration of cervical mucus by sperm.  IgG and IgA antibodies may impair the acrosome reaction and the binding of sperm to the zona pellucida of the oocyte.

Therapy usually involves in vitro fertilization (see later). Development of antisperm antibodies is not always a primary phenomenon because the antibodies have been identified in men with both bilateral and unilateral obstruction of the vas deferens and after vasectomy.

Defects Associated with Systemic Diseases

Temporary impairment of semen quality, particularly decreased sperm density, is common after acute febrile illness.  This is one reason that several semen analyses must be obtained in the work-up for men with infertility to be confident that true basal parameters have been determined (see earlier). Men with celiac disease have a distinct testicular abnormality, namely, endocrine features typical of androgen resistance with elevated plasma testosterone and LH levels.

Improvement in gluten enteropathy may reverse the androgen resistance-like state.

As discussed earlier, spinal cord injury is commonly associated with isolated infertility.

Androgen Resistance

Androgen resistance may cause infertility without underandrogenization (see earlier).  It was originally thought on the basis of functional studies that androgen receptor defects might account for 10% to 20% of idiopathic azoospermia or severe oligospermia, but loss-of-function mutations in the androgen receptor gene are present in only about 2% of such men.  Expansion of the CAG repeat sequence in the N-terminal region of the androgen receptor (see earlier) appears to be more common in men with azoospermia or severe oligospermia.

Impairment of Sperm Transport

Disorders of sperm transport may cause as much as 6% of male infertility.  Such disorders can be unilateral or bilateral, congenital or acquired. Infertility in men with unilateral obstruction may be due to antisperm antibodies. Obstructive azoospermia at the level of the epididymis also occurs in association with chronic infections of the paranasal sinuses and lungs.

In polycystic kidney disease, dilated cysts of the seminal vesicles may obstruct semen transport.  Tuberculosis, leprosy, and gonorrhea can obstruct the ejaculatory system, and sperm transport can be obstructed by deep midline müllerian cysts.  Congenital defects of the vas deferens can cause azoospermia or oligospermia in sons of women given diethylstilbestrol during pregnancy.

Congenital bilateral absence of the vas deferens is common in men with cystic fibrosis, and mutations in the gene responsible, the transmembrane conductor regulator (CFTR) gene, can cause bilateral absence of the vas deferens without other manifestations of the disease.  Congenital unilateral absence of the vas deferens may be an incomplete form of the bilateral disorder.  Thus, congenital absence of the vas deferens and cystic fibrosis are variable manifestations of mutations of the same gene.
Idiopathic Infertility

In large series, known causes were identified for only about 60% of cases of infertility in males, with the remainder classified as idiopathic (Table 18-5). Because at best only about half of fertile men with a varicocele achieve fertility after surgical repair, it is likely that even a larger fraction of infertile men have idiopathic infertility. Some may have androgen resistance, and as many as a fifth may have disorders involving the \( \text{AZF} \) gene (see earlier). Others have oligosperma or azoospermia with normal plasma LH and testosterone but elevated FSH levels in the absence of cryptorchism, radiation, or drug exposure.

Studies of such men indicate that isolated FSH elevation may be associated with a decreased GnRH pulse frequency and that FSH levels may be corrected with pulsatile GnRH therapy. Men with oligosperma and normal FSH levels may also have altered pulsatile secretion of gonadotropins and testosterone, and testosterone production rates are said to be low in selected infertile men with isolated FSH elevations and normal total serum testosterone. Whether these abnormalities are of causative significance is unknown.

A subset of men with severe idiopathic oligospermia associated with a decreased ratio of serum testosterone to estradiol have shown an increase in sperm density after treatment with an aromatase inhibitor. Testicular biopsies of men with idiopathic infertility have shown increased apoptosis associated with maturation arrest and hypospermatogenesis, decreased expression of the c-kit receptor, and increased mutations consistent with abnormal DNA repair mechanisms.

Management of Infertility

The management of infertility is usually unsatisfactory because the number of potentially correctable causes is small. When appropriate, however, associated hormonal disorders and coexisting medical conditions may be treated and offending drugs can be discontinued.

Although claims of success have been made for a variety of empirical therapies for infertility with oligospermia, most such claims fail to take into account the spontaneous fertility rate in untreated oligospermic men (25% in 1 year). The fact that treatment-independent pregnancy occurs in all forms of human infertility (male and female factors) makes it necessary for all therapies to be evaluated by randomized clinical trials. When several forms of empirical therapy—including testosterone rebound, nonaromatizable androgen (mesterolone), gonadotropin, antiestrogen (clomiphene), antibiotics, bromocriptine, varicocele repair, artificial insemination, and no therapy—were compared in one large retrospective analysis of oligospermic men, none were effective.

In Vitro Fertilization

The only effective empirical therapy for male infertility is in vitro fertilization. Standard techniques of in vitro fertilization require about 500,000 motile sperm/mL of ejaculate. Although the fertilizing capacity of sperm from men with abnormal sperm parameters is diminished, conventional techniques can result in 10% or more live births per attempt in men with mild to moderate abnormalities. Such rates are threefold to fivefold higher than natural impregnation rates in such men. However, standard in vitro fertilization is ineffective in men with more severe defects in spermatogenesis. In the Melbourne experience, fertilization rates were low in men with severe oligospermia, poor motility, and increased numbers of abnormal forms.

Intracytoplasmic Sperm Injection

Better results have been obtained with the development of intracytoplasmic sperm injection (ICSI)namely, fertility rates of 50% or more using poor quality semen, including decreased sperm number, impaired motility, increased abnormal forms, and combined defects. In men with obstructive azospermia in whom sperm must be aspirated from the epididymis, fertilization rates are nearly normal.

In the past, men with nonobstructive azoospermia, typically with elevated plasma FSH levels, were not treatable. Such patients include men with maturation arrest, postcryptorchidism tubular atrophy, mumps orchitis, and Klinefelter's syndrome. However, when minute amounts of sperm could be identified by testicular biopsy, fertilization and impregnation have been achieved with rare spermatooza or spermatids retrieved from such biopsy specimens using ICSI. ICSI should not be undertaken until men with abnormal semen undergo a complete work-up so that hypogonadotrophic hypogonadism or some other treatable condition is not missed. Furthermore, ICSI may increase the chances of transmitting the father's disorder to offspring, as has been reported in men with Klinefelter's syndrome and men with AZF mutations.
Old Age

The role of the decrease in total and bioavailable testosterone and the increase in estradiol in the decline of male sexual function with aging is not clear (see the following). However, this changing hormonal milieu may be involved in the pathogenesis of breast enlargement in elderly men and in the development of prostatic hyperplasia.

Prostatic Hyperplasia

Enlargement of the prostate to the extent that it obstructs urethral outflow is common in elderly men. The gland weighs a few grams at birth, and at puberty androgen-mediated growth causes the prostate to reach the adult size of about 20 g by age 20. This growth is accompanied by transformation of the cuboidal epithelium of the acini to a columnar, secretory epithelium and by initiation of secretion of the prostatic component of the ejaculate. The weight of the gland remains stable for about 25 years.

Beginning in the fifth decade of life, a hyperplastic phase of prostatic growth ensues in most men. The second growth phase, unlike the earlier growth, which involves the gland diffusely, typically begins in the periurethral region as a localized proliferation of glandular and stromal elements. The hyperplasia may remain limited in scope, but in many men growth continues and eventually compresses the remaining normal portion of the prostate. The progressive increase in gland size can cause lower urinary tract symptoms and urinary obstruction, but the correlation between symptoms and anatomic changes can be unpredictable. On the one hand, hyperplasia primarily of the periurethral region can obstruct urine outflow in the absence of gross prostatic enlargement; on the other hand, men with gross enlargement of the gland may be asymptomatic.

The second growth spurt, like the growth at puberty, requires a functioning testis. Dihydrotestosterone formed within the prostate from testosterone mediates the embryonic development, pubertal growth, and hyperplastic growth of the prostate. Administration to animals of a 5-reductase inhibitor to block dihydrotestosterone formation caused involution of the prostate despite elevation of testosterone levels within the gland. Furthermore, although plasma testosterone may decline with age, the level of dihydrotestosterone in the hyperplastic gland either remains constant or increases.

Prostatic hyperplasia also occurs in the aging male dog, and most research on its pathogenesis has been done in that species. Administration to the castrated dog of androgens that cause an increase in the prostatic dihydrotestosterone level caused prostatic enlargement comparable to that seen in the naturally occurring disorder. Estrogen acts synergistically with dihydrotestosterone to enhance prostatic growth in the dog because estrogen increases the amount of androgen receptor in the tissue. Thus, two hormones participate in the development of prostatic hyperplasia in the dog; dihydrotestosterone is responsible for prostate growth, and estradiol enhances dihydrotestosterone action.

Three types of evidence suggest that dihydrotestosterone and estradiol are also involved in human prostatic hyperplasia:

1. Either surgical or pharmacologic castration causes a decrease in the size of the hyperplastic prostate gland, indicating that continuing androgen action is essential to maintain the hyperplastic state.
2. Inhibition of prostatic 5-reductase with agents such as finasteride causes a profound decrease in prostatic dihydrotestosterone levels and a 20% to 30% decrease in prostate volume after 3 to 6 months of therapy. This effect is maintained for up to 4 years and is associated with few side effects.
3. There is a temporal relation between the development of prostatic hyperplasia and the increase in plasma estradiol with age, but a causal relation between estradiol and human prostatic hyperplasia has not been established.

Demonstration that hormones play a role in prostatic hyperplasia does not necessarily provide insight into its pathogenesis because their action could be permissive rather than causal, and the reason that the disorder varies so markedly in its manifestations is unclear. Similarly, the therapeutic role of 5-reductase inhibitors is not established because there is a strong placebo effect on urinary symptoms, there is no clear-cut relation between symptoms and urine flow, and the natural history of the disorder particularly how to predict which subset of men will develop significant obstruction is not understood. As a consequence, although a variety of minimally invasive or medical therapies are now available, the indications for surgical or medical management, compared with watchful waiting, are sometimes unclear.

Prostatic Cancer

The endocrine aspects of prostatic cancer are discussed in Chapter 39.
Disorders of All Ages

Testicular Tumors

Tumors of the testes occur with an incidence of 2 to 3 per 100,000 men per year in the United States and account for about 1% of cancer deaths in men. The incidence in most Western countries has risen since the 1930s, particularly in adults, but mortality rates have declined. The frequency shows a trimodal curve, with peaks in childhood (embryonal carcinomas and teratocarcinomas), young adulthood, and old age (seminomas). The incidence in blacks is a sixth or less that in whites, but overall the tumors are the second most common malignancy (after leukemia) in men between ages 20 and 35 years. The tumors are commonly bilateral (either simultaneous or sequential, e.g., a seminoma developing in one testis many years after the removal of the other).

Occurrence is familial in 1% to 2% of cases. The presence of an isochromosome of the short arm of chromosome 12 is characteristic of germ cell tumors of all subtypes.

Several factors predispose to testicular malignancy. Men with cryptorchidism have a fivefold increased risk of development of such tumors, men with intra-abdominal testes being more at risk than those with high inguinal testes, so that in one series 10 of 131 men with testicular cancer had antecedent maldescent. Three fourths of tumors associated with maldescent are seminomas, the remainder being other germ cell tumors.

Early orchiopexy facilitates detection, but whether it reduces the incidence of tumor development is not clear. Testicular malignancy may be more frequent in individuals with abnormal sexual development (i.e., 45,X/46,XY mixed gonadal dysgenesis or 46,XY testicular feminization) than with other forms of testicular maldescent. Occupational exposure to extremely high or low temperature can increase the risk. Estrogen administration to pregnant women may be a predisposing factor in male offspring, and Down's syndrome, Klinefelter's syndrome, and HIV infection are associated with an increased incidence.

**TABLE 18-6 - Classification of Testicular Tumors**

<table>
<thead>
<tr>
<th>I. Germ cell tumors (95%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Single-celltype tumors (60%)</td>
</tr>
<tr>
<td>1. Seminomas</td>
</tr>
<tr>
<td>2. Yolk sac tumors (embryonal cell tumors)</td>
</tr>
<tr>
<td>3. Teratomas</td>
</tr>
<tr>
<td>4. Choriocarcinoma</td>
</tr>
<tr>
<td>B. Combination tumors (40%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>II. Tumors of gonadal stroma (12%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Leydig cell</td>
</tr>
<tr>
<td>B. Sertoli cell</td>
</tr>
<tr>
<td>C. Primitive gonadal structures</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>III. Gonadoblastomas</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Germ cell + stroma cell</td>
</tr>
</tbody>
</table>


Most testicular tumors in men with congenital adrenal hyperplasia related to steroid 21-hydroxylase (CYP21) deficiency consist of adrenal cell rests, are dependent on cortisol for growth and secretion, and occur in men who are inadequately treated and hence have elevated plasma corticotropin levels. However, the tumors can be difficult to separate histologically from interstitial cell tumors.

**Diagnosis**

Most testicular cancers produce local symptoms, but delay in making a diagnosis is common because of oversights by both physicians and patients. Testicular cancers usually occur before age 45, and men should be educated about the need to seek prompt medical advice for any change in a previously normal testis, including enlargement, pain or a feeling of heaviness, swelling, or other unusual findings. Pain occurs in half of affected men, and to reduce delay physicians should consider any testicular mass to be a tumor until proved otherwise and to obtain surgical consultation if symptoms and signs persist.

**Classification**

The most widely used classification is that of Mostofi (Table 18-6) and is based on the cell type from which the tumor originates.

**Germ Cell Tumors**

Germ cell tumors are the most common types.

Seminomas are characterized by large cells with clear cytoplasm in a delicate fibrovascular stroma infiltrated with lymphocytes; the granulomatous reaction around the tumor can be so intense as to suggest a graft-versus-host reaction. These tumors account for at least half of all testicular neoplasms and can be subdivided into spermatocytic seminomas, which occur in older men and are associated with a 90% to 95% 5-year survival, and anaplastic seminomas, which have a poor prognosis.

Embyronal carcinomas are the most common testicular tumors in boys, resemble embryonal carcinomas of the ovary, and are associated with 5-year survivals of about 70% in infants and 25% in adults.

Choriocarcinomas contain syncytiotheliotrophic cells and usually occur in the second and third decades of life; the prognosis is poor.

Teratomas contain at least two germ cell layers and may be either benign or malignant; they are second in frequency to embryonal carcinomas in childhood and are unusual in adults.

Tumors that contain combinations of germ cell types account for 40% of germ cell tumors; the biology of such tumors is determined by the least differentiated (most malignant) element. Of mixed tumors that contain cells of germinal and stromal origin, perhaps the most distinctive is the gonadoblastoma, which consists of germ
cells, sex cords, and, usually, Leydig cells. Gonadoblastomas usually occur in dysgenetic testes containing a Y chromosome and synthesize androgen.

Germ cell tumors of all types can also originate in extragonadal sites, including the mediastinum and the brain. These extragonadal tumors are presumed to arise from aberrant migration of germ cells early in embryogenesis; from some common precursor stem cell line that normally gives rise to germ cells, thymus, and pineal gland; or from migration of transformed gonadal germ cells.

Testicular germ cell tumors usually occur as a nodule or painless swelling of the testis but may be identified as the result of metastases or because of the peripheral manifestations of hCG secretion by the tumor. After diagnosis, the tumors are staged either by surgical exploration or by computed tomographic scanning or magnetic resonance imaging. Stage I is limited to the testes, stage II involves metastases to infradiaphragmatic lymph nodes but not beyond, stage III involves supradiaphragmatic lymph nodes, and stage IV involves extralymphatic metastases.

Germinomas can secrete several distinct tumor cell markers into plasma, including hCG and its subunit, -fetoprotein, lactate dehydrogenase, carcinoembryonic antigen, and placental alkaline phosphatase. Virtually all germ cell tumors synthesize hCG and its subunits, but the hormone is secreted in large amounts only by some nonseminoma germ cell tumors (choriocarcinomas, teratocarcinomas, and yolk sac tumors). -Fetoprotein is a marker of tumors containing yolk sac elements, and teratomas can secrete carcinoembryonic antigen. An elevated level of one of these tumor markers in the plasma of a patient whose tumor has been classified as a pure seminoma usually indicates that the tumor is actually a combination tumor. These markers are particularly useful for following the response to therapy. Secreted hCG may be endocrinologically active and cause enhanced formation of testosterone and, more important, of estradiol by the testes. The net result can be a feminizing syndrome and inhibition of the secretion of LH and FSH by the pituitary (see later).

The treatment of germ cell tumors constitutes a major triumph of cancer therapy. Appropriate therapeutic strategies include debulking of the tumor mass, resection of involved lymph nodes, administration of chemotherapy (usually combinations of cisplatin, vinblastine, etoposide, and bleomycin), radiation, and monitoring of tumor cell markers. The cure rates for patients with seminomas are approximately 90% for stage I disease, and individuals with stage III nonseminoma tumors, which were previously uniformly lethal, now have good survival rates.

Because young men with germ cell tumors may have infertility related to castration, radiation, chemotherapy, or a combination, cryopreservation of semen before treatment has been advocated as a means of preserving fertility, but many men have adequate sperm production after chemotherapy. Treatment is associated with a small risk of recurrence and the development of secondary solid tumors and leukemia and with the late sequelae of chemotherapy such as nephrotoxicity and neurotoxicity.

Stromal Cell Tumors

Stromal tumors (Leydig cell tumors, Sertoli cell tumors) account for 1% to 2% of testicular tumors, and both cell types may coexist within the same tumor. Rarely, adrenal rest tumors occur in the testes. As would be expected, Leydig cell tumors commonly secrete testosterone and thus may cause virilization in prepubertal boys (precocious pseudopuberty); many of the tumors secrete estradiol as well and cause mixed signs of feminization and virilization during the prepubertal years and feminizing signs in adult men. The hormones from such tumors can suppress levels of endogenous gonadotropins and testosterone and can cause azoospermia and decreased size of the contralateral testis. Because the tumors may be so small as to be recognized only by ultrasonography, documenting that the testis is the site of increased estrogen production may require selective catheterization of the testicular veins.

Sertoli cell tumors show a bimodal age distribution, most patients being younger than 1 year or between ages 20 and 45 years. The tumors are frequently bilateral and familial (usually as a component of the Peutz-Jeghers syndrome). Gynecomastia occurs in about 25% of patients, and estrogen secretion can impair spermatogenesis and cause shrinkage of the contralateral testis. Leydig cell hyperplasia can occur in the area around the tumor, implying either that the tumor is of mixed cell origin or that Sertoli cells secrete some factor that stimulates Leydig cell development. Complete cure and regression of feminizing signs usually follow surgical resection. Approximately 10% of stromal tumors are malignant.

Rate Tissue Tumors

Adenocarcinoma of the rete testis is rare but tends to be highly malignant.

Summary

Testicular tumors can enhance production of estradiol and testosterone by more than one mechanism. When tumors produce steroid hormones autonomously, plasma gonadotropin levels and androgen secretion by uninvolved portions of the testes are depressed, and azoospermia is common. When hCG is secreted by the tumor, production of estradiol and testosterone is increased in unaffected areas of the testes, and azoospermia is uncommon. Furthermore, occasional chorionicarcinomas that cannot synthesize steroids de novo nevertheless convert circulating androgens to estrogens. When hormones are formed directly or indirectly by the tumors, the response varies depending on the pattern of hormones produced and the age of the subject.
ABNORMALITIES IN ESTROGEN METABOLISM

Gynecomastia

Administration of large amounts of estrogen to men, as for carcinoma of the prostate or in preparation for sex change surgery, causes a variety of side effects, including fluid retention and congestive heart failure, hypertension, electrocardiographic changes, myocardial infarction, and thromboembolic disease. At lower levels, as can occur with estrogen-secreting testicular tumors, estrogen excess suppresses gonadotropin secretion, secondarily impairs testosterone production, and inhibits spermatogenesis. However, the most common manifestation of estrogen excess in men is gynecomastia (breast enlargement).

Sexual dimorphism in breast development at the time of puberty is due to the ovarian secretion of estrogen, and estrogen excess at any stage of life can cause breast enlargement in men. In the absence of a progestagen, the breast acini and lobules do not undergo complete female development, probably explaining why galactorrhea is unusual in men.

Clinical Features

At the clinical level, gynecomastia is complicated by problems of definition. The common view has been that any palpable breast tissue in men is abnormal except for three situations: transient gynecomastia of the newborn, pubertal gynecomastia, and gynecomastia that occasionally occurs in elderly men. However, this view was challenged by Nuttall and Niewoehner and Nuttall, who reported that 36% of normal adult men and two thirds of hospitalized men have palpable breast tissue. The prevalence may have increased because of some unrecognized cause.

A confounding problem is that it can be difficult to distinguish enlargement of breast tissue from lipomastia, in which enlargement is caused by adipose tissue. True gynecomastia can be separated from lipomastia by mammography or sonography. Autopsy data are not of much help in establishing the frequency of gynecomastia because they do not provide information about what fraction of gynecomastialactive or inactivestheoretically palpable. For the purposes of this discussion, breast enlargement in men (other than in the three so-called physiologic states) may be indicative of an underlying endocrinopathy and deserves at least a limited evaluation.

Histopathology and Etiology

Gynecomastia is frequently asymmetrical, and unilateral gynecomastia can be temporary in that one breast may enlarge months or years before the other. The process begins with proliferation of the stroma and the duct system, which elongates, buds, and duplicates. With time, progressive fibrosis and hyalinization are accompanied by regression of epithelial cells so that the ducts decrease in number. On correction of the cause, resolution involves reduction in size and cell content of the epithelium followed by gradual disappearance of the ducts, leaving hyaline bands that may persist or eventually disappear.

Gynecomastia is generally viewed as the consequence of absolute or relative estrogen excess, and it can be classified as either physiologic or pathologic (Table 18-7).

Physiologic Gynecomastia

During three phases of male life, breast enlargement can be a normal finding.

Gynecomastia in the Newborn

Enlargement of the breast in the newborn is probably due to maternal or placental estrogens, or both. The swelling may or may not be associated with milk production and usually disappears in a few weeks but can persist longer.

Adolescent Gynecomastia

Transient enlargement of the breast occurs in about 40% of adolescent boys. The median age at onset is 14 years, the breasts may be asymmetrical and tender, and by age 20 only a small number of men have palpable vestiges of gynecomastia. The most severe disorder, termed pubertal gynecomastia (breast tissue > 4 cm), may persist into adulthood and is more commonly associated with an underlying endocrinopathy.

The cause of pubertal breast enlargement is uncertain. Plasma estradiol levels in boys normally reach the adult range before plasma testosterone, and by age 20 only a small number of men have palpable vestiges of gynecomastia. The most severe disorder, termed pubertal gynecomastia, may persist into adulthood and is more commonly associated with an underlying endocrinopathy.

Pathologic Gynecomastia

Relative estrogen excess

Congenital defects

Congenital anorchia

Klinefelter’s syndrome

Androgen resistance (testicular feminization and Reifenstein’s syndrome)

Defects in testosterone synthesis

Secondary testicular failure (viral orchitis, trauma, castration, neurologic and granulomatous diseases, renal failure)

Increased estrogen production

Increased testicular estrogen secretion

Testicular tumors

Bronchogenic carcinoma and other tumors producing hCG
Women with complete testicular feminization and men with Reifenstein's syndrome have normal or elevated production of estrogen (hypospadias and gynecomastia) or men with undervirilization or infertility, or both. Phenotypic women with testicular feminization. If the impairment of receptor function is less complete, the phenotype is that of men with Reifenstein's syndrome and male testosterone levels but who are resistant to their own and to exogenous androgens.

Hereditary defects in the X-linked gene that encodes the androgen receptor cause a spectrum of syndromes of incomplete virilization in 46,XY men who have testes. Estrogen/androgen ratios.

Serum testosterone and estrogen levels begin to decline as plasma LH, which also causes enhanced estradiol secretion by the testes. Testicular function becomes progressively impaired with time, so that after age 15 years, the causes of elevated plasma estradiol are complex.

Pathologic Gynecomastia

Pathologic gynecomastia can be due to a relative (as in testosterone deficiency) or absolute increase in estrogen formation (as in Leydig cell tumors), to drugs, or to unknown causes.

Relative Estrogen Excess

Failure of testosterone synthesis or action causes elevated plasma gonadotropin levels, and relative estrogen excess ensues because of the extraglandular aromatization of adrenal androgens and on occasion a secondary increase in testicular estrogen secretion. Plasma FSH and LH levels are high, and the average plasma testosterone level is half normal, although some have normal testosterone levels. Variations in plasma levels of testosterone and estradiol are associated with variable degrees of androgenization and feminization in the disorder.

Congenital Defects

Congenital Anorchia.

Congenital anorchia is a rare, often familial disorder in which the testes are missing in phenotypically normal 46,XY males. Affected individuals are thought to have bilateral cryptorchidism at birth, but no tests are located on surgical exploration of the abdomen. Because testicular hormones are necessary for male phenotypic development and because the penis is normal in this disorder, it is believed that testes are present and function normally until late in embryonic life and then regress for unknown reasons. Approximately half of anorchid men develop gynecomastia.

In some anorchid men, Leydig cells secrete small amounts of testosterone into the circulation even if testes cannot be found at surgery. Other men with congenital anorchia have profound testosterone deficiency and a small amount of estradiol formed by the indirect pathway androstenedione estrone estradiol in extraglandular aromatization of adrenal androgens and on occasion a secondary increase in testicular estrogen secretion. These findings imply that the critical factor for feminization is not the absolute level of estrogen but rather some ratio of testosterone to estradiol to induce breast enlargement in the absence of disease.

Other men with congenital anorchia have profound testosterone deficiency and a small amount of estradiol formed by the indirect pathway. Phenytoin, theophylline, diazepam, marijuana, heroin) appear to block estrogen action by competing with estradiol for binding to the estrogen receptor.

Klinefelter's Syndrome

Approximately half of nonmosaic and a third of a mosaic men with Klinefelter's syndrome have gynecomastia after the expected time of puberty. Plasma FSH and LH levels are high, and the average plasma testosterone level is half normal, although some have normal testosterone levels. Variations in plasma levels of testosterone and estradiol are associated with variable degrees of androgenization and feminization in the disorder.

The causes of elevated plasma estradiol are complex. Early in adolescence, plasma testosterone is usually in the normal male range as the result of elevated plasma LH, which also causes enhanced estradiol secretion by the testes. Testicular function becomes progressively impaired with time, so that after age 15 years, serum testosterone and estrogen levels begin to decline and the end stage resembles anorchia (see earlier). Diminished estrogen clearance may further increase estrogen/androgen ratios.

Androgen Resistance (Testicular Feminization and Reifenstein's Syndrome)

Hereditary defects in the X-linked gene that encodes the androgen receptor cause a spectrum of syndromes of incomplete virilization in 46,XY men who have testes and male testosterone levels but who are resistant to their own and to exogenous androgens. In the most severe form, affected individuals are phenotypic women with testicular feminization. If the impairment of receptor function is less complete, the phenotype is that of men with Reifenstein's syndrome (hypospadias and gynecomastia) or men with undervirilization or infertility, or both.

Women with complete testicular feminization and men with Reifenstein's syndrome have normal or elevated production
rates for testosterone and estradiol, presumably because of increased secretion by the testes in response to elevated plasma gonadotropin levels. FSH and LH levels are elevated because of resistance at the hypothalamic-pituitary level to negative-feedback control by testosterone. However, there is no direct relation between the rates of estrogen secretion in these disorders and the degree of feminization that results, probably because the degree of feminization is influenced by other factors such as the severity of the androgen resistance and the variable elevation of plasma androgen levels.

Defects in Testosterone Synthesis.

Five inherited defects impair testosterone synthesis and prevent normal virilization of the male embryo. Each of the defects involves a critical biochemical step in the conversion of cholesterol to testosterone. The completeness of the defects and the severity of clinical manifestations vary, but feminization is common in two of the disorders, 3-HSDII deficiency and 17-HSDIII deficiency.

Feminization in these disorders can arise from more than one mechanism. For example, normal or low levels of plasma estrogen can cause feminization in the presence of diminished androgen production, analogous to the situation in congenital anorchia. Alternatively, estrogen production may be increased because of increased availability for extragonadal aromatization of steroids such as androstenedione that accumulate proximal to the enzymatic block. Partial deficiency of 17-HSDIII and late-onset 3-HSDII deficiency are rare causes of gynecomastia in otherwise phenotypically normal men.

Testicular Failure

Viral orchitis is the most common cause of testicular failure after puberty, and mumps is the most frequent etiology (see earlier). In men with gynecomastia and bilateral testicular atrophy related to orchitis, testosterone production is severely impaired, whereas production of estradiol and estrone is normal, arising almost entirely from extraglandular sources. 

The second most common cause of acquired testicular atrophy in the adult is trauma, and gynecomastia can result. Trauma to the testes can be associated with elevated levels of plasma estradiol many years later. Neurologic disease, including myotonic dystrophy and spinal cord injury, can also cause testicular atrophy. 

Testicular atrophy, decreased plasma testosterone levels, elevated gonadotropin levels, and gynecomastia are also common in proximate block.

Gynecomastia is present in approximately half of men undergoing hemodialysis for renal failure. Plasma LH and FSH levels are elevated, the plasma testosterone level is low, and plasma prolactin levels are elevated. Estradiol levels may also be high (see earlier).

Increased Estrogen Production

Estrogen production in men can increase because of (1) increased testicular secretion, (2) increased availability of substrate for extraglandular formation, or (3) increased activity of extraglandular aromatase itself.

Increased Testicular Estrogen Secretion (see Fig. 18-18 B)

Testicular Tumors.

Testicular tumors can feminize in three ways.

First, germinal cell tumors (embryonal carcinomas, choriocarcinomas, teratomas, and rarely seminomas) can produce hCG or fragments of hCG, which can act in uninvolved areas of the testes to stimulate the synthesis of estradiol and testosterone, which in turn suppress plasma LH and FSH.

Second, stromal cell tumors (Leydig and Sertoli cell tumors) can secrete testosterone and estradiol autonomously. About 20% of men with Leydig cell tumors have gynecomastia, and gynecomastia may be even more common with Sertoli cell tumors. 

Feminization can occur before such tumors are detectable by physical examination, but even small tumors can usually be identified by ultrasonography. Similarly, in choriocarcinomas and in hepatocellular carcinomas, aromatase in the tumor tissue can convert circulating adrenal and testicular androgens to estrogens.

Third, some Sertoli cell tumors stimulate adjacent Leydig cells to secrete androgens that serve as substrate for aromatase in the tumor cells. (See earlier for discussion of diagnosis and management of testicular tumors.)

Bronchogenic Carcinoma.

Lung cancer can cause an increase in hCG levels in plasma, and gynecomastia in this condition correlates with the amount of estradiol secreted by the testes. Indeed, hCG secretion by any tumor, such as by transitional cell tumors of the urinary tract, can cause feminization.

True Hermaphroditism.

In true hermaphroditism (see Chapter 22), both the ovarian and the testicular components of the gonads are endocrinologically active and cause a mixed pattern of feminization and virilization at puberty. Gynecomastia is due to gonadal estrogen secretion (see Fig. 18-18 B), presumably by the ovarian elements of the ovotestes.

Increased Substrate for Peripheral Aromatase (see Fig. 18-18 C)

Adrenal Disease.

In feminizing adrenal carcinoma, estrogen production is usually due to massive increases in the levels of the adrenal androgens androstenedione and DHEA, which serve as substrates for extraglandular aromatization. In rare instances, adrenal tumors secrete estrogen. 

Feminization in boys with congenital adrenal hyperplasia (as in CYP21 or CYP11A2 deficiency) is usually the consequence of increased production of androstenedione by the adrenal glands and hence of increased substrate for peripheral aromatase. In some instances, decreased testosterone levels may play a role in the gynecomastia. Increased androstenedione is also the usual cause of feminization in men with 17-HSDIII deficiency.

Androgen Administration.

Administration of testosterone to children commonly causes gynecomastia, correlating with an increase in estrogens, whereas replacement with conventional doses in adult men increases plasma estradiol but rarely causes gynecomastia. In contrast, testosterone administration to men with impaired liver function can cause
profound increases in plasma estradiol levels. In addition, administration of supraphysiologic amounts of aromatizable androgens can increase estradiol levels as much as sevenfold in normal men, and gynecomastia is common in users of anabolic steroids. In probing patients’ histories for possible causes of gynecomastia, it should be remembered that some androgens are not aromatizable or are weak substrates for the enzyme (see the following).

Liver Disease.

Liver disease is a common cause of feminization. Gynecomastia is thought to be largely a result of over-production of estrone. However, the liver is not the direct source of the estrogen, which are mainly due to decreased hepatic catalysis of androstenedione and the consequent increased availability of androstenedione for extrahepatic aromatization. In carcinoma of the liver, feminization can be the consequence of increased aromatase activity in the tumor itself.

Starvation.

Gynecomastia was common in American prisoners of war during World War II. About a third of the cases occurred during refeeding after release, other instances were associated with temporary improvements in the food supply during imprisonment, and most regressed within a few months. The pathophysiology of starvation gynecomastia may be similar to that with liver disease (see earlier).

Thyrotoxicosis.

Thyrotoxicosis can cause gynecomastia. Elevation of plasma estradiol levels is probably due to increased androstenedione production rates and increased formation of estrogen in extraglandular sites.

Increase in Extraglandular Aromatization (See Fig. 16-13B)

Increased activity of aromatase enzymes in peripheral tissues can increase estrogen production as much as 50-fold, and in at least one family the trait appeared to be inherited in an autosomal dominant pattern through three generations, being manifested in females by precocious puberty and macromastia and in males by gynecomastia. A characteristic feature is that the onset of gynecomastia correlates with the onset of adrenarche and occurs before the time of normal puberty. A similar trait is present in the Sebright bantam chicken, in which an autosomal dominant gene increases extraglandular aromatization more than 100-fold.

Drugs

Drugs can cause gynecomastia by direct action as estrogens, by enhancement of testicular production of estrogens, by inhibition of testosterone synthesis or action, or by unknown mechanisms.

Estrogens and Estrogen Mimetics

Estrogens given to men in any form can cause gynecomastia, as in the treatment of prostatic cancer with diethylstilbestrol and of transsexual men with estrogens. Young men and boys are particularly sensitive to estrogens, and gynecomastia can develop as a result of industrial exposure or dermalointments containing estrogens. Identifying the source may require a high index of suspicion, as in the case of a barber who massaged the scalps of customers with ointment containing estrogen, factory workers who manufacture oral contraceptives, children of workers in a diethylstilbestrol manufacturing plant who absorbed the drug from the clothing of their fathers,

and offspring of women who use topical estrogen preparations. Sufficient estrogen to induce gynecomastia can be absorbed by men during sexual intercourse with partners who use vaginal creams containing estrogens. In the United States, no federal regulations cover estrogens in cosmetics, and estradiol levels may be as high as 18 ng/ml in creams and 50 mg/dl in lotions.

Epidemics of gynecomastia among children have resulted from the ingestion of milk or meat from estrogen-treated cows, raising the possibility that long-term exposure to small amounts of estrogens may be a cause of idiopathic gynecomastia. Sources may include meat and dairy products from animals treated with estrogens other than diethylstilbestrol, endogenous estrogens in animal tissues, or plant or fungal estrogens in foods.

About 10% of men given digitalis for a year had gynecomastia; however, abnormal liver function is common in such men, and gynecomastia is said to correlate better with congestive heart failure than with administration of digitalis. Nevertheless, digitalis preparations associated with gynecomastia also have estrogenic effects on the vaginal epithelium in postmenopausal women. Digitalis binds to the estrogen receptor and may act as a direct estrogen agonist.

Drugs That Enhance Endogenous Estrogen Formation

Administration of hCG can cause gynecomastia as a consequence of increased estradiol secretion by the testes. Clomiphene citrate (both an estrogen agonist and antagonist) has been used to treat gynecomastia in boys, but paradoxically it can cause gynecomastia on withdrawal, presumably by increasing LH secretion and consequently increasing estradiol secretion by the testes.

Drugs That Inhibit Testosterone Synthesis or Action

The antifungal drug ketoconazole and other imidazoles block steroid hormone synthesis.

Gynecomastia is presumably due to altered ratios of estradiol to testosterone.

Antineoplastic agents can cause long-term impairment of testosterone synthesis, presumably through toxic effects on Leydig cells; such damage may occur when the therapy is for systemic neoplasms (e.g., alkylating agents for Hodgkin’s disease) or for testicular cancers. The cause of gynecomastia has not been elucidated, but it may be due to elevated plasma gonadotropin levels secondary to testicular damage and enhancement of testicular estrogen synthesis.

Gynecomastia is common in men treated with spironolactone. At low doses, the drug prevents the binding of androgen to its receptor, and at high dose it inhibits testosterone synthesis.

Antiangiogens, including cyproterone, flutamide, zanoterone, and bicalutamide, inhibit testosterone binding to the receptor and can cause gynecomastia.

Gynecomastia is a common side effect of treatment with cimetidine, which also blocks the binding of androgen to the androgen receptor and may inhibit the catabolism of estradiol. Gynecomastia is less common in subjects receiving ranitidine. Suggestive evidence for induction of gynecomastia by an environmental antiandrogen has come from studies of an epidemic of temporary gynecomastia that affected Haitian refugees in five detention centers in the United States in 1981. The detouring agent used in these centers binds to the androgen receptor and acts as an antiandrogen in rats. All antiandrogens are believed to impair the feedback control of gonadotropin production and cause elevation of plasma LH, which in turn increases estradiol secretion from the testes.

Drugs That Act by Unknown Mechanisms

A variety of drugs cause gynecomastia by unknown mechanisms. For example, gynecomastia occurred in boys and in men given human growth hormone. Many drugs are associated with gynecomastia with a frequency that is probably not coincidental; these include busulfan, calcium channel-blocking agents, angiotensin-converting enzyme inhibitors, diazepam, isoniazid, methyl dopa, omeprazole, penicillamine, tricyclic antidepressants, and a variety of antimicrobial agents, particularly protease inhibitors used for the treatment of HIV. Some of these agents may act by altering liver function. Both marijuana and
A careful drug history that encompasses potential environmental and indirect exposures to endocrine substances.

A limited endocrine work-up, including (a) measurement of plasma DHEAS or urinary 17-ketosteroids (usually elevated in adrenal feminizing states), (b)

A detailed physical examination including the testes (the finding of small testes bilaterally suggests testicular insufficiency, and asymmetrical testes raise the

The extent to which minor endocrine disorders are not recognized with current methodologies is uncertain. The fact that gynecomastia can develop as the result of subtle environmental exposure to estrogens or antiandrogens (as described earlier) raises the possibility that a large fraction of idiopathic gynecomastia may be the consequence of unrecognized exposure to endocrine disruptors. The critical clinical point is that, whatever the cause, the diagnosis of idiopathic gynecomastia carries no known import related to health.

Lack of Role of Prolactin in Gynecomastia

Plasma prolactin levels are usually normal in men with gynecomastia of diverse causes, and men who have prolonged elevation in plasma prolactin secondary to use of psychotropic drugs do not commonly have gynecomastia. Consequently, prolactin is not believed to play a direct role in the disorder. This conclusion is in keeping with the fact that prolactin is not a growth hormone for the breast. Furthermore, when gynecomastia develops in men with prolactin-secreting tumors of the pituitary gland and high plasma prolactin levels or in men taking psychotropic agents, the gynecomastia is probably the consequence of secondary testicular failure as a result either of the effects of the tumor mass or of inhibition of LH secretion by prolactin.

Diagnosis

The dilemma is to distinguish men with significant endocrine disease from those with idiopathic gynecomastia. In general, only men with symptomatic gynecomastia are evaluated; however, if there is a question about whether the gynecomastia is real, the issue is best solved by mammography or ultrasonography.

Most of the known causes of gynecomastia can be identified by a work-up that includes the following:

1. A careful drug history that encompasses potential environmental and indirect exposures to endocrine substances.
2. A detailed physical examination including the testes (the finding of small testes bilaterally suggests testicular insufficiency, and asymmetrical testes raise the possibility of testicular tumors).
3. Evaluation of liver function.
4. A limited endocrine work-up, including (a) measurement of plasma DHEAS or urinary 17-ketosteroids (usually elevated in adrenal feminizing states), (b) measurement of plasma estradiol (helpful if elevated but usually normal), (c) assessment of plasma hCG (sometimes elevated with testicular tumors), and (d) measurement of plasma LH and testosterone.

If these parameters are normal, as is frequently the case, the usual recourse is to observe the patient without treatment. If the symptoms persist or worsen and if the enlargement is progressive, a more extensive evaluation may have to be undertaken.

Treatment

The difficulty in treating gynecomastia is inherent in its natural history. If the feminizing process persists for a long period, the initial glandular hyperplasia is replaced by progressive fibrosis and hyalinization that do not regress after the source of excess estrogen is corrected. Consequently, surgery remains the mainstay of therapy and is frequently indicated for psychological and cosmetic reasons. Such surgery is usually accomplished through a circumareolar approach.

Medical management is most successful when it is addressed to gynecomastia of recent onset or to prevention of its development. Testosterone administration has inconsistent effects in Klinefelter’s syndrome but can cause dramatic improvement in other forms of testicular failure (e.g., anorchia, viral orchitis).

Several drugs have been tried for gynecomastia, including the antiestrogens tamoxifen and clomiphene, the aromatase inhibitor testolactone, and danazol, a weak androgen that inhibits gonadotropin secretion. In one study, tamoxifen was about twice as effective in treating idiopathic gynecomastia as danazol; in another study, tamoxifen was uniformly effective in treating the gynecomastia induced by antiandrogen treatment in men with prostatic carcinoma. Treatment with dihydrotestosterone (which cannot be aromatized to estrogen) was also reported to provide symptomatic improvement.

Perhaps the most effective therapy for gynecomastia is to prevent its development by radiating the breasts before the institution of estrogen therapy in men with prostatic carcinoma or of antiandrogen therapy in male sex offenders. This therapy is about 90% effective, and the complication rate is low.
Impairment of Estrogen Formation or Action

The study of men with single-gene mutations that impair estrogen formation or action has provided insight into the role of estrogen in male physiology. These forms of estrogen deficiency are rare, but the fact that the phenotypes in the two disorders are similar establishes the importance of this role.

Aromatase Deficiency

Aromatase deficiency is the consequence of autosomal recessive loss-of-function mutations in the CYP19 gene. In the two reported men with this disorder, childhood development was considered normal but skeletal growth continued into the 20s despite pubertal maturation and resulted in tall stature. This growth pattern was associated with failure of epiphyseal closure, marked delay in bone age, and osteopenia.

One man had undetectable estrogen in plasma and elevated levels of testosterone, dihydrotestosterone, and gonadotropins; testicular volume was normal, and semen analysis was declined. The other affected man was evaluated because of tall stature, infertility, and skeletal pain associated with severe osteopenia; the testicular volume was 8 mL bilaterally, the sperm density was very low with many immotile sperm, and testicular biopsy revealed a maturation arrest at the spermatocyte stage. The etiology of the infertility in the latter patient was not clear because a brother was infertile in the presence of a normal CYP19 gene, suggesting that the infertility in this family may be due to some other disorder. The man responded dramatically to estradiol therapy with an increase in bone density to the normal range, resolution of bone pain, and lowering of elevated levels of total and LDL cholesterol and triglycerides.

These men had different homozygous missense mutations in the CYP19 gene that caused single amino acid substitutions and resulted in the formation of mutant enzymes with 0.2% to 0.4% of the activity of normal enzyme.

Estrogen Receptor Deficiency

An autosomal recessive, loss-of-function mutation in the estrogen receptor gene has been described in a man with tall stature, unfused epiphyses, osteopenia, and acanthosis nigricans. Virilization was normal, and he had a normal level of plasma testosterone. Plasma levels of estradiol, estrone, FSH, and LH were elevated, and semen analysis revealed a normal sperm density but decreased sperm motility. Serum lipoprotein levels were normal, but the presence of hyperinsulinemia and impaired glucose tolerance indicated insulin resistance. He did not respond to treatment with high-dose, transdermal estradiol sufficient to raise the plasma free estradiol 10-fold above normal, as indicated by no change in plasma gonadotropins or in bone density after 6 months. He was homozygous for a missense mutation in exon 2 of the estrogen receptor gene that resulted in a premature termination codon and hence precluded the formation of functional receptor.

In summary, the evidence from these rare disorders of estrogen formation and action indicates that estrogen plays a major role in controlling skeletal maturation and both the accrual and maintenance of bone mass in men. In both conditions, there was no pubertal spurt in growth; growth was instead steady and continued in association with failure of epiphyseal closure. Despite testosterone levels that were normal or increased, gonadotropin levels were elevated. These findings are in accord with studies of aromatase inhibitors, which indicate that estrogens are important for feedback control of gonadotropin secretion at the level of the pituitary and the hypothalamus. Abnormalities of carbohydrate and lipid metabolism in these patients appear to be inconsistent. The fact that gender identity and gender role behavior are male in both conditions indicates that estrogen does not play a critical role on these parameters.
HORMONAL THERAPY

Androgen Therapy

When administered by mouth, testosterone is absorbed into the portal blood and degraded by the liver so that only a small fraction reaches the systemic circulation. Parenterally injected testosterone is also rapidly absorbed and degraded. As a consequence, effective androgen therapy requires either administration of testosterone in a slow-release form (transdermal or micronized oral preparations) or administration of chemically modified analogues. Such chemical alterations either retard absorption or catabolism to maintain effective blood levels or enhance the androgenic potency of each molecule so that physiologic effects can be achieved at a lower plasma level of drug.

Three general types of modification of testosterone are clinically useful:

1. Esterification of the 17-hydroxyl group (type A).
2. Alkylation at the 17 position (type B).
3. Modification of the A, B, or C rings, particularly substitutions at the 1, 2, 9, and 11 carbons (type C) (Fig. 18-19).

Most agents actually contain combinations of ring structure alterations and either 17 alkylation or esterification of the 17-hydroxyl group.

Esterification of testosterone with various carboxylic acids decreases the polarity of the steroid, makes it more soluble in the fat vehicles that are used for injection, and hence slows release of the injected steroid into the circulation. The esters of 19-nortestosterone have particularly slow release and turnover rates. The longer the carbon chain in the ester, the more fat soluble the steroid becomes and hence the more prolonged the action. For example, testosterone propionate must be injected daily, whereas testosterone cypionate and testosterone enanthate can be administered every 2 or 3 weeks. Even more slowly hydrolyzed esters are under investigation.

![Figure 18-19 Types of androgen preparations available for clinical use. Type A derivatives are esterified in the 17 position. Type B steroids have alkyl substitutions in the 17 position. Type C derivatives involve a variety of alterations of ring structure that enhance activity, impede catabolism, or influence both functions. Most androgen preparations involve combinations of type AC or type BC changes.](Image)

Figure 18-19

such as testosterone buciclate, which is administered every 12 weeks, and testosterone undecanoate, which is administered every 6 weeks. Testosterone cypionate or enanthate was for many years the treatment of choice for male hypogonadism.

Although testosterone esters can be detected in plasma, they must be hydrolyzed before the hormone acts so that the effectiveness of therapy can be monitored by assaying the plasma level of testosterone after administration. Most esters must be injected, but twomethanolone acetate and testosterone undecanoate can be administered by mouth. Testosterone undecanoate is absorbed through the lymphatic system into the systemic circulation, and physiologic blood levels of testosterone can be achieved at doses of approximately 120 mg/day. Because of rapid turnover, testosterone undecanoate must be administered two to three times a day. The reason for the oral effectiveness of methenolone acetate (and of mesterolone) is not entirely clear.

The use of transdermal testosterone formulations makes it possible to sustain serum testosterone levels in the normal male range while avoiding the necessity for parenteral administration. These formulations include a scrotal patch, two nonscrotal patches, and a gel. The scrotal patch Testoderm is designed to deliver either 4 or 6 mg of testosterone over 24 hours and takes advantage of the fact that absorption across the scrotal skin is efficient in the absence of permeation enhancers. After application in the morning, serum levels peak in 2 to 3 hours and are maintained throughout the day.

The nonscrotal patches, Androderm and Testoderm TTS, differ in recommended application sites and times to peak serum levels, but both provide physiologic testosterone levels throughout the day. Androderm, available in 2.5- and 5-mg doses, is applied at bedtime; peak levels are achieved in 8 hours, and the application site must be rotated to avoid skin irritation. Testoderm TTS delivers 5 mg of testosterone from a larger surface area, is applied in the morning, results in a maximal serum level in 2 to 3 hours, and appears to cause minimal skin irritation. In a randomized comparative study, Androderm therapy did not cause the temporary supraphysiologic levels of estradiol and of total and bioavailable testosterone that occur after injections of testosterone enanthate. The transdermal preparation resulted in more normal testosterone levels, less frequent suppression of plasma LH, and less frequent elevation of plasma hemoglobin levels.

A transdermal 1% testosterone gel has been developed for the application of 50 to 100 mg of testosterone to the shoulders or abdomen each morning, the usual dose being 50 mg. The absorption of testosterone appears to be largely independent of the surface area to which it is applied, and steady-state serum levels are achieved within a few days. The hands must be washed carefully after each application to avoid inadvertent transmission of the hormone, but the application site should not be washed for 6 hours to maintain absorption efficiency. Application site skin-to-skin transfer may occur with close physical contact. In comparison with the 5-mg nonscrotal patch, 50 mg of testosterone gel causes somewhat higher serum testosterone levels.

Administration of a 5-mg preparation of testosterone cypoxodrin sublingually three times a day also results in normal plasma testosterone levels in hypogonadal men. 17-Alkylated androgens, such as methyltestosterone and methandienolone, are effective by mouth because alkylated steroids are absorbed into the portal circulation, are slowly catabolized by the liver, and reach the systemic circulation in effective amounts. For this reason, 17-methyl or 17-ethyl substitution is present in most orally active androgens. Because 17-alkylated androgens are believed to act within the cell as such (i.e., the alkyl groups are not removed), because assays are not routinely available for monitoring blood levels, and because they can cause abnormal liver function, these steroids have a limited role in medicine.

Other alterations of the ring structure either alter the metabolism or enhance the potency of a given molecule. For example, the potency of fluoxymesterone, 19-nortestosterone, and 1-methylsubstituted steroids may be enhanced because they are poor precursors for estrogen formation in extraglandular tissues. Similarly, 19-nortestosterone is a more potent androgen than testosterone because its more planar ring structure, like that of dihydrotestosterone, fits more tightly into the binding site of the androgen receptor. 7-Methyl-19-nortestosterone cannot be 5-reduced and may be useful when androgen replacement is needed with minimal effects on the prostate.

As with 17-alkylated steroids, androgens with ring alterations are not converted to testosterone in vivo, and specific assays for each must be used to monitor blood levels. One orally effective androgen, mesterolone, is neither esterified nor alkylated in the 17 position and cannot be aromatized to estrogens in peripheral tissues. Consequently, effective androgen replacement can be achieved by oral administration without causing abnormalities of liver function; unfortunately, the steroid is ineffective in regulating gonadotropin secretion and consequently is a poor agent for routine androgen replacement therapy.
The subcutaneous implantation of testosterone-filled silicone elastomer (Silastic) capsules results in slow release of hormone into plasma for long periods. However, this is impractical because of the large size of such capsules. When large amounts of testosterone are given by mouth in microcrystalline form (200 to 400 mg/day), physiologic blood levels can be achieved, but the preparation has to be taken several times a day and these doses induce hepatic drug-metabolizing enzymes, the long-term effects of which are uncertain.

**Androgens for Normal Men**

When the plasma testosterone level is raised above the normal range, both the basal levels of LH and FSH and the peak levels after GnRH administration are diminished. As a consequence, the testicular volume is decreased about 20%, sperm production is uniformly decreased by 90% or more, and the volume of the ejaculate remains unchanged. Acne is common, and the serum estradiol level increases twofold.

Administration of usual replacement doses of testosterone enanthate (100 mg/week) to normal men caused significant decreases in triglycerid and total body fat and increases in BMD in the spine, effects that are probably the consequence of the temporary increases in testosterone levels above the normal range after the injections. Administration of six times this dose (600 mg/week) of testosterone enanthate to normal men caused an increase in fat-free body mass, triceps and quadriceps muscle size, and muscle strength.

In a similar study, there was no change in levels of prostate-specific antigen (PSA). In a dose-response study in healthy young men, testosterone enanthate caused increases in fat-free body mass, muscle strength, and hemoglobin levels and decreases in fat mass and HDL cholesterol levels, beginning with a dose that was just above replacement levels (125 mg/week). Sexual function, visual-spatial cognition, mood, and PSA levels did not change at any dose.

**Androgens for Hypogonadal Men**

The aim of androgen therapy in hypogonadal men is to restore or normalize male secondary sexual characteristics (beard, body hair, external genitalia) and male sexual behavior to promote normal male somatic development (hemoglobin, voice, muscle mass, nitrogen balance, and epiphyseal closure). Because a reliable assay for plasma testosterone is widely available for monitoring therapy, the treatment of androgen deficiency is straightforward and almost universally successful.

The parenteral administration of a long-acting testosterone ester, such as 100 to 300 mg of testosterone enanthate at 1- to 3-week intervals, results in a sustained increase in plasma testosterone concentration to the normal male range or slightly above. The usual replacement regimen is 200 mg every 2 weeks. Similar effects are obtained with the transcutaneous administration of testosterone.

Such regimens usually reduce the plasma LH level (if elevated) and maintain serum testosterone within the normal range. Serum testosterone should be measured after 4 to 6 weeks of therapy to assess adequacy of dosage; the trough level is measured in men receiving intramuscular testosterone, and midmorning levels are assessed in men receiving transdermal formulations. If the hypogonadism is primary and of long duration (as in Klinefelter's syndrome), suppression of the plasma LH value to the normal range may not occur for many weeks, if at all. In postpubertal testicular failure, even of many years' duration, resumption of normal sexual activity is usual after adequate replacement, primarily because of increased libido and increased frequency of erections.

Androgen therapy does not restore spermatogenesis in hypogonadal states, but the volume of the ejaculate, derived largely from the prostate and seminal vesicles, and other secondary sexual characteristics return to normal. Treatment of hypogonadal men with testosterone replacement in growth of the prostate to the same degree as that of age-matched controls. Testosterone replacement in such men can cause dramatic changes in body composition, strength, and BMD, although maximal effects may not be seen for as long as 2 years. Improvement in BMD in BMD involves both trabecular and cortical bone and is independent of the age at which replacement is started.

In men of all ages in whom hypogonadism develops before expected puberty (such as in hypogonadotropic hypogonadism), it is appropriate to bring plasma testosterone into the adult range slowly. When therapy is begun at the time of expected puberty, the normal events of male puberty proceed in the usual fashion. If therapy is delayed until after the time of expected puberty, the degree of virilization is variable, but many such men undergo a late but relatively complete anatomic and functional male maturation. Intermittent androgen therapy is sometimes given to prepubertal hypogonadal boys with microphallus to stimulate penile growth.

In boys of pubertal age with either isolated hypogonadotropic hypogonadism or primary testicular deficiency, the initial administration of small doses of testosterone esters follows by a gradual increase to doses of 100 to 150 mg/m² of body surface area per month results in a normal pubertal growth spurt. Penile growth, deepening of the voice, and appearance of other secondary sexual characteristics usually commence during the first year of treatment. Puberty in normal boys extends over several years, and treatment that is designed to replicate normal development cannot shorten the process greatly.

In postpubertal testicular failure, even of many years' duration, resumption of normal sexual activity involves both trabecular and cortical bone and is independent of the age at which replacement is started.

**Androgens for Healthy Older Men with Decreased Bioavailable Testosterone Levels**

Because many of the changes in body composition, libido, and erectile function with aging also occur with male hypogonadism, androgen administration has been evaluated in healthy older men. The criterion for inclusion in such studies is a low level of total testosterone or of bioavailable (nonSHBG-bound) testosterone.

In one study of older men with a total serum testosterone less than 14 nmol/L (<420 ng/dL), administration of testosterone enanthate at 100 mg/week returned plasma testosterone levels to normal, increased lean body mass and hemoglobin levels, and decreased total and LDL cholesterol levels.
levels of bioavailable testosterone, treatment with testosterone enanthate at 200 mg every 2 weeks increased muscle strength and hemoglobin levels without increasing PSA levels. 672 In a larger, long-term placebo-controlled study of older men, the use of a testosterone patch increased BMD significantly in the men with low levels of total testosterone but had minimal effects in men with normal levels of total testosterone. 673 674 Lean body mass and hemoglobin levels were increased and fat mass was decreased in all men receiving testosterone. The PSA levels increased by 0.5 ng/mL and then were stable after 6 months.

Although these studies suggest that such treatment is beneficial in men with low plasma testosterone, many questions are still unanswered. 675 For example, it will be necessary to perform long-term studies to be certain that the risk-to-benefit ratio is favorable before such treatment can be recommended routinely.

**Use of Androgens for Purposes Other Than Replacement Therapy**

Administration of testosterone to hypogonadal men has systemic effects in addition to those on the male urogenital tract, including reduction in the urinary excretion of nitrogen, sodium, potassium, and chloride and induction of weight gain. A major component of androgen-induced weight gain and nitrogen retention involves an increase in skeletal and muscle mass. In several species, including humans, the skeletal muscles that support the forelimbs, namely the muscles of the pectoral and shoulder region, show the greatest response, but most muscles probably respond to androgen administration. Such muscles enlarge because of formation of new myofilaments along the myofibrils and because of division of the enlarging myofibrils; the net consequence is an increase in the diameter of muscle fibers and fibrils. 676

The effects of androgens on muscle are not due to different actions of the same hormone but represent the same action in different tissues. It is theoretically possible that a steroid might be devised that would be taken up by or retained selectively by muscle, 677 but no anabolic hormone devoid of androgenic effects has been found. Indeed, all anabolic agents tested in men so far are also androgens and in appropriate doses could be used for androgen replacement. Androgens have been tried in a variety of clinical situations other than hypogonadism with the hope that improvement in nitrogen balance and muscle development could outweigh any deleterious side effects.

**Attempts to Improve Nitrogen Balance in Catabolic States**

After injury, infection, or surgery, the breakdown of body protein is accelerated, and as a consequence excess nitrogen is excreted in the urine. During the recovery phase, nitrogen deficits are replaced. Anabolic steroids can improve the nitrogen balance during the first few days after relatively minor operations in well-nourished individuals, 678 but the decrease in nitrogen loss is minimal and does not appear to be of therapeutic benefit. Likewise, any effect of androgens on weight in undernourished, debilitated, or elderly men is complicated by the fact that many such men, including some men with AIDS, also have secondary testosterone deficiency (see earlier). 679

There is no convincing evidence that testosterone supplementation improves strength or outcome in wasting disorders in the absence of androgen deficiency. 680 Androgens are also of no proven value in the management of nitrogen accumulation in chronic renal failure.

**Androgens and Athletic Performance**

The use of androgens by athletes who believe that athletic performance will be improved constitutes a widespread form of drug abuse. Weight-lifters and body-builders began to use them in the 1950s, and the practice spread to all levels of athletic competition from high school to professional. Several lines of evidence suggest that androgens may have a beneficial effect on strength:

1. In a meta-analysis of 16 studies in athletes, Elashoff and colleagues 681 concluded that androgen administration to trained athletes results in about 5% improvement in strength.
2. Forbes deduced that administration of a total dose of about 20 g of exogenous androgen causes an increase of about 18 kg of lean body mass.
3. Griggs and colleagues 682 reported that large amounts of testosterone increase muscle protein synthesis and muscle mass in normal men.
4. Bhasin and co-workers 683 showed that pharmacologic amounts of testosterone esters increase lean body mass, muscle size, and strength.

Considered together, these studies support the views of athletes and their trainers that such agents are effective in adult male athletes. Although controlled studies have not been carried out in women and boys, it is clear on the basis of uncontrolled studies in the German Democratic Republic that the agents are even more effective in these groups. 683

Unfortunately, the side effects of the drugs are also more striking in women and children. The doses of androgens taken by athletes are often 10 to 100 times ordinary replacement doses. At these doses, androgens may promote anabolism by functioning as antagonists to the catabolic effects of glucocorticoids and hence promote nitrogen retention independent of the androgen receptor.

The question of efficacy, interesting though it may be, is independent of the side effects of the drugs. Because many athletes take oral agents such as nandrolone phenpropionate and stanozolol along with testosterone esters by injection, the potential toxic side effects are formidable.

**Stimulation of Erythropoiesis**

The difference in the hematocrit between men and women is the result of enhancement of erythropoietin formation and erythropoiesis by testosterone. After castration of men, red blood cell mass decreases 10%, red blood cell diameter decreases 40%, and osmotic fragility increases. Occasionally, the resulting anemia may be severe. In normal men given pharmacologic doses of testosterone esters, the average increase in hemoglobin is about 10 g/L (1 g/dL). As a consequence, androgens have been used in the treatment of refractory anemia. 684

The mechanism by which androgens stimulate erythropoietin formation by the kidneys involves the same receptor mechanism that has been documented for other androgen actions, and all androgens have the capacity to enhance erythropoiesis. In humans, some erythropoietin is synthesized outside the kidneys, and the presence of renal tissue is not an absolute requirement for stimulation of erythropoiesis by androgens. 685 686 Androgen has been given to treat anemias associated with bone marrow failure but has been used infrequently since erythropoietin became available for therapy.

Occasional dramatic increases in hemoglobin level occur when administration of androgens to individuals with bone marrow failure. 687 In unselected patients who were treated with androgens, approximately 50% appeared to respond. 688 Improvement appears to be more common when the bone marrow is hypoplastic or when there is myelofibrosis than when the marrow is hypercellular. Furthermore, in a prospective randomized trial of androgen therapy in patients with aplastic anemia, the use of oral androgens at high dosages (1 mg/kg body weight per day) was associated with hematologic improvement and increased survival, primarily in the less severe cases. 689

Androgens have also been used for the anemia of renal failure. Androgen-induced increases in erythropoietin and hemoglobin levels are less marked in the anephric state, as would be expected if the beneficial effect were due to increased erythropoietin formation. 686 688 In most studies, androgen therapy increased the hemoglobin level (10 to 50 g/L [1 to 5 g/dL]) and red blood cell volume (325 to 350 mL), provided that dialysis was adequate and stores of iron and folate were normal. 690 691 Whether the benefits of such treatment outweigh the potential adverse effects is unclear, 692 and the use of androgens for this purpose has largely been replaced by erythropoietin therapy.

**Hereditary Angioneurotic Edema**

In this autosomal dominant disorder, the serum inhibitor of the first component of complement is nonfunctional or absent and unopposed activation of the complement cascade generates factors that enhance the permeability of vessels and produce attacks of angioedema. A variety of 17-alkylated steroids can increase the activity of
the inhibitor in serum and restore the complement components that are depleted secondarily in the disorder.  
Orally active androgens are effective, and steroids such as danazol that are weak androgens appear to be as effective as or more effective than potent androgens.

The response in men and women appears to be the same. Because 17-alkylated androgens (but not testosterone or testosterone esters) cause elevations of several plasma glycoproteins, including haptoglobin, protein-bound sialic acid, plasminogen, and the inhibitor of the first component of complement, the beneficial effect of oral androgens in this disorder is probably the result of the side effect of 17-alkylated steroids on liver function rather than androgen action per se. No reports of the effect of testosterone esters in angioneurotic edema have been published.

Short Statute

Androgens have been used in the management of growth retardation of various causes other than pituitary insufficiency. Their administration before the epiphyses close accelerates linear growth, and the mean height age may be advanced more than skeletal age. Administration of androgens for short periods (6 months or less) has no permanent effects on hypothalamic-pituitary or gonadal maturation. The acceleration of growth may be due to increased levels of plasma growth hormone, but such treatment does not appear to increase final height. Indeed, such therapy in short children before the age of 9 years may actually have a deleterious effect on adult height.

Carcinoma of the Breast

See Chapter 39.

Side Effects of Androgens

Some side effects of androgens are due to physiologic actions of the hormones (through the androgen receptor) but in an inappropriate setting. For example, the virilizing actions are desirable in hypogonadal men but undesirable in women and young boys. In some older hypogonadal men, androgen therapy may cause previously unrecognized prostate cancer to become clinically apparent. Other side effects are the results of actions of androgen metabolites, and because different androgens are metabolized differently, the side effects vary. Testosterone can be metabolized in estrogens and can have feminizing as well as virilizing effects, whereas 5-reduced androgens such as dihydrotestosterone cannot be converted to estrogens and consequently do not have feminizing effects.

Normal people vary in the frequency of side effects, just as there is variability in the degree of virilization of males at puberty. There are also age differences in the occurrence of some side effects. For example, androgens in children may cause premature closure of the epiphyses, may induce gynecomasia, or may produce virilization, even when used in small amounts and for limited periods. The incidence of side effects may also be increased by coexisting clinical conditions. Hepatoma may occur more frequently after androgen treatment of individuals with Fanconi's anemia, sodium retention is worse in patients with congestive heart failure, and feminizing side effects are more prominent in men with hepatic cirrhosis.

A physiologic effect of androgen administration is an increase in the hematocrit, and some men develop polycythemia with such treatment. In one retrospective study, nine of nine older men with high hematocrit values had cerebrovascular events after discontinuation of testosterone therapy.

In a randomized study, significant increases in hematocrit values were more common in men given testosterone esters than a testosterone patch and the increase in hematocrit was related to age, bioavailable testosterone levels, and estradiol levels.

Virilizing Side Effects

All androgens involve the risk of virilizing women and children of both sexes. Coarsening of the voice, hirsutism, and menstrual irregularities are common. If treatment is discontinued as soon as these effects are noticed, the manifestations may slowly subside. With prolonged treatment, male-pattern baldness, hirsutism, coarsening of the voice and enlargement of the cricoïd cartilage, and hypertrophy of the clitoris become largely irreversible.

Feminizing Side Effects

The feminizing side effects of androgens are poorly understood. Testosterone can be converted (aromatized) in peripheral tissues to estrogens and can have feminizing effects, whereas 5-reduced androgens such as dihydrotestosterone cannot be converted to estrogens and consequently do not have feminizing effects.

The conversion of all androgen analogues to estrogens has not been documented, it is presumed that most, if not all, C19 steroids with a 3-keto configuration can be converted to estrogens and that feminization is the effect of estrogenic metabolites of the parent steroids. Administration of testosterone esters to men increases plasma estrogen levels. The most common manifestation of feminization, development of gynecomasia, is unpredictable and in adult men usually occurs only after high-dose androgen therapy. However, in children given androgens, gynecomasia is common (see earlier).

Toxic Side Effects

Some degree of sodium retention is a common consequence of androgen therapy. The amount of retained sodium is usually minor but can cause edema in the presence of underlying heart disease or renal failure.

17-Alkylated androgens impair liver function as evidenced by elevation of plasma alkaline phosphatase and conjugated bilirubin levels during therapy. Such changes are rare in men given parenteral testosterone esters. The predominant effect on hepatic function appears to be at the site of transport of metabolites from hepatocyte into bile. The clinical consequences of abnormal liver function probably depend on the previous integrity of the liver, but jaundice can occur in the absence of preexisting liver disease because of a hypersensitivity reaction. The changes in liver function induced by 17-alkylated drugs include increases in a variety of plasma proteins and decreased conjugation of adrenal steroids.

The most serious complications of oral androgen use are development of peliosis hepatitis (blood-filled cysts in the liver) and hepatoma. These disorders are more common in patients with aplastic anemia but also occur in persons given oral androgens for other reasons, including hypogonadism. Although they are usually benign and regress after discontinuation of the drugs, the tumors may undergo malignant transformation.

Occasional hyperlipidemia has been reported with oral androgens. All androgens including testosterone cause decreases in serum HDL cholesterol levels, and testosterone is the major cause of the differences in serum levels of HDL between men and women. 17-Alkylated androgens may cause a further suppression of HDL and raise LDL levels, but replacement with physiologic levels of testosterone in hypogonadal men does not suppress HDL cholesterol below levels seen in normal men. Sleep apnea has been reported in occasional men given pharmacologic amounts of testosterone esters, possibly the consequence of increased collapse of the upper airways during sleep. Obese men at risk for sleep apnea should be monitored for related symptoms when receiving testosterone therapy. Priapism has occurred in men treated with testosterone enanthate.
Gonadotropin Therapy

Gonadotropin treatment can establish or restore fertility in men who have gonadotropin deficiency either as an isolated disorder or as a part of more extensive anterior pituitary failure. Because men with hypogonadotropic hypogonadism may become resistant to gonadotropins after long-term treatment (presumably as the result of the development of neutralizing antibodies), the customary strategy is to treat such individuals initially with testosterone esters as described earlier and to reserve gonadotropin therapy until fertility is desired. Prior androgen therapy does not impair subsequent gonadotropin induction of spermatogenesis in men with hypogonadotropic hypogonadism.

Two gonadotropin preparations are available: hMG and hCG. The usual preparation of hMG (menotropins), purified from the urine of postmenopausal women, contains 75 IU of FSH and 75 IU of LH per vial. There are several sources of hCG, which is available in vials containing 5000 to 20,000 IU. The hCG is devoid of FSH activity and resembles LH in its ability to stimulate Leydig cells. Because of the expense of hMG, treatment is usually begun with hCG alone, and hMG is added later to stimulate the FSH-dependent stages of spermatid development.

A high ratio of LH to FSH activity and a long duration of treatment are necessary to bring about the maturation of prepubertal testes. In hypophysectomized adult men with long-term suppression of spermatogenesis, it is not predictable whether administration of preparations with both FSH and LH activities is necessary to initiate spermatogenesis. However, after spermatogenesis has been restored in hypophysectomized patients or initiated in hypogonadotropic hypogonadal men by combined therapy, sperm production can usually be maintained by hCG alone.

In men with hypogonadotropic hypogonadism, the dose of hCG required to maintain a normal plasma testosterone level varies from 1000 to 6000 IU/week. Most regimens for the induction of spermatogenesis involve starting with doses of 2000 IU three times or more a week until most of the clinical parameters, including normal male plasma testosterone values, indicate an optimal effect. During initial treatment, the testis volume may reach only 8 mL. Then hMG is added, with as little as 12.5 IU of FSH being required three times a week to complete the development of spermatogenesis and cause further growth of the testes. Optimal spermatogenesis may require treatment for 12 to 24 months.

In most men with hypogonadotropic hypogonadism and no history of cryptorchidism, such a regimen brings sperm counts to the fertile range; for those in whom sperm production is not sufficient for fertility, in vitro fertilization may be successful. The addition of hMG may not be necessary in individuals with partial hypogonadotropic hypogonadism who presumably have some endogenous FSH secretion. Anti-hCG antibodies may develop after long-term hCG treatment, but development of resistance to the action of the hormone is less common.

Recombinant human FSH (rhFSH) is similarly successful in inducing puberty and spermatogenesis in gonadotropin-deficient adolescent and adult men but is ineffective in the treatment of idiopathic infertility.

Treatment with hCG has been used to attempt to promote permanent descent of inguinal testes into the scrotum. Although there are discrepancies among studies because of varying definitions of cryptorchidism, such therapy appears to be successful in about a fifth of patients and makes it possible to identify unambiguously the cryptorchid boys who should be treated surgically (see earlier). Such therapy is associated with a variety of virilizing and feminizing side effects in boys because of enhancement of the testicular production of estradiol and testosterone (see Chapter 24).
Gonadotropin-Releasing Hormone Therapy

GnRH agonists can produce diametrically opposite effects depending on the mode of administration. When given in a pulsatile fashion to mimic physiologic secretory patterns, such therapy enhances gonadotropin secretion. In contrast, the tonic administration of the same agonist inhibits gonadotropin secretion and causes a physiologic (reversible) castration ([741] [742] (see discussion of antiandrogens later). In addition, antagonists have been designed that have no agonist action, regardless of the pattern of administration. ([742]

Agonistic effects are of benefit in hypogonadotropic hypogonadism and cryptorchidism. Because the most common cause of isolated gonadotropin deficiency is defective synthesis or release of GnRH, this agent is the most physiologic treatment of the disorder. In such individuals, the long-term pulsatile administration of GnRH through a portable infusion pump induces normal pubertal development. Normal levels of plasma testosterone, LH, and FSH can be attained with small doses of gonadorelin administered subcutaneously every 90 to 120 minutes. ([743] [744]

Pulsatile GnRH therapy does not appear to have advantages over gonadotropin therapy in men with hypogonadotropic hypogonadism. ([745] GnRH and its analogues have also been used to treat boys with cryptorchidism. The success (or failure) rates for descent of inguinal testes in boys so treated appear to be comparable to those achieved with gonadotropins (see earlier). ([746]
Antiandrogens and 5-Reductase Inhibitors

Agents that block the synthesis or action of androgens are used for treatment of hyperplasia and carcinoma of the prostate, acne, male-pattern baldness, virilizing syndromes in women, precocious puberty in boys, and male sex offenders.

Inhibitors of Testosterone Synthesis

The most effective inhibitor of testosterone synthesis is either GnRH itself or a GnRH agonist or antagonist (see Chapter 8 for details). These agents cause a decline in the plasma levels of LH and testosterone and induce a pharmacologic (and reversible) castration. Such therapy provides a medical alternative to castration for producing androgen deprivation in men with prostatic cancer and produces fewer deleterious side effects than estrogen therapy. \[747\]

In individuals with prostatic carcinoma, such agents are usually administered in conjunction with androgen receptor antagonists, which block the action of androgens of adrenal origin, to produce total androgen deprivation. \[747\] In contrast, monotherapy with GnRH analogues is effective in individuals with precocious puberty. \[748\] Some men with paraphilias, \[750\] and selected men with prostatic hyperplasia. \[752\] These agents are administered parenterally. The frequency of administration and dosage vary with the agent, but long-acting agents such as depot leuprolide acetate are effective when given at monthly intervals in doses of 7.5 mg. \[754\] Vasomotor symptoms characteristic of the castrated state are common and may be severe; management of the latter symptoms is frequently unsuccessful, but they may respond to megestrol acetate. \[756\]

The progestagen medroxyprogesterone acetate inhibits testosterone biosynthesis by at least two mechanisms; at low concentrations it inhibits synthesis directly, and at higher concentrations it inhibits LH secretion. \[757\] Dosages of 400 mg/week by parenteral injection induce a pharmacologic castration and have been used successfully in some men with paraphilia. \[759\]

Antifungal agents of the imidazole class, such as ketoconazole and itraconazole, secondarily block cytochrome P450 enzymes in steroid hormone biosynthesis, effects that have been used to induce androgen deprivation in some prostate cancer. \[760\] Gastrointestinal side effects, short duration of action, and inhibition of the biosynthesis of adrenal steroids limit the usefulness of these agents.

Spironolactone, an aldosterone antagonist (see Chapter 14 and Chapter 15), is also a weak inhibitor of the binding of androgen to the androgen receptor and impairs androgen biosynthesis. In some women with hirsutism, the drug decreases the growth rate and mean diameter of facial hair; \[761\] because it may cause menstrual irregularity, it is commonly given together with an oral contraceptive. \[762\] In efficacy studies, spironolactone was about as effective as finasteride in improving hirsutism scores. \[763\]

5-Reductase Inhibitors

Because the conversion of testosterone to dihydrotestosterone is essential for certain androgen actions, inhibition of steroid 5-reductase selectively blocks androgen action in the tissues (e.g., prostate, certain hair follicles) in which continuing production of dihydrotestosterone is essential. The azasteroid finasteride is an orally active inhibitor that preferentially blocks steroid 5-reductase 2 but may also inhibit enzyme 1. \[764\] At ordinary dosages, the agent causes a profound decline in dihydrotestosterone levels in serum and prostate but has little if any effect on serum testosterone or LH levels. It also has no significant effect on potency or libido. \[765\]

In men with prostatic hyperplasia, finasteride at a dosage of 5 mg/day causes a consistent decrease in prostate size and improvement in urine flow and symptoms in a third, and thus provides an alternative to surgery in men with moderate disease manifestations. \[766\] Finasteride is also effective at a dose of 1 mg/day for male-pattern baldness. \[767\] Additional 5-reductase inhibitors, including agents specific for the individual isozymes, are under development.

Androgen Receptor Antagonists

Several anilide derivatives block the binding of androgen to its receptor. Bicalutamide, a nonsteroidal antianrogen that blocks binding of dihydrotestosterone to the androgen receptor, is devoid of other hormonal activity and has a half-life that allows once-daily oral dosing. Bicalutamide appears to have a low incidence of adverse effects. \[768\] Flutamide, given at doses of 750 mg/day, is converted in vivo to 2-hydroxyflutamide, a potent antienrogen, but because of rapid turnover must be administered three times daily. \[769\] Nilutamide has similar effects at doses of 300 mg/day. \[770\]

In normal men, all three agents block the inhibitory feedback of androgen on LH production and result in an increase in the frequency of LH secretory bursts and hence an increase in serum LH and testosterone. The rise in serum testosterone serves to limit the antianrogenic effectiveness. Consequently, they are most useful in inhibiting androgen action in castrated men, in men in whom LH is inhibited (e.g., in men receiving GnRH agonists), or in women (in whom LH production is not under androgenic control). The principal use is for treatment of prostatic cancer, usually in conjunction with GnRH blockade or estrogen. \[771\]

In men with advanced prostate cancer, bicalutamide and flutamide, each used in combination with a GnRH analogue, have similar effects. \[772\] If the agent crossed the placenta, it would be expected to cause male pseudohemaphroditism in male embryos and consequently should always be given to women in conjuction with an oral contraceptive. Flutamide can cause diarrhea and hepatotoxicity, and nilutamide is associated with side effects similar to those seen with disulfiram (Antabuse) and visual disturbances.

The progestagen cyproterone acetate is an antianrogen antagonist that also suppresses the secretion of gonadotropins and thus interferes with testosterone production. \[773\] The principal effect is believed to be competition with dihydrotestosterone for binding to the androgen receptor. \[774\] In castrated animals, dosages about five times that of testosterone blocked androgenic responses by about 50% and with larger dosages the antagonism was almost complete. \[775\] The administration of cyproterone acetate at 100 mg/day to normal men caused a 50% decrease in serum levels of LH and FSH and a 75% decrease in serum testosterone. \[776\] The agent has been used for the treatment of acne, male-pattern baldness, and hirsutism and virilizing syndromes in women \[777\] and to induce chemical castration in men with paraphilias. \[778\] The agent has orphan drug status in the United States. The fact that it can cross the plaencta and induce male pseudohemaphroditism in male embryos \[779\] and that its use has been associated with severe liver damage, including development of hepatocellular carcinoma, \[780\] limits its usefulness.
Aromatase Inhibitors and Antiestrogens

Aromatase Inhibitors

Second-generation aromatase inhibitors that are both selective and potent include letrozole and anastrozole, both of which are orally active and can be administered once daily. At a dose of 1 mg/day, anastrozole lowered plasma estradiol in men by about half and caused reciprocal increases in plasma LH and testosterone levels, and letrozole had similar effects in male monkeys.

Although not approved for these indications, such agents may be useful in the treatment of familial testotoxicosis in boys, selected cases of gynecomastia, and the aromatase excess syndrome. Potential adverse effects include asthenia, nausea and vomiting, and headache.

Estrogen Receptor Antagonists (see Chapter 39)

Of the two estrogen receptors identified, current antagonists were designed to block the receptor, and it is assumed that dual or selective -receptor antagonists would have different consequences. Tamoxifen at a dose of 20 mg/day was reported to provide resolution in about 75% of cases of idiopathic gynecomastia, and 16 of 18 men with mastalgia or gynecomastia associated with spironolactone therapy for hepatic cirrhosis improved significantly with a dose of 20 mg of tamoxifen twice a day. Similar beneficial effects of tamoxifen have been reported in flutamide-induced gynecomastia and in the gynecomastia that follows castration or leuprolide therapy. A variety of side effects have been reported, including development of fatty liver.
References


123. Evans RM. The steroid and thyroid hormone receptor superfamily. Science 1988; 240:88995.


References


172. George FW, Noble JF. Androgen receptors are similar in fetal and in adult rabbits. Endocrinology 1984; 115:14511458.


195. Hamilton JB. The role of testicular secretions as indicated by the effects of castration in man and by studies of pathological conditions and the short lifespan associated with maleness. Recent Prog Horm Res 1948; 3:257322.


799


References


References

455. Page DC, Silber S, Brown LG. Men with infertility caused by AZFc deletion can produce sons by intracytoplasmic sperm injection but are likely to transmit the deletion and infertility. Hum Reprod 1999; 14:17221726.
468. Farnsworth WE. Estrogen in the etiopathogenesis of BPH. Prostate 1999; 41:263274.


References

1. Larsen. Williams Textbook of Endocrinology, 10th ed. Copyright © 2003 Elsevier

References


796


Erectile dysfunction (ED), previously referred to as impotence, is the inability of the male to attain or maintain a penile erection sufficient for satisfactory sexual intercourse. Sexual dysfunction is a more general term that also includes libidinal, orgasmic, and ejaculatory dysfunction in addition to the inability to attain or maintain penile erection. [1]
MALE SEXUAL DYSFUNCTION

PREVALENCE AND INCIDENCE RATES

The best data on the prevalence of ED in men have emerged from two cross-sectional studies that have used probability sampling techniques, namely the Massachusetts Male Aging Study (MMAS) and the National Health and Social Life Survey (NHSLS). The MMAS was a cross-sectional, community-based random sample epidemiologic survey in which 1709 men in the greater Boston area, 40 to 70 years of age, were first surveyed between 1987 and 1989. Of these, 847 men were resurveyed between 1995 and 1997. This survey revealed that 52% of men between the ages 40 and 70 were affected by ED to some degree; 17.2% of surveyed men reported minimal ED, 25.2% moderate ED, and 9.6% complete ED.

The NHSLS was a national probability survey of English-speaking Americans, 18 to 59 years of age living in the United States in 1992. In this survey, 7% of men between 18 to 29 years of age, 9% between 30 and 39 years of age, 11% between 40 and 45 years, and 18% between 50 and 59 years reported ED, based on self-reports of difficulty in obtaining or maintaining erections.

These two landmark studies and data from several other studies suggest ED is a common problem affecting 20 to 30 million men in the United States alone. The prevalence of ED increases with age; it affects fewer than 10% of men younger than 45 years of age but 75% of men older than 80 years of age. Men with other medical problems, such as hypertension, diabetes, cardiovascular disease, and end-stage renal disease, have a significantly higher prevalence of ED compared with healthy men.

There is a paucity of longitudinal data on the annual incidence rates of ED in men. Most of the available information has been derived from two studies. In the MMAS, of the 1297 men 40 to 70 years of age who were originally surveyed in 1987 to 1989, follow-up information was gathered from 847 men between 1995 and 1997. In this study, the crude incidence rate of ED in white men in the Boston area was 25.9 cases per thousand man-years. The incidence rates increased from 12.4 cases per thousand man-years for men aged 40 to 49 years to 29.8 cases per thousand man-years for men aged 50 to 59 years, and 46.4 per thousand man-years for men aged 60 to 69 years.

In another study, incidence rates were derived from a survey of 3250 men, 6 to 83 years of age, seen at a preventive medicine clinic between 1987 and 1991. This study found the incidence rates of ED to be fewer than 3 cases per 1000 man-years among men younger than 45 years of age and 52 cases per thousand man-years among men 65 years of age or older. On the basis of these two studies, it is estimated that there are 600,000 to 700,000 new cases of ED each year in the United States alone.
REGULATION OF MALE SEXUAL FUNCTION AND PHYSIOLOGY OF PENILE ERECTION

Sexual function is a complex, multifunctional biologic process that comprises central mechanisms for regulation of libido and arousability as well as local mechanisms for the generation of penile tumescence, rigidity, orgasm, and ejaculation. Androgen-deficient men have decreased overall sexual activity but can achieve normal erections in response to visual erotic stimuli. These observations have led to the prevalent dogma that libido is testosterone-dependent and that the local mechanisms for penile erection are androgen-independent. There is emerging evidence that testosterone is a regulator of nitric oxide synthase (NOS) activity in the cavernosal smooth muscle. Therefore, it is possible that physiologically normal testosterone concentrations might be required for optimal penile rigidity. Orgasm and ejaculation are not androgen-dependent and can occur without a full penile erection.

Normal penile erection requires coordinated involvement of intact central and peripheral nervous systems and the corpora cavernosa and spongiosa as well as a normal arterial blood supply and venous drainage (Fig. 19-1) (Figure Not Available). The cavernosal arteries and their branches (the helicine arteries) provide blood flow to the penis. Helicine arteries deliver blood into the cavernosal sinuses. Dilatation of the helicine arteries increases the blood flow and pressure in the cavernosal sinuses. Relaxation of the cavernosal smooth muscle that surrounds the cavernosal sinuses along with increased blood flow results in pooling of blood in the cavernosal spaces and penile engorgement. The expanding corpora cavernosa compress the veins against the rigid tunica albuginea, restricting the venous outflow from the cavernosal spaces. This facilitates entrapment of blood in the cavernosal sinuses and achievement of a rigid erection.

The erectile state of the penis is determined by the tone of the corporal smooth muscle cells. When the cavernosal smooth muscle cells are relaxed, the tone is low and the penis is engorged with blood and erect. Conversely, when the cavernosal smooth muscle tone is high, the penis is flaccid. The smooth muscle tone in the corpora cavernosa is maintained by agonist-stimulated release of intracellular calcium into the cytoplasm and influx of calcium through membrane channels. An increase in intracellular calcium through its binding to calmodulin activates myosin light chain kinase, resulting in phosphorylation of myosin light chain, actin-myosin interactions, and muscle contraction. The transmembrane and intracellular calcium flux in the cavernosal smooth muscle cells is regulated by a number of cellular processes and signaling molecules such as potassium ion (K⁺) flux through potassium channels, connexin 43-derived gap junctions, norepinephrine, prostaglandin E₁, and nitric oxide (NO).

Movement of K⁺ across the membrane determines the membrane potential of the cavernosal smooth muscle cells; at least four subtypes of potassium channels mediate this K⁺ efflux. The adjacent smooth muscle cells are interconnected in a syncytium through connexin 43-derived gap junctions. Therefore, changes in K⁺ channel activity in one myocyte affect the membrane potential of adjacent cells, resulting in rapid transmission of electrical and biochemical signaling throughout the syncytium.

The vascular smooth muscle contraction in the corpora cavernosa is regulated by the noradrenergic pathway. Norepinephrine binds to adrenergic receptors, resulting in generation of diacylglycerol and inositol triphosphate (IP₃). Diacylglycerol activates protein kinase C, which in turn can inhibit K⁺ channels, whereas IP₃ increases intracellular calcium and calcium influx through the membrane. The net increase in intracellular calcium promotes actin-myosin interaction, resulting in smooth muscle contraction.

PGE₂, activation of its receptor results in generation of cyclic adenosine monophosphate (cAMP), which activates protein kinase A (PKA). Activated PKA stimulates K⁺ channels, resulting in K⁺ efflux from the cell. In addition, PKA-mediated processes also result in a net decrease in intracellular calcium, favoring smooth muscle cell relaxation. Nitric oxide, through stimulation of synthesis of intracellular cyclic guanosine monophosphate (cGMP), also decreases intracellular calcium and K⁺ efflux.

The relaxation of the cavernosal smooth muscle trabeculae is under the regulation of the autonomic nervous system. A number of cholinergic, adrenergic, and noradrenergic, noncholinergic (NANC) mediators regulate cavernosal smooth muscle relaxation. The NANC mediators include vasoactive intestinal peptide (VIP), calcitonin gene-related peptide (CGRP), and nitric oxide. Nitric oxide is derived from the nerve terminals innervating the corpora cavernosa, the endothelial lining of penile arteries and cavernosal sinuses, and is an important biochemical regulator of cavernosal smooth muscle relaxation. Nitric oxide also induces arterial dilatation. The actions of nitric oxide on the cavernosal smooth muscle and the arterial blood flow are mediated through the activation of guanylyl cyclase and production of cyclic GMP. The latter acts as an intracellular second messenger and causes smooth muscle relaxation by lowering intracellular calcium.

A class of enzymes, called cyclic nucleotide phosphodiesterases, degrades cGMP into an inactive form, GMP. Researchers have now identified at least 11 different isoforms of cyclic nucleotide phosphodiesterases. These isoforms are widely distributed throughout the body; the predominant isoform of this enzyme in the cavernosal smooth muscle is cyclic nucleotide phosphodiesterase type 5 (PDE5).

Hydrolysis of cGMP by this enzyme results in reversal of the smooth muscle relaxation and reversal of penile erection. Sildenafil (Viagra, Pfizer) is a potent and selective inhibitor of type 5 PDE activity that prevents breakdown of cGMP and thereby enhances penile erection.

Figure 19-1: Anatomy and mechanism of penile erection. The corpora cavernosa are made up of trabecular spaces that are surrounded by cavernosal smooth muscle. Helicine arteries provide the arterial supply to the cavernosal spaces. The dorsal nerve provides the sensory innervation to the penis. During erection, the relaxation of the trabecular smooth muscle and the increased blood flow result in engorgement of the sinusoidal spaces in the corpora cavernosa. The expansion of the sinusoids compresses the venous return against the tunica albuginea, resulting in entrapment of blood. This imparts rigidity to the tumescent penis. (Adapted from Lue T. Erectile dysfunction. N Engl J Med 2000; 342:1802-1813.)
PHYSIOLOGIC, LIFESTYLE, AND PSYCHOSOCIAL CORRELATES AND RISK FACTORS FOR ERECTILE DYSFUNCTION

Epidemiologic studies indicate that the best predictors of the risk of ED are age, a history of diabetes mellitus, hypertension, medication use, and cardiovascular disease. Advancing age is an important risk factor for ED in men: fewer than 10% of men below age 40 years and more than 50% of men older than age 70 years are anticipated to be affected by ED. In both studies mentioned earlier (MMAS and NHSLS), the prevalence of ED increased with each decade of life.

Among the chronic diseases associated with ED, diabetes mellitus is the most important risk factor. In the MMAS, the age-adjusted risk for development of complete ED was three times higher in men with history of treated diabetes mellitus than in men without such a history. Of men with diabetes mellitus, 50% are expected to experience ED at some time during the course of their illness. In the MMAS, treated heart disease, treated hypertension, and hyperlipidemia were associated with significantly increased risk of ED. Among men with treated heart disease and hypertension, the probability of ED was more than two times greater for smokers than for nonsmokers. Smoking also increases the risk of ED in men taking medications for cardiovascular diseases. Cardiovascular disorders, including hypertension, stroke, coronary artery disease, and peripheral vascular disease, are all associated with increased risk of ED.

Several reviews have emphasized the relationship of prescription medications and the occurrence of ED. In the MMAS, the use of antihypertensive agents, cardiac medication,

Figure 19-2 Biochemical mechanisms of penile smooth muscle relaxation. A, Relaxation of the cavernosal smooth muscle is regulated by intracellular cyclic adenosine monophosphate (cAMP) and cyclic guanidine monophosphate (cGMP). The intracellular second messengers, by activation of specific protein kinases, cause sequestration of intracellular calcium (Ca^{2+}), closure of calcium channels, and opening of potassium (K^+) channels. This results in a net decrease in intracellular calcium, causing smooth muscle relaxation. Nitric oxide, released by noradrenergic, noncholinergic nerve endings (NANC), stimulates guanyl cyclase. By inhibiting phosphodiesterase type 5 (PDE5), sildenafil increases the amount of intracellular cGMP. Prostaglandin E, (PGE), stimulates generation of cAMP. Papaverine inhibits PDE, PDE, and PDE and thereby increases the amount of intracellular cAMP. AMP, adenosine monophosphate; GTP, guanosine triphosphate; IP, inositol 1,4,5-triphosphate; NO, nitric oxide; PKA, protein kinase A; PKG, cGMP-specific protein kinase. B, Interconnection of cavernosal smooth muscle cells in the penis. The smooth muscle cells in the corpora cavernosa are interconnected through connexin 43-derived gap junctions. Therefore, alterations in action potential and potassium channel activity in any myocyte affect the adjacent myocytes. (A and B, Adapted and redrawn from Melman A, Christ GJ. Integrative erectile biology: the effects of age and disease on gap junctions and ion channels and their potential value to the treatment of erectile dysfunction. Urol Clin North Am 2001; 28:217-231.) and oral hypoglycemic drugs was associated with an increased risk of ED. Thiazide diuretics and psychotropic drugs used in the treatment of depression may be the most common drugs associated with ED, simply because of the high prevalence of their use. A variety of drugs, however, including almost all antihypertensive agents, digoxin, histamine-2 receptor antagonists, anticholinergics, cytotoxic agents, and androgen antagonists, have been implicated in the pathophysiology of ED.
PATIENT EVALUATION

History

The diagnostic work-up of the patient with ED should start with an evaluation of general health (Table 19-1). The general medical history should be directed at identifying etiologic factors as well as factors that might affect the selection and response to therapy. The presence of diabetes mellitus, coronary artery disease, peripheral vascular disease, and hypertension may suggest a vascular cause. A history of stroke, spinal cord or back injury, multiple sclerosis, or dementia may point to a neurologic disorder. Also relevant is any history of pelvic trauma, prostate surgery, or priapism.

The social history must include ascertainment of recreational drug abuse particularly tobacco, use of cocaine, marijuana and alcohol. Information about medications, particularly antihypertensive agents, antiandrogens, antidepressants, and antipsychotic drugs is important because almost 25% of all cases of impotence can be attributed to medications. Psychiatric illnesses (e.g., depression, psychosis) or drugs used to treat these disorders may be associated with sexual dysfunction.

A detailed sexual history, including the nature of relationships, partner expectations, situational erectile failure, performance anxiety, and marital discord, needs to be elicited (see Table 19-1). It is important to distinguish between an inability to achieve erection, changes in sexual desire, failure to achieve orgasm and ejaculation, and dissatisfaction with the sexual relationship. The physician should inquire about the onset, quality, and duration of erections and the presence of nocturnal and early morning erections.

<table>
<thead>
<tr>
<th>TABLE 19-1 — Diagnostic Evaluation of Erectile Dysfunction</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>History</strong></td>
</tr>
<tr>
<td>1. Ascertain psychosexual history of:</td>
</tr>
<tr>
<td>a. The strength of marital relationship and marital discord</td>
</tr>
<tr>
<td>b. Depression</td>
</tr>
<tr>
<td>c. Stress</td>
</tr>
<tr>
<td>d. Performance anxiety</td>
</tr>
<tr>
<td>2. Ascertain etiologic factors, such as:</td>
</tr>
<tr>
<td>a. The presence of diabetes mellitus, hypertension, end-stage renal disease, peripheral vascular disease</td>
</tr>
<tr>
<td>b. History of spinal cord injury, stroke, or Alzheimer's disease</td>
</tr>
<tr>
<td>c. Prostate or pelvic surgery</td>
</tr>
<tr>
<td>d. Pelvic injury</td>
</tr>
<tr>
<td>e. Concomitant medications, such as antihypertensive, antidepressant, and antipsychotic agents; antiandrogens, such as flutamide, bicalutamide, cyproterone acetate, and cimetidine; and inhibitors of androgen production, such as ketoconazole and GnRH agonists</td>
</tr>
<tr>
<td>f. The use of recreational drugs such as alcohol, cocaine, opiates, and tobacco</td>
</tr>
<tr>
<td>3. Ascertain factors that might affect choice of therapy and the patient's response to it, such as:</td>
</tr>
<tr>
<td>a. Coexisting coronary artery disease and its symptoms and severity</td>
</tr>
<tr>
<td>b. The use of nitrates for angina</td>
</tr>
<tr>
<td>c. Exercise tolerance</td>
</tr>
<tr>
<td>d. The use of vasodilators for hypertension or congestive heart failure</td>
</tr>
<tr>
<td><strong>Physical Examination</strong></td>
</tr>
<tr>
<td>1. Ascertain signs of androgen deficiency, such as loss of secondary sex characteristics, eunuchoidal proportions, small testicular volume, or breast enlargement</td>
</tr>
<tr>
<td>2. Exclude neurologic findings of spinal cord lesion, previous stroke, or peripheral neuropathy; genital and perineal sensation</td>
</tr>
<tr>
<td>3. Palpate femoral and pedal pulses, and exclude evidence of lower extremity ischemia</td>
</tr>
<tr>
<td>4. Perform penile examination to exclude Peyronie's disease</td>
</tr>
<tr>
<td><strong>Laboratory Evaluation</strong></td>
</tr>
<tr>
<td>1. Brachial penile blood pressure index</td>
</tr>
<tr>
<td>2. Intracavernosal injection of vasodilator</td>
</tr>
<tr>
<td>3. Duplex Doppler ultrasonography</td>
</tr>
<tr>
<td>4. Pelvic arteriography</td>
</tr>
<tr>
<td>5. Cavernosography</td>
</tr>
</tbody>
</table>
Physical Examination

A directed physical examination should assess secondary sex characteristics, the presence or absence of breast enlargement, and testicular volume. An evaluation of femoral and pedal pulses can provide clues to the presence of peripheral vascular disease (see Table 19-1). The neurologic examination focuses on the presence of motor weakness, perineal sensation, anal sphincter tone, and bulbocavernosus reflex. The examination of the penis notes any unusual curvature, palpable plaques, or superficial lesions.
Self-Reporting Questionnaires

Over the last decade, there has been a general shift in most male sexual dysfunction clinics away from expensive, time-consuming, and invasive techniques (e.g., dynamic infusion cavernosometry, penile duplex Doppler ultrasonography, and Rigiscan studies) toward the use of simple, noninvasive, self-reporting questionnaires. These questionnaires have been found to be of value because many men with ED, for a variety of reasons, do not voluntarily come forward to their physicians and state their sexual complaints. Many men with ED feel embarrassed, whereas others consider ED an inevitable concomitant of the aging process. Some physicians themselves feel uncomfortable while discussing issues of such personal nature; this creates an atmosphere that is not conducive to effective communication. These self-reporting questionnaires can help break the ice and facilitate communication. These instruments are widely available and easy to complete, and they can complement or enhance the work-up of sexual dysfunction.

The International Index of Erectile Function (IIEF) is a multidimensional scale consisting of 15 questions that address relevant domains of male sexual function. The IIEF has been validated in several languages, used in many multinational clinical trials, and has been found to have adequate sensitivity and specificity for detecting treatment-related changes, including response to oral erectogenic agents in men with ED.

Considerable effort has been invested in the development of abbreviated questionnaires that take less time than the IIEF but more concisely address similar aspects of male sexuality (e.g., Sexual Health Inventory for Men Questionnaire [SHIM]).
Laboratory Tests

The diagnostic evaluation of a man with ED starts with general health evaluation. This may include measurements of hemoglobin, white blood cells, blood glucose, aspartate transaminase (AST), alanine transaminase (ALT), bilirubin, alkaline phosphatase, blood urea nitrogen (BUN), and creatinine.

Measurement of serum total testosterone concentrations can help detect androgen deficiency. Although there is no consensus on this issue, it is important to exclude androgen deficiency in men presenting with ED because this deficiency may be a manifestation of serious underlying illness, such as a pituitary tumor or human immunodeficiency virus (HIV) infection. In addition, testosterone replacement is desirable in men with androgen deficiency not only for restoration of sexual function but also for maintaining bone mineral density, muscle mass and protein metabolism, and a sense of well-being.

If the history, physical examination, and ED questionnaire do not identify any obvious medical concerns warranting further work-up, a more cost-effective approach for a busy practitioner is to prescribe a trial of oral medication (e.g., sildenafil) if there are no contraindications (e.g., nitrate use). See algorithm A.
Evaluation of Penile Vasculature and Blood Flow

Several tests are available to assess the integrity of penile vasculature and blood flow. Of these, the penile brachial blood pressure index is a simple and specific, but not a very sensitive, indicator of vascular insufficiency. It is of historical interest but is rarely used today.

Intracavernosal injection of a vasoactive agent such as PGE₁ can be used as both a diagnostic and potential therapeutic modality. This procedure can show whether the patient will respond to this therapeutic modality and can facilitate patient education about the procedure and its potential side effects. Failure to respond to intracavernosal injection sometimes suggests vascular insufficiency or a venous leak that might need further evaluation and treatment.

Most men with ED do not need duplex color sonography, cavernosography, or pelvic angiography. These procedures should be reserved for patients in whom the results of these tests would alter the management or prognosis and only by physicians with considerable experience in their use. For instance, angiography may be useful in a young man with arterial insufficiency associated with pelvic trauma. Similarly, suspicion of congenital or traumatic venous leak in a young man presenting with ED would justify cavernosography. In each instance, confirmation of the vascular lesion might lead to consideration of surgery. Duplex ultrasonography can provide a noninvasive evaluation of vascular function.
Nocturnal Penile Tumescence

Although recording of formal nocturnal penile tumescence (NPT) in a sleep laboratory for successive nights can help differentiate organic from psychogenic impotence, this test is expensive and labor-intensive and is not required in most men with ED. In most cases, formal NPT studies are reserved for medical-legal documentation.

The introduction of portable Rigiscan devices in 1985 has provided clinicians with a reliable means of continuously monitoring penile tumescence and rigidity at home. The patient wears this multicomponent device at bedtime for two to three nights. Two wiregauge loops are placed around the base and tip of the penis to record changes in penile circumference and rigidity. Data are stored and then downloaded via a software program that allows for sophisticated interpretation.

NPT testing is not needed for most men with ED. It is recommended only for patients with a high clinical suspicion of psychogenic ED or situational problems or to document preoperatively poor penile rigidity. For most cases, a careful history eliciting nighttime or early morning erections provides a reasonable correlation with formal NPT and Rigiscan studies.
Diagnostic Tests to Exclude Androgen Deficiency and Hypothalamic-Pituitary Lesions

There is considerable debate about the usefulness and cost-effectiveness of hormonal evaluation and the extent to which androgen deficiency should be investigated in men presenting with ED. Of all men with ED, 8% to 10% have low testosterone levels; the prevalence of androgen deficiency increases with advancing age. The prevalence of low testosterone levels is not significantly different in men who present with ED and in an age-matched population. Urologic studies report that in 6% to 8% of men with ED there is an endocrine basis to the condition. These data are consistent with the proposal that ED and androgen deficiency are two common but independently distributed disorders.

Yet, it is important to exclude androgen deficiency in this patient population. Androgen deficiency is a correctable cause of sexual dysfunction, and some men with ED and low testosterone levels do respond to testosterone replacement. Hypogonadism can have additional deleterious effects on the patient’s health; for instance, hypogonadism might contribute to osteoporosis, loss of muscle mass and function, and increased risk of disability, falls, and fracture. In addition, in cross-sectional epidemiologic studies, low testosterone levels are associated with increased risk of midsegment obesity, insulin resistance, type 2 diabetes mellitus, and coronary artery disease; however, we do not know whether testosterone replacement can reduce visceral fat and improve insulin sensitivity and cardiovascular risk in middle-aged men with midsegment obesity. Regardless of the presence of sexual dysfunction, androgen deficiency should be corrected by appropriate hormone replacement therapy. Further, androgen deficiency may be a manifestation of a serious underlying disease, such as HIV infection or a hypothalamic-pituitary space-occupying lesion.

In large studies, only a small fraction of men with ED and low testosterone levels have been found to have space-occupying lesions of the hypothalamic-pituitary region. In one large survey, all of the hypothalamic-pituitary lesions were found in men with serum testosterone levels below 150 ng/dL. Therefore, the cost-effectiveness of the diagnostic work-up to rule out an underlying lesion of the hypothalamic-pituitary region can be increased by limiting the work-up to men with serum testosterone levels less than 150 ng/dL.
TREATMENT

The selection of the therapeutic modality should be based on the underlying etiology, the patient's preference (goal-directed approach), the nature and strength of relationship with the man's sexual partner, and the absence or presence of underlying cardiovascular disease and other co-morbid conditions. In current practice, it is common to employ a step approach that first utilizes minimally invasive therapies that are easy to use and produce fewer adverse effects and then progresses to more invasive therapies that may require injections in some circumstances or surgical interventions after the first-line choices have been exhausted (Table 19-2). The physician needs to discuss the risks, benefits, and alternatives of all the diagnostic procedures and therapies with the couple.

It is intuitive that, in the execution of good medical practice, all associated medical disorders need to be optimized. In men with diabetes mellitus, efforts to optimize glycemic control are instituted, although improving glycemic control may not necessarily improve sexual function. In men with hypertension, control of blood pressure is optimized, and, if possible, the therapeutic regimen may need to be modified to discontinue antihypertensive drugs that impair sexual function. This strategy is not always possible because almost all antihypertensive agents have been associated with sexual dysfunction; the frequency of this adverse event is less with angiotensin-converting enzyme (ACE) inhibitors than with other agents.

All patients with ED can benefit from psychosexual counseling. Unfortunately, many couples are reluctant to pursue this avenue. When there is latent marital discord, the sensitive and astute clinician should direct affected couples appropriately.

First Line Therapies

Psychosexual Counseling

Counseling can be of benefit in both psychogenic and organic causes of sexual dysfunction (see Table 19-2). It can help decrease performance anxiety and increase the patient's ability to cope with the problem. Involving the partner in the counseling process can help dispel misperceptions about the problem, decrease stress, enhance intimacy and ability to talk about sex, and increase the chances of successful outcome of therapy. Counseling sessions are also helpful in uncovering conflicts in relationships, psychiatric problems, alcohol and drug abuse, and significant misperceptions about sex. Although psychobehavioral therapy has been claimed to relieve depression and anxiety, there is a striking paucity of outcome data on the effectiveness of this therapeutic modality.

TABLE 19-2 – A Stepwise Approach to Treatment of Erectile Dysfunction

<table>
<thead>
<tr>
<th>Step</th>
<th>Therapies</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>All patients and their sexual partners can benefit from and should receive psychosexual counseling.</td>
</tr>
<tr>
<td>2.</td>
<td>First-line therapies</td>
</tr>
<tr>
<td>a.</td>
<td>Sildenafil citrate</td>
</tr>
<tr>
<td>b.</td>
<td>Vacuum constriction devices</td>
</tr>
<tr>
<td>3.</td>
<td>Second-line therapies</td>
</tr>
<tr>
<td>a.</td>
<td>Intracavernosal injection of alprostadil</td>
</tr>
<tr>
<td>b.</td>
<td>Intracavernosal injections of other vasoactive amines</td>
</tr>
<tr>
<td>4.</td>
<td>Third-line therapies</td>
</tr>
<tr>
<td>a.</td>
<td>Penile prosthesis</td>
</tr>
<tr>
<td>b.</td>
<td>Vascular surgery</td>
</tr>
</tbody>
</table>

Sildenafil

Sildenafil (Viagra) is the first effective oral agent for the treatment of ED. It was introduced to the United States market in March 1998, and since that time millions of tablets have been dispensed. Sildenafil is a selective type 5 PDE inhibitor that is a safe and effective first-line, oral treatment for ED (Table 19-3 and Table 19-4). Therefore, sildenafil action requires an intact nitric oxide response as well as constitutive synthesis of cGMP by the smooth muscle cells of the corpora cavernosa. By selectively inhibiting cGMP catabolism in the cavernosal smooth-muscle cells, sildenafil restores the natural erectile response to sexual stimulation but, importantly, does not produce an erection in the absence of sexual stimulation. The fidelity of the nitric oxide production pathway and sexual stimulation are both necessary requirements for sildenafil to induce an erection.

Efficacy

The efficacy of sildenafil was demonstrated in a randomized dose-response study in which 532 men with organic, psychogenic, or mixed ED were randomized to receive placebo or 25, 50, or 100 mg of sildenafil for 24 weeks. In this dose-response study, patients taking sildenafil performed better in terms of increased rigidity, frequency of vaginal penetration, and maintenance of erection. Increasing doses of sildenafil were associated with higher mean scores for the questions assessing frequency of penetration and maintenance of erections after sexual penetration.

In a follow-up dose escalation study, 329 men were randomly assigned to receive placebo or 50 mg of sildenafil for 12 weeks. At each follow-up, the dose of sildenafil was increased or decreased by 50%, depending on the therapeutic response.

TABLE 19-3 – Recommendations for the Use of Sildenafil by Men with Cardiac Disease (American College of Cardiology and American Heart Association)

<table>
<thead>
<tr>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Sildenafil (Viagra) is absolutely contraindicated in men taking long-acting or short-acting nitrate drugs on a regular basis.</td>
</tr>
<tr>
<td>2. If the patient has stable coronary artery disease, is not taking long-acting nitrates, and uses short-acting nitrates only infrequently, the use of sildenafil should be guided by careful consideration of risks.</td>
</tr>
<tr>
<td>3. All men taking nitrates should be warned about the risks of the potential interaction between nitrates and sildenafil. The patients should also be warned that concurrent recreational use of inhaled nitrates or “poppers” may result in marked hypotension that may be serious or even fatal.</td>
</tr>
<tr>
<td>4. Sildenafil is contraindicated within 24 hours of ingestion of any form of nitrate.</td>
</tr>
<tr>
<td>5. In men with preexisting coronary artery disease, the physician should assess the risks of inducing cardiac ischemia during sexual activity before prescribing sildenafil. This assessment may include a stress test.</td>
</tr>
</tbody>
</table>
6. Men who are taking a combination antihypertensive medication should be warned about the possibility of sildenafil-induced hypotension. This is of particular concern in men with congestive heart failure who have borderline low blood volume or who are receiving complex regimens that include vasodilators or diuretics.  


<table>
<thead>
<tr>
<th>Table 19-4 – Common Adverse Events Associated with the Use of Sildenafil in men with Erectile Dysfunction.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Adverse Event</strong></td>
</tr>
<tr>
<td>Headache</td>
</tr>
<tr>
<td>Flushing</td>
</tr>
<tr>
<td>Dyspepsia</td>
</tr>
<tr>
<td>Rate of discontinuation</td>
</tr>
</tbody>
</table>

Sildenafil Guidelines in Men with Coronary Artery Disease

In a separate randomized clinical trial, 268 men with diabetes mellitus and ED received either placebo or sildenafil for 12 weeks. Fifty-six percent of men receiving sildenafil reported improved erections compared with 10% of those receiving placebo (P < .001). The rate of men reporting at least one successful attempt at intercourse was 61% for the sildenafil group versus 22% for the placebo group. The study demonstrated that sildenafil was an effective treatment for ED in patients with diabetes mellitus.  

Sildenafil is also effective in men with ED due to a variety of other causes, including spinal cord injury and postradical prostatectomy. In general, baseline sexual function correlated positively with response to sildenafil, and patients with diabetes mellitus or previous prostate surgery responded less well than patients with psychogenic or vasculogenic ED. Because there is no baseline characteristic that predicts the likelihood of failure to respond to sildenafil therapy, a therapeutic trial of sildenafil is warranted in all patients except in those in whom it is contraindicated.  

Adverse Effects Associated with Sildenafil Therapy

In clinical trials, the adverse effects that have been reported with greater frequency in sildenafil-treated men than placebo-treated men include headache, flushing, dyspepsia, respiratory tract disorders, and visual disturbances (see Table 19-4). Sildenafil does not affect semen characteristics. No cases of priapism were noted in any of the pivotal clinical trials.

Hemodynamic Effects of Sildenafil Cite

In postmarketing surveillance, several instances of myocardial infarction and sudden death were reported in men using sildenafil. Forty-four of the 120 deaths reported by the U.S. Food and Drug Administration, from March to November 1998, occurred in temporal relation to the ingestion of sildenafil; 16 of these deaths occurred in individuals who were taking nitrates. Because most men presenting with ED also have high prevalence of cardiovascular risk factors, it is unclear whether these events were causally related to the ingestion of sildenafil, underlying heart disease, or both.  

In a rigorously controlled study, oral administration of 100 mg sildenafil to men with severe coronary artery disease produced only small decreases in systemic blood pressure and no significant changes in cardiac output, heart rate, coronary blood flow, or coronary artery diameter. This led the American Heart Association to conclude that the preexistence of coronary artery disease by itself does not constitute a contraindication for the use of sildenafil.  

Sildenafil Guidelines in Men with Coronary Artery Disease

Before sildenafil is prescribed to a patient, it is crucial to assess cardiovascular risk factors (see Table 19-3). If the patient has hypertension or symptomatic coronary artery disease, the treatment of those clinical disorders must be addressed first. The physician must inquire about the use of nitrates because sildenafil is absolutely contraindicated in individuals taking any form of nitrates; sildenafil should not be used within 24 hours of the use of nitrates.  

In men with preexisting coronary artery disease, sexual activity can induce coronary ischemia; these individuals should undergo assessment of exercise tolerance. One practical way to assess exercise tolerance is to have the patient climb one or two flights of stairs. If the patient can safely climb one or two flights of stairs without angina or excessive shortness of breath, he can probably engage in sexual intercourse with a stable partner without similar symptoms. Exercise testing before prescribing sildenafil may be indicated in some men with significant heart disease to assess the risk of inducing cardiac ischemia during sexual activity.  

In men with congestive heart failure, men receiving vasodilator drugs, or taking complex regimens of antihypertensive drugs, it is advisable to monitor blood pressure after initial administration of sildenafil.  

A number of reviews have been published on the safety of sildenafil therapy.  

Drug-Drug Interactions

Sildenafil is metabolized mainly by the P450 2C9 and the P450 3A4 pathways. Cimetidine and erythromycin, inhibitors of P450 3A4, increase the plasma concentrations of the drug. Protease inhibitors may also alter the activity of the P450 3A4 pathway and can affect the clearance of sildenafil.  

Conversely, sildenafil is an inhibitor of the P450 2C9 metabolic pathway, and its administration has the potential to affect the metabolism of drugs metabolized by this system, such as warfarin and tolbutamide. The most serious interactions of sildenafil occur with nitrates. The vasodilator effects of nitrates are augmented by sildenafil; this effect also applies to inhaled forms of nitrates such as amyl nitrate or nitrites sold under the street name "poppers." Concomitant administration of the two drugs can cause a potentially fatal decrease in blood pressure.  

Therapeutic Regimens

In most men with ED, sildenafil can be started at an initial dose of 50 mg. In patients with significant coronary artery disease, an initial trial of 25 mg of sildenafil may be appropriate. If this dose does not produce any adverse effects, the dose can be titrated to 100 mg. Further dose adjustment should be guided by the therapeutic response and the occurrence of adverse effects. Typically, unit doses higher than 100 mg are not recommended. Sildenafil is taken at least 1 hour before...
sexual intercourse and not more than once in any 24-hour period.

Cost-Effectiveness of Sildenafil

A number of studies have evaluated the economic cost of treating ED in men in managed care organizations. These analyses, using prevalence-based cost-of-illness approach, have concluded that sildenafil and vacuum constriction devices are the most cost-effective of all the available therapeutic options in managed care setting and should be considered first-line strategies.

Vacuum Devices for Inducing Erection

Commercially available vacuum devices consist of a plastic cylinder, a vacuum pump, and an elastic constriction band. The plastic cylinder fits over the penis and is connected to a vacuum pump. The negative pressure created by the vacuum within the cylinder draws blood into the penis, producing an erection. An elastic band slipped around the base of the penis traps the blood in the penis, and an erection is maintained as long as the rubber band is retained around the base. The constriction band should not be left in place for more than 30 minutes.

These devices are safe, relatively inexpensive, and reasonably effective. They can impair ejaculation, however, resulting in entrapment of semen, and are difficult and awkward for some patients to use. Some couples dislike the lack of spontaneity engendered by the use of these devices. Partner cooperation is usually important for their successful use.

Testosterone Replacement in Androgen-Deficient Men with Erectile Dysfunction

Testosterone replacement in healthy, young, androgen-deficient men restores sexual function. In healthy young men, relatively low normal levels of serum testosterone can maintain sexual function. In male rats, a decrease in serum testosterone concentrations to castrate levels is associated with marked impairment of all measures of mating behavior. Testosterone replacement of castrated male rats to levels that are at the lower end of the adult male range normalizes all measures of mating behavior. In general, supraphysiologic doses of testosterone do not further improve sexual function. Although increasing testosterone levels above the physiologic range may increase arousability, this outcome has not been conclusively demonstrated.

As previously stated, androgen deficiency and ED are two common but independently distributed clinical disorders in middle-aged and older men that often coexist in the same patient. Of men presenting with ED, 8% to 10% have low testosterone levels. The prevalence of low testosterone levels is not significantly different between middle-aged and older men with impotence and those without impotence. Testosterone administration is unlikely to improve sexual function in men with normal testosterone levels. Therefore, indiscriminate use of testosterone replacement in all older men with ED is not warranted. However, it is important to exclude testosterone deficiency in older men presenting with ED. Androgen deficiency may be a manifestation of an underlying disease such as a pituitary tumor. Additionally, therapies directed just at ED in men do not correct androgen deficiency, which, if left uncorrected, would have deleterious effects on bone, muscle, energy level, and sense of well-being.

Many, but not all, of the men with ED and low testosterone levels experience improvements in libido and overall sexual activity with androgen replacement therapy. The response to testosterone supplementation, even in this group of men, is variable because of the coexistence of other disorders, such as diabetes mellitus, hypertension, cardiovascular disease, and psychogenic factors. A meta-analysis of the usefulness of androgen replacement therapy has concluded that testosterone administration is associated with greater improvement in sexual function than that associated with placebo in men with ED and low testosterone levels.

In middle-aged and older men, ED is often a multifactorial disorder. Common causes of ED in men include diabetes mellitus, hypertension, medications, peripheral vascular disease, psychogenic factors, and end-stage renal disease. Many of these factors often coexist in the same patient. Therefore, it is not surprising that testosterone treatment alone may not improve sexual function in all men with androgen deficiency.

Testosterone treatment does not improve sexual function in impotent men with normal testosterone levels. It is not known whether testosterone replacement improves sexual function in impotent men with borderline serum testosterone levels.
Second-Line Therapies

Intracavernosal Therapies

An intracavernosal system for delivery of alprostadil called MUSE (medicated urethral system for erection, VIVUS) was released in 1997. Alprostadil is a stable, synthetic form of PGE$_1$, which causes an increase in cAMP levels and thereby promotes cavernosal smooth muscle relaxation and penile erection.

Alprostadil, when applied into the urethra, must be absorbed through the ventral side of the tunica albuginea and into the corpus cavernosum to cause an erection. High concentrations of PGE$_1$ must be used to maintain efficacy.

Typically, the clinician applies the initial MUSE dose of 500 µg in the office to observe any changes in blood pressure or urethral bleeding secondary to misapplication of the MUSE device into the urethra. Depending on the erectile response, the clinician can either increase the alprostadil dose to 1000 µg or decrease the dose to 250 µg.

Common side effects of transurethral alprostadil are penile pain and urethral burning in up to 30% of patients. Initial randomized, placebo-controlled studies reported 40% to 60% success rates, defined as having at least one successful sexual intercourse during a 3-month study period.

The use of intracavernosal injections of vasoactive agents has been a cornerstone of the medical management of ED since the early 1980s. Patients are taught how to self-inject a vasoactive agent into the corpus cavernosum with a 27- or 30-gauge needle up to 3 times a week for sexual intercourse. Erections occur typically 15 minutes after intracorporal injection and ideally last 45 to 90 minutes. When titration is correct, the success rate of this therapy in producing a rigid erection is 80% to 90%. Early studies with intracavernosal injection therapy report patient and partner satisfaction rates of 70% and 67%, respectively.

The main adverse effects include penile pain, occurrence of hematoma, formation of corporal nodules, and the possibility of prolonged erections (priapism if longer than 4 hours). Despite the effectiveness of this approach in producing rigid erections, most patients do not relish injecting a needle into the penis; therefore, it is not surprising that long-term drop-out rates approach 60% to 70%.

Intracavernosal injection therapy can be used as first-line therapy but is often reserved for patients who have failed medical therapy or for those who need more control over their erections. The use of a constriction device (Actis, VIVUS) at the time of application of transurethral alprostadil has been shown to increase efficacy. VIVUS is currently investigating a combination of alprostadil and prazosin (an -blocking agent) to further promote cavernosal smooth muscle relaxation and erectile response.

<table>
<thead>
<tr>
<th>TABLE 19-5</th>
<th>Checklist Before Administration of Intracavernosal Therapy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The patient should be instructed on how to inject the medication and should be educated about the risks of this form of therapy.</td>
<td></td>
</tr>
<tr>
<td>2. Physicians who wish to prescribe intracavernosal injections must have contingency plans and a designated urologist to handle emergencies related to complications such as priapism.</td>
<td></td>
</tr>
<tr>
<td>3. It is advisable to administer the first injection in the physician’s office and to observe the blood pressure and heart rate response. This provides an excellent opportunity for educating the patient, observing adverse effects, and determining whether the patient will respond to this form of therapy.</td>
<td></td>
</tr>
<tr>
<td>4. The dose of the vasoactive substance should be adjusted to achieve an erection that is sufficient for sexual intercourse but that does not last more than 30 minutes.</td>
<td></td>
</tr>
<tr>
<td>5. The patient should be advised that priapism and fibrosis are potential complications of intracavernosal therapy.</td>
<td></td>
</tr>
<tr>
<td>6. After the injection, the patient should compress the injection site to minimize the risk of hematoma formation and subsequent fibrosis.</td>
<td></td>
</tr>
<tr>
<td>7. If the erection does not abate in 30 minutes, the patient should be instructed to take a pseudoephedrine tablet or an intracavernosal injection of phenylephrine. If this is not effective, the patient should either call a designated urologist or come to the emergency department.</td>
<td></td>
</tr>
<tr>
<td>8. Intracavernosal therapy is not suitable for patients with psychiatric disorders, hypercoagulable states, or sickle cell disease and for those who are receiving anticoagulant therapy.</td>
<td></td>
</tr>
</tbody>
</table>

Three different agents are commonly used alone or in combination by clinicians who prescribe injection therapy for the treatment of ED:

1. **Papaverine (PGE$_1$)**: Papaverine, a nonspecific PDE inhibitor, which increases both intracellular cAMP and cGMP. As a single agent, it is efficacious and inexpensive and needs no refrigeration. It does not cause penile pain but tends to induce priapism and fibrosis with long-term use.

2. **Phentolamine**: Phentolamine is a competitive , -adrenergic and , -adrenergic antagonist that contributes to smooth muscle relaxation. As a single agent, it is minimally efficacious but is commonly used in combination to potentiate papaverine and/or PGE$_1$ action.

In an attempt to maximize efficacy and minimize side effects, many clinicians use a combination of PGE$_1$, papaverine, and phentolamine as a tri-mix, which allows the use of a lower dose of each agent. A common mixture is 2.5 mL papaverine (30 mg/mL), 0.5 mL phentolamine (5 mg/mL), and 0.05 mL alprostadil (500 mg/mL). The reliable patient can titrate the intracavernosal dose from 0.2 to 0.5 mL to optimize his erectile response.

The biggest concern with intracavernosal injection therapy is priapism. Early studies with intracavernosal injection therapy report patient and partner satisfaction rates of 70% and 67%, respectively.
Third-Line Therapy: The Penile Prosthesis

In the early part of the 20th century, the treatment of ED was virtually nonexistent. It was the introduction of the semirigid and inflatable penile prostheses in the early 1970s that initiated the great strides that have been made in recent years in the treatment of ED. To some, penile prostheses are considered invasive and costly, but for many patients with advanced organic disease who are unresponsive to any contemporary form of therapy, have significant structural disorders of the penis (e.g., Peyronie’s disease), or have suffered corporal loss from cancer or traumatic injury, prostheses remain a highly effective and predictable method for restoring erectile function.

Implantation surgery usually takes less than an hour and in most cases can be done as an outpatient procedure with general or regional anesthesia. Recent studies have reported that more than 80% of patients and 70% of partners are pleased with their prosthesis and the togetherness that it brings to their relationship.

Penile implants are paired supports that are placed one in each of the two erectile bodies. There are two basic types: (1) hydraulic or fluid-filled (inflatable) implants and (2) semirigid implants, which are bendable but always remain firm in the penis.

With a number of recent modifications incorporating newer materials and designs, the chance of mechanical malfunction is only 5% to 10% in the first 10 years. Penile prostheses have a higher reliability rate than any other mechanical device implanted in the human body. The most feared complication by surgeons with prosthesis implantation is infection, which occurs in 1% to 3% of cases, but this can be higher in revision surgery, especially in diabetic patients.
Oral Therapeutic Agents under Development

It has been clearly established, in North America at least, that the preferred route of administration for an ED treatment is by mouth. The huge success of sildenafil attests to the demand for an effective oral erectogenic agent. Research and development in this field by other pharmaceutical enterprises has introduced a number of new PDE type 5 inhibitors with supposedly increased potency and fewer side effects.

PDEs have a ubiquitous presence in the human body. PDE type 5 inhibitors are recognized to enhance the effects of sexual stimulation through nitric oxide increases in cGMP in the penis. Besides acting in the corpus cavernosum of the penis, PDE type 5 inhibition is active in skeletal muscle, vascular smooth muscle, platelets, and visceral smooth muscle. It is the PDE type 5 inhibition cross-reacting in other tissues that causes most of the recognized side effects (e.g., headache, gastroesophageal reflux, muscle cramps, visual acuity changes).

The efficacy and side effects of PDE inhibitors are a function of pharmacologic specificity and bioavailability. Vardenafil (Bayer) and Cialis (Lilly-ICOS) are two new PDE type 5 inhibitors with better tissue specificity and pharmacokinetic profiles than sildenafil and are in advanced stages of clinical development (Table 19-6). Because of their reported increased selectivity, overall efficacy, and decreased number of adverse effects in Phase II and III studies, they show clinical promise. Their availability in the United States market is anticipated.

The PDE type 5 inhibitors are recognized to act as peripheral conditioners (i.e., enhance a local pathway to cause an erection). Another exciting area of research are drugs that initiate erections by actions directed within the central nervous system. Apomorphine SL (TAP Pharmaceuticals, Inc.) is an aporphine (not an opiate) that acts as a dopaminergic agonist. It is effective centrally at picomolar concentrations and has actions on the paraventricular and supraoptic nuclei of the midbrain. It has been reformulated for sublingual absorption with an onset of action within 15 to 20 minutes. Adverse effects include nausea, vomiting, and, rarely, syncope. Ongoing clinical research studies and experience from Europe since its approval in June 2001 may bring it to the United States in the near future.

Another central initiator of erection in early development is -melanocyte-stimulating hormone (melanotan-II). It not only has dopaminergic agonist activity but also has beneficial effects on libido.

In a double-blind, placebo-controlled crossover study, there was significant improvement in Rigiscan events, penile rigidity, and sexual desire in 10 patients with documented ED risk factors who were taking melanotan-II compared to those taking placebo. Nausea was the most common adverse event.

### Table 19-6 -- Specificity of Phosphodiesterase (PDE) Type 5 Inhibitors

<table>
<thead>
<tr>
<th>Isoenzyme</th>
<th>Distribution</th>
<th>Sildenafil (nM)</th>
<th>Cialis (nM)</th>
<th>Vardenafil (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDE Type 5</td>
<td>Corpus cavernosum, skeletal muscle, vascular and visceral smooth muscle</td>
<td>3.5</td>
<td>0.94</td>
<td>0.7</td>
</tr>
<tr>
<td>PDE Type 6</td>
<td>Retina</td>
<td>34</td>
<td>730</td>
<td>157</td>
</tr>
</tbody>
</table>

IC₅₀ is the concentration of a given agent to inhibit 50% of a given phosphodiesterase.
Gene Therapy and Erectile Dysfunction

Gene therapy can be defined as the introduction of genetic material (ribonucleic acid [RNA] or deoxyribonucleic acid [DNA]) into an appropriate cell type, thus altering gene expression of that cell in order to produce a therapeutic effect. The goal of gene therapy for ED is to introduce novel genetic material into an appropriate cell in an attempt to restore normal cellular and physiologic function.

Gene therapy has been proposed as a viable treatment option for diseases that have a vascular origin, such as arteriosclerosis, congestive heart failure, and pulmonary hypertension. This, by biologic extension, suggests that gene therapy may also be employed to treat vascular diseases of the penis; ED in most cases is a manifestation of vascular disease.

One advantage of applying gene therapy for the treatment of ED is the easily accessible external location of the penis. Hence, a tourniquet can be placed around the base of the penis and the desired gene can be administered directly into the corpora cavernosa without entering the systemic circulation. Further, in the penis only a small number of cells need to be transfected because smooth muscle cells of the corpus cavernosum are interconnected by gap junctions that allow second messenger molecules and ions to be transferred to a number of interconnected smooth muscle cells, thus causing physiologically relevant changes in erectile function. Moreover, the vascular smooth muscle cells of the penis have a relatively low turnover rate, thus allowing a desired gene to be expressed for long periods of time.

The concept of gene therapy for ED treatment focuses on preventing cavernosal tissue degradation and increasing cavernosal smooth muscle tone. Smooth muscle relaxation is the necessary step for achieving a normal erection. Therefore, molecules, enzymes, or growth factors that influence the signal transduction pathway of corporal smooth muscle relaxation represent potential targets for ED gene therapy. Nitric oxide has been recognized to be the principal mediator of penile erection. However, other diverse mediators, such as the prostaglandins, VIP, and CGRP, play a role in erectile physiology.

Garban and associates first demonstrated that gene therapy could be performed in the penis by using naked complementary DNA (cDNA) encoding the penile inducible NOS gene, leading to physiologic benefit in the aging rat. Christ and colleagues later showed that injection of hSlo cDNA, which encodes the human smooth muscle maxi-K⁺ channel, into the rat corpora cavernosa can increase gap junction formation and enhance erectile responses to nerve stimulation in the aged rat. More recently, Bivalacqua and colleagues used an adenoviral gene transfer approach in which an adenoviral construct encoding the eNOS and CGRP genes were shown to reverse age-related ED in rats.

In these studies, cytomegalovirus and Rous sarcoma adenoviruses were utilized, and both eNOS and CGRP expression were sustained for at least 1 month in the corpora cavernosa of the rat penis. Five days after transfection with the AdCMVeNOS or AdRSVeNOS viruses, aged rats had significant increases in erectile function, as determined by cavernosal nerve stimulation and pharmacologic injection with the endothelium-dependent vasodilator acetylcholine and the type 5 PDE inhibitors zaprinast and sildenafil.

Lue and colleagues demonstrated that intracavernous injection of adeno-associated virus brain-derived neurotrophic factor could improve erectile function after cavernosal nerve injury. This neurotrophic factor purportedly restored neuronal NOS in the major pelvic ganglion, thus enhancing the recovery of erectile function after bilateral cavernous nerve injury. These early but innovative studies provide evidence that in vivo gene transfer can have beneficial physiologic effects on penile erection; this approach still requires a significant amount of basic research before in vivo gene therapy techniques can be applied to humans.
FEMALE SEXUAL DYSFUNCTION

Female sexual dysfunction is a multicausal and multidimensional medical problem that adversely affects physical health and emotional well-being. There has been a conspicuous paucity of basic science research on female sexuality. Thus, our knowledge and understanding of the anatomy and physiology of normal female sexual response and the pathophysiology of female sexual dysfunction are limited. On the basis of our understanding of the physiology of the male erectile response, recent advances in technology, and heightened interest in women's health, the study of female sexual function and dysfunction is evolving.

PREVALENCE AND INCIDENCE RATES

Sexual dysfunction in women is age-related, progressive, and highly prevalent, affecting 30% to 50% of American women. In the National Health and Social Life Survey, which included 1749 women, 43% of adult women had complaints of sexual dysfunction. Although this study had a large sample size, with minority representation, and used modern probability sampling, it was limited by its cross-sectional design. In addition, the NHLSLS did not include women over age 60 years and did not make any adjustment for menopausal status or medical risk factors.

Another study that surveyed 448 women over 60 years of age demonstrated that two thirds of these women were sexually inactive, 12% of married women had difficulty with intercourse, and 14% experienced pain with intercourse. Sexual activity was strongly correlated with marital status. Women over age 60 years were less likely to have sex if their partners were in poor health and if they had feelings of low self-worth.

The incidence rates of sexual dysfunction in women are not known. The same disease processes and risk factors that are associated with ED in men such as aging, hypertension, cigarette smoking, and hypercholesterolemia, are also associated with sexual dysfunction in women.
PHYSIOLOGY OF THE FEMALE SEXUAL RESPONSE CYCLE

Masters and Johnson first characterized the female sexual response as consisting of four successive phases: excitement, plateau, orgasm, and resolution. During sexual arousal, the clitoris and the labia minora become engorged with blood and vaginal and clitoral length and diameter increase. These authors observed that the labia minora increase in diameter by two to three times during sexual excitement and consequently become everted, exposing their inner surface.

Kaplan proposed the aspect of "desire," and the three-phase model, consisting of desire, arousal, and orgasm. In this model, desire is the factor that incites the overall response cycle. This three-phase model is the basis for the Diagnostic and Statistical Manual (DSM IV) definitions of female sexual dysfunction, and the recent reclassification system proposed by the American Foundation of Urologic Disease (AFUD) Consensus Panel in October 1998.

Others have suggested that sexual function should be viewed as a circuit, with four main domains: libido, arousal, orgasm, and satisfaction. Each of these four domains may overlap and feed back negatively or positively upon the other three domains.
PELVIC ANATOMY IN WOMEN

An understanding of female pelvic anatomy and physiology is essential for the evaluation and treatment of female sexual dysfunction. Although the female pelvis is composed of a continuum of organs interrelated in structure and function, it is helpful to group the pelvic organs into two categories: the external and internal genitalia.

The organs of the external genitalia are collectively known as the vulva, which is bound anteriorly by the symphysis pubis, posteriorly by the anal sphincter and laterally by the ischial tuberosities. The vulva consists of the labia, interlabial space, clitoris and vestibular bulbs.

The internal genitalia consist of the vagina, uterus, fallopian tubes, and ovaries.

Vagina

The vagina is a midline cylindrical organ that connects the uterus with the external genitalia. It usually measures 7 to 15 cm in length, depending on the position of the uterus. It can easily dilate and expand for intercourse and childbirth. Anteriorly, two pleats of sensitive tissue, the labia minora (inner lips), surround the opening of the vagina and are protected by the larger folds known as the labia majora (outer lips). The labia minora enclose an area called the vestibule, which contains the clitoris, the urethral opening and vaginal opening. The portion of the labia minora that covers the clitoris is known as the prepuce, or clitoral hood.

The wall of the vagina consists of three layers: (1) an inner aglandular mucous membrane epithelium; (2) an intermediary, richly supplied, vascular muscularis layer; and (3) an outer adventitial supportive mesh. Vaginal mucosa is a mucus type, stratified, nonkeratinized, squamous cell epithelium that undergoes hormone-related cyclical changes during the menstrual cycle. The middle muscularis layer is highly infiltrated with smooth muscle and an extensive tree of blood vessels, which engorges during sexual arousal. The surrounding outer fibrosa layer provides structural support to the vagina.

The vagina has many rugae, which are necessary for distensibility of the organ and are more prominent in the lower third of the vagina. Smaller ridges increase frictional tension during intercourse.

Blood Supply to the Vagina

The vascular system of the vagina consists of an extensive anastomotic network of blood vessels throughout its length. The main arterial supply to the vagina arises from vaginal branches of the uterine arteries, pudendal arteries (vaginal branches), and ovarian arteries.

Innervation of the Vagina

The autonomic innervation of the vagina originates from two separate plexuses: the superior hypogastric and the pelvic. Sympathetic fibers originate in the lateral gray column of T1-L2 and form the hypogastric plexuses. Parasympathetic fibers originate in the intermediolateral cell column of S2-S4 and synapse in the pelvic plexus. Sympathetic and parasympathetic nerve fibers leave the pelvic plexus and travel within the uterosacral and cardinal ligaments, along with the vessels, to supply the proximal two thirds of the vagina and the corporal bodies of the clitoris.

Somatic motor fibers originate from the anterior horns of sacral cord levels S2-S4 and travel within the pudendal nerve to innervate the bulbocavernosus and ischiocavernosus muscles. Sensory fibers innervating the introitus and perineum travel within the perineal and posterior labial nerves to the pudendal nerve (Fig. 10.3). Immunohistochemical studies have revealed an abundance of nerve fibers in the distal vagina, as compared with the more proximal part. This area of increased innervation, which plays an important role in sexual function, can be damaged during performance of bladder suspension procedures and vaginal hysterectomies.

Physiologic Changes in the Vagina during Sexual Arousal

During sexual arousal, genital vascongestion occurs as a result of increased blood flow. The vaginal canal is lubricated by secretions from uterine glands, and by a transudate that originates from the subepithelial vascular bed. This is passively transported through the subepithelial spaces, which are sometimes referred to as "intercellular channels." ENGORGEMENT of the vaginal wall raises pressure inside the capillaries and increases the transudation of plasma through the vaginal epithelium. This vaginal lubricative plasma flows through the epithelium onto the surface of the vagina, initially forming sweat-like droplets that coalesce to form a lubricative film that covers the vaginal wall.

Additional moistening during intercourse comes from secretions of the paired, greater vestibular or Bartholin’s glands, although some believe that these glands have a more primal function of emitting an odoriferous fluid to attract the male. In addition to its lubrication, the vagina lengthens and dilates during sexual arousal as a result of relaxation of the smooth muscle in its wall. In human and animal models, sexual stimulation results in increased vaginal blood flow and decreased vaginal luminal pressure.
Clitoris

The clitoris is an erectile organ similar to the penis and arises from the same embryologic structure, the genital tubercle. It is composed of three parts: the outermost glans or head, the middle corpus or body, and the innermost crura. The glans and body of the clitoris are 2 to 4 cm long, and the crura are 9 to 11 cm. The clitoris consists of fused midline erectile bodies (corpora cavernosa) that give rise to bilateral crura. The glans clitoris is visible as it emerges from the labia minora. The labia minora bifurcate to form the upper prepuce anteriorly and the lower frenulum posteriorly.

The clitoris consists of fused midline erectile bodies (corpora cavernosa) that give rise to bilateral crura. The glans clitoris is visible as it emerges from the labia minora.

Each corpus cavernosum is ensheathed by a thick, fibrous connective tissue structure, the tunica albuginea, that covers the lacunar sinusoids, which are surrounded by a trabecula of vascular smooth muscle and collagen fibers. The two separate crura of the clitoris, formed from the separation of the most proximal parts of the corpora in the perineum, attach bilaterally to the undersurface of the pubis along the ischiopubic rami.

The main arterial supply to the clitoris is via the iliohypogastric-pudendal arterial bed. With sexual stimulation, increased blood flow to the clitoral cavernosal arteries results in increased clitoral intracavernous pressure and tumescence and protrusion of the glans. Studies show that, unlike the penis, the clitoris lacks a subalbugineal layer between the erectile tissue and the tunica albuginea layer. In the male, this layer possesses a rich venous plexus that, during sexual excitement, expands against the tunica albuginea, reducing venous outflow and making the penis rigid. The absence of this venous plexus in the clitoris allows this organ to achieve tumescence, but not rigidity, during sexual arousal.

Duplex ultrasound of the clitoris reveals that during sexual simulation the clitoris increases in length and diameter and blood flow almost doubles.
Vestibular Bulb

The vestibular bulbs are 3-cm-long, paired structures that lie along the sides of the vaginal orifice, directly beneath the skin of the labia minora (Fig. 19-3A and B). Although they are homologous to the corpus spongiosum of the penis and are composed of vascular smooth muscle, they are distinct, in that they are separated from the clitoris, urethra, and vestibule of the vagina. Recent cadaver dissections reveal that in young premenopausal women, the bulbs lie on the superficial aspect of the vaginal wall and do not form the core of the labia minora. Furthermore, there are considerable age-related variations in the dimensions of the erectile tissue between young premenopausal specimens and older, postmenopausal specimens. (118)

The main arterial supply to the vestibular bulbs is via bulbar, inferior perineal, and posterior labial branches of the internal pudendal artery (see Fig. 19-3). The somatic sensory innervation of the labia minora travels via the perineal and posterior labial branches of the pudendal nerve. Autonomic innervation consists of sympathetic and parasympathetic fibers that travel with the vessels to reach the vestibular bulb.

In the labia minora, blood flow increases during sexual stimulation, particularly to the vestibular bulbs. This causes a twofold to threefold increase in diameter and eversion of the labia with exposure of its inner surface.
Uterus

The uterine and cervical glands secrete mucus during sexual arousal to lubricate the vagina. Uterine and other pelvic surgical procedures can have a significant impact on female sexual response and function. The innervation of the uterus is closely proximate to the bladder and vagina, and pelvic dissection, as it is currently performed, can adversely affect a woman's later sexual health by damaging the uterine innervation. Surgical menopause brought on by hysterectomy and oophorectomy affects sexual function by multiple mechanisms. Even hysterectomy alone, without the removal of the ovaries, can result in sexual dysfunction. Symptoms that women commonly experience postoperatively after hysterectomy include decreased desire, decreased arousal, decreased genital sensation, and orgasmic dysfunction.

The anatomic and physiologic basis for sexual dysfunction after hysterectomy is poorly understood. Understanding of female neurovascular anatomy is limited but vital to the understanding of normal sexual arousal and function. The sexual dysfunction symptoms women experience after hysterectomy are probably the result of nerve or vascular injuries as well as the loss of ovarian estrogens and androgens. Removal of the uterus and ligation of the arterial supply at the uterine pedicles can result in ovarian atrophy as well as fibrosis of vaginal wall and clitoral cavernosal smooth muscle. Disruption of the pelvic autonomic and cervical plexuses as well as the uterosacral and cardinal ligaments, which are associated with autonomic nerve fibers, results in difficulties in genital arousal and orgasm. This is an area of current focus and research, and in the future, perhaps women will be offered nerve-sparing pelvic procedures similar to those now routinely performed in men.
Pelvic Floor Muscles

The pelvic floor is a collection of tissues that span the opening within the bony pelvis. In addition to supporting the abdominal and pelvic organs and maintaining continence of urine and stool, the pelvic floor allows for intercourse and parturition. The pelvic floor musculature in particular, the pelvic diaphragm that is formed by the levator ani muscles, the urogenital diaphragm, and the perineal membrane is important for pelvic support. The perineal membrane that consists of the ischiocavernosus, bulbocavernosus, and superficial transverse perineal muscles is closely related to the vestibular bulbs and clitoris and plays a role in sexual response. These muscles, when voluntarily contracted, can intensify orgasm of both the female and her mate partner.

The levator ani muscle has two different parts, the pubococcygeus and the iliococcygeus. These muscles can be palpated during pelvic examination as a distinct ridge just above the hymenal ring along each lateral wall of the pelvis. The function of this group of muscles is to pull the rectum, vagina, and urethra anteriorly toward the pubic bones to compress the lumen closed. Nonvoluntary pelvic floor spasm associated with vaginal penetration, or even examination with a speculum, is referred to as vaginismus. This disorder prevents sexual intercourse and is associated with dyspareunia and other sexual pain disorders. If the opposite problem exists, consisting of laxity and hypotonia of the pelvic floor musculature due to aging, menopause, or childbirth, for instance, symptoms of vaginal hypoanesthesia, coital anorgasmia, as well as incontinence during sexual intercourse or orgasm can develop. Women with pelvic floor disorders often have coexisting urologic and sexual dysfunction complaints. As a result of this overlap, all women who present with voiding dysfunction should be questioned about their sexual function as well.
NEUROGENIC MEDIATORS OF FEMALE SEXUAL RESPONSE

Within the central nervous system, the medial preoptic, anterior hypothalamic region and related limbic-hippocampal structures are responsible for sexual arousal. Upon activation, these centers transmit electrical signals through the parasympathetic and sympathetic nervous system. The neurogenic mechanisms modulating vaginal and clitoral smooth muscle tone, and vaginal and clitoral vascular smooth muscle relaxation are currently under investigation.

Noradrenergic, Noncholingeric (NANC)-Mediated Responses

Immunohistochemical studies in human vaginal tissues have shown the presence of nerve fibers containing neuropeptide Y (NPY), VIP, NOS, CGRP and substance P. Preliminary studies suggest that VIP and nitric oxide are involved in modulating vaginal relaxation and secretory processes. In the clitoris, nitric oxide has been identified in human tissue and is thought to be the primary mediator of clitoral and labial engorgement. Organ bath analysis of rabbit clitoral cavernosal smooth muscle strips demonstrates enhanced relaxation in response to sodium nitroprusside and L-arginine (both nitric oxide donors), which supports this hypothesis (Berman et al, unpublished data). Recently, PDE type 5, the enzyme responsible for degradation of cGMP, has been isolated in human clitoral, vestibular bulb, and vaginal smooth muscle culture and is inhibited by sildenafil citrate. Human and rabbit vaginal smooth muscle cells, treated with the nitric oxide donor sodium nitroprusside, in the presence of sildenafil have enhanced intracellular cGMP synthesis and accumulation. PGE1 and forskolin also produce a marked increase in intracellular cGMP.

In organ bath studies, sildenafil causes dose-dependent relaxation of female rabbit clitoral and vaginal smooth muscle strips (Berman), further suggesting a role for nitric oxide as a mediator of clitoral cavernosal and vaginal wall smooth muscle relaxation. The exact identity of the relaxant NANC neurotransmitter, however, remains unclear. VIP is a NANC neurotransmitter that, like nitric oxide, may play a role in enhancing vaginal blood flow, lubrication, and secretions. The vagina is heavily innervated with VIP-immunoreactive nerve fibers in close relation to the epithelium and blood vessels. In organ bath studies, VIP also causes dose-dependent relaxation of rabbit clitoral cavernosal and vaginal smooth muscle tissue, suggesting a similar role for endogenous VIP as a NANC neurotransmitter in clitoral and vaginal tissues (Berman et al, unpublished data).
Alpha₁-Adrenergic and Alpha₂-Adrenergic Responses

In men, adrenergic receptors exist in the brain centers and are associated with penile erection, libido, and ejaculation. Agents that affect these receptors have been extensively studied and used in the treatment of male ED. -Adrenergic agonists such as norepinephrine activate sympathetic nerve terminals, resulting in contraction of penile trabecular smooth muscle and detumescence. In addition, α₂ agonists cause similar responses. Activation of β receptors results in a decrease in intracellular adenosine 3′5′-monophosphate concentrations and in potent contraction in blood vessels. -Adrenergic mediators also appear to play a physiologic role in female sexual arousal.

Preliminary organ chamber experiments using rabbit vaginal tissue suggest that adrenergic mechanisms modulate smooth muscle tone. Exogenous norepinephrine (α₁ and α₂ agonist) causes dose-dependent contraction of vaginal smooth muscle and α₁ (prazosin and tamsulosin) and α₂ (delequamine) selective antagonists inhibit the contraction (Berman et al, unpublished data). These observations suggest that adrenergic nerves mediate contractile response. Furthermore, there appears to be a difference in the quality of the contractile responses in upper and lower vaginal segments, which is consistent with their different innervation and embryologic origin.
HORMONAL REGULATION: THE ROLES OF ESTROGEN AND TESTOSTERONE

Estrogen

Estrogen plays a significant role in regulating female sexual function. Estradiol affects cells throughout the peripheral and central nervous systems and influences nerve transmission. A decline in serum estrogen levels results in thinning of vaginal mucosal epithelium and atrophy of vaginal wall smooth muscle. Decreased estrogen levels also result in a less acidic environment in the vaginal canal. These changes, associated with estrogen deficiency, can predispose women to vaginal infections, urinary tract infections, incontinence, and sexual dysfunction.

In animal models, estradiol administration results in expanded touch receptor zones along the distribution of the pudendal nerve, suggesting that estrogen affects sensory thresholds. In postmenopausal women, estrogen replacement restores clitoral and vaginal vibration and pressure thresholds to levels close to those of premenopausal women. Estrogens also have vasoprotective and vasodilatory effects, which result in increased vaginal, clitoral, and urethral arterial flow, resulting in maintenance of the female sexual response by preventing atherosclerotic compromise to pelvic arteries and arterioles.

With menopause and the decline in circulating estrogen levels, a majority of women experience some degree of change in sexual function. Common sexual complaints include loss of desire, decreased frequency of sexual activity, painful intercourse, diminished sexual responsiveness, difficulty achieving orgasm, and decreased genital sensation. Masters and Johnson first published their findings of the physiologic changes occurring in menopausal women that related to sexual function in 1966. We have since learned that symptoms related to alterations in genital sensation and blood flow are, in part, secondary to declining estrogen levels and that there is a direct correlation between the presence of sexual complaints and levels of estradiol below 50 pg/mL. Some of the symptoms associated with menopause, such as vaginal dryness and diminished genital sensation, improve with estrogen replacement. However, some postmenopausal women continue to experience lack of sexual interest, fantasy and satisfaction, flat mood, and diminished well-being despite appropriate estrogen replacement; these symptoms are often attributed to androgen deficiency. As discussed later, it remains unclear whether physiologic testosterone replacement can improve sexual function in older women or whether this symptom complex is even due to androgen deficiency.

Estrogen also plays a role in regulating vaginal and clitoral NOS expression, the enzyme responsible for the production of nitric oxide. Aging and surgical castration result in decreased vaginal and clitoral NOS expression and apoptosis of vaginal wall smooth muscle and mucosal epithelium. Estrogen replacement restores vaginal mucosal health, increases vaginal NOS expression, and decreases vaginal mucosal cell death. These findings suggest that medications such as sildenafil, which mediate vascular and nonvascular smooth muscle relaxation via nitric oxide, may have a potential role in the treatment of female sexual dysfunction, in particular that associated with sexual arousal disorder.
Testosterone deficiency in the pathophysiology of female sexual dysfunction remains controversial. Most commercial assays for the measurement of total and free testosterone levels were developed to measure the much higher circulating concentrations in men; these assays lack the sensitivity and precision required to measure the low levels prevalent in androgen-deficient women. Because of the paucity of normative data on serum total and free testosterone concentrations in healthy, menstruating women, there is no consensus on the thresholds that can be used to define androgen deficiency in women. Until recently, the available testosterone formulations were designed to deliver the much higher dose required in hypogonadal men. However, two new formulations, the testosterone transdermal matrix patch for women and the testosterone gel are now undergoing phase 1 and phase 2 studies. There is agreement that androgens, by acting within the brain, influence sexual behavior, although a woman’s libido is also determined by environmental, emotional, cultural, and hormonal factors. The effects of androgens in the brain are mediated, in part, directly through the androgen receptor and through aromatization of testosterone to estradiol. Androgen receptors have been identified in the cortex, pituitary, hypothalamus, preoptic region, thalamus, amygdala, and brain stem. In addition, both testosterone and estrogen have been shown to exert nongenomic effects in the central nervous system.

Testosterone might have additional effects on the genitalia in women. Androgen receptors have been reported in the vaginal wall, leading to speculation that testosterone might induce vaginal smooth muscle relaxation. We do not know whether NOS activity in the clitoris is under testosterone regulation as it is in the cavernosal smooth muscle in the penis.
CLASSIFICATIONS AND DEFINITIONS OF FEMALE SEXUAL DYSFUNCTION (1998)

The Sexual Function Health Council of the American Foundation for Urologic Disease convened the AFUD Consensus Panel, an interdisciplinary conference panel, consisting of a multinational group of experts in female sexual dysfunction. The panel included specialists from many relevant disciplines, including endocrinology, family medicine, gynecology, nursing, pharmacology, physiology, psychiatry, psychology, rehabilitation medicine, and urology. The panel's objective was to evaluate and revise existing definitions and classifications of female sexual dysfunction. Specifically, medical risk factors and etiologic mechanisms of female sexual dysfunction were incorporated with the preexisting psychologically based definitions to generate an updated, unified classification system.

The AFUD Consensus Panel classified female sexual dysfunction into several categories, described next (Table 19-7).

Hypoactive Sexual Desire Disorder

Hypoactive sexual desire disorder refers to the persistent or recurring deficiency (or absence) of sexual fantasies or thoughts and/or desire for or receptivity to sexual activity that causes personal distress. This disorder may result from psychological or emotional factors or may be secondary to physiologic problems (e.g., hormone deficiency) and medical or surgical interventions. Any disruption of the female hormonal milieu caused by natural menopause, surgically or medically induced menopause, or endocrine disorders can result in inhibited sexual desire.

<table>
<thead>
<tr>
<th>TABLE 19-7 -- AFUD Classification of Female Sexual Dysfunction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Hypoactive Sexual Desire Disorder</td>
</tr>
<tr>
<td>2. Sexual Aversion Disorder</td>
</tr>
<tr>
<td>3. Sexual Arousal Disorder</td>
</tr>
<tr>
<td>4. Orgasmic Disorder</td>
</tr>
<tr>
<td>5. Sexual Pain Disorders</td>
</tr>
<tr>
<td>a. Dyspareunia</td>
</tr>
<tr>
<td>b. Vaginismus</td>
</tr>
<tr>
<td>c. Other Sexual Pain Disorders</td>
</tr>
</tbody>
</table>

In 1998, the Sexual Function Health Council of the American Foundation for Urologic Disease convened a consensus panel that proposed a new classification for female sexual dysfunction. (see text for definitions).
Sexual Aversion Disorder

Sexual aversion disorder is the persistent or recurrent phobic aversion to and avoidance of sexual contact with a sexual partner that causes personal distress. This disorder is generally a psychologically or emotionally based problem that can result from experiences such as physical or sexual abuse, or childhood trauma.
Sexual Arousal Disorder

Sexual arousal disorder refers to the persistent or recurring inability to attain or maintain adequate sexual excitement that causes personal distress. It may be experienced as lack of subjective excitement or lack of genital (lubrication or swelling) or other somatic response.

Disorders of arousal include, but are not limited to, lack of or diminished vaginal lubrication, decreased clitoral and labial sensation, decreased clitoral and labial engorgement, or lack of vaginal smooth muscle relaxation. These conditions may occur as a result of psychological factors; however, often there is a medical/physiologic basis, such as diminished vaginal or clitoral blood flow, previous pelvic trauma, pelvic surgery, or medications.
Orgasmic Disorder

The orgasmic disorder refers to the persistent or recurrent failure, difficulty, or delay in attaining orgasm following adequate sexual stimulation and arousal that causes personal distress. This may be a primary condition (never achieved orgasm) or a secondary condition that may occur as a result of surgery, trauma, or hormone deficiency. Primary anorgasmia can be secondary to emotional trauma or sexual abuse; however, other medical problems as well as medications (e.g., serotonin reuptake inhibitors) can contribute to or exacerbate the problem.
Sexual Pain Disorders

Three sexual pain disorders are described:

1. **Dyspareunia.** Dyspareunia refers to the recurrent or persistent genital pain associated with sexual intercourse. Dyspareunia can develop secondary to medical problems such as vestibulitis, vaginal atrophy, or vaginal infection. It can be either physiologically or psychologically based, or due to a combination of the two.

2. **Vaginismus.** Vaginismus refers to the recurrent or persistent involuntary spasms of the musculature of the outer third of the vagina that interferes with vaginal penetration, and causes personal distress. Vaginismus usually develops as a conditioned response to painful penetration, or secondary to psychological and emotional factors.

3. **Other sexual pain disorders.** Recurrent or persistent genital pain induced by noncoital sexual stimulation can be caused by anatomic and inflammatory conditions, including infections (e.g., HSV), vestibulitis, prior genital mutilation, trauma, and endometriosis.

Each of these categories can be further subtyped according to whether the disorder has been lifelong or acquired; generalized or situational; and organic, psychogenic, or mixed in its pathophysiology. The etiology of any of these disorders may be multifactorial, and often the disorders overlap. At present, sexual arousal disorder is the focus of clinical and basic science research as well as treatment interventions. Human and animal models have been established to assess sexual arousal and validated sexual arousal instruments are now available.

An emerging hypothesis is that decreased sexual arousal, as manifested by diminished genital engorgement, may be related to inadequate pelvic arterial blood flow. Decreased genital blood flow ultimately leads to fibrosis of vaginal wall and clitoral cavernosal smooth muscle. Women with sexual arousal disorder complain of diminished vaginal lubrication, pain with intercourse, decreased vaginal and clitoral sensation, and difficulty achieving orgasm. Because different domains of sexual function in women are interlinked, enhancement of sexual arousal by improvements in vaginal lubrication, clitoral and labial engorgement, and genital sensation would be expected to secondarily improve orgasmic ability as well as libido. The potential role of pharmacotherapy for the treatment of female sexual arousal complaints remains poorly understood.
ETIOLOGY OF FEMALE SEXUAL DYSFUNCTION

The causes of sexual dysfunction in women are complex and multifactorial. Even in women with a known organic disease, additional medical, psychosexual, and relational factors are operative that complicate the management of these patients. Similarly, although the vasculogenic, neurogenic, psychogenic, and hormonal factors are discussed separately to highlight their contributions to the pathophysiology of sexual dysfunction, in reality it is often difficult to clinically establish causal roles of vasculogenic and neurogenic factors in individual patients. Furthermore, hormone deficiencies may secondarily lead to anatomic, vascular, and neural changes that, in turn, may cause sexual aversion because of pain.

Vasculogenic Causes

High blood pressure, high cholesterol levels, smoking, and heart disease are associated with ED in men and with sexual dysfunction in women. The recently named clitoral and vaginal vascular insufficiency syndromes are directly related to diminished genital blood flow secondary to atherosclerosis of the iliohypogastric/pudendal arterial bed. Traumatic injury to this structure from pelvic fractures, blunt trauma, surgical disruption, or chronic perineal pressure from bicycle riding can result in diminished vaginal and clitoral blood flow. Histomorphometric evaluation of clitoral erectile tissue from atherosclerotic animals demonstrates clitoral cavernosal artery wall thickening, loss of corporal smooth muscle, and increase in collagen deposition.

In human clitoral tissue, there is a similar loss of corporal smooth muscle with replacement by fibrous connective tissue in association with atherosclerosis of clitoral cavernosal arteries. It is possible that the atherosclerotic changes that occur in clitoral vascular and trabecular smooth muscle may interfere with normal relaxation and dilation responses to sexual simulation.
Neurogenic Causes

The same neurogenic disorders that cause ED in men can also cause sexual dysfunction in women. These include spinal cord injury; diseases of the central or peripheral nervous system, including diabetes; and complete upper motor neuron injuries that affect the sacral spinal segments. Women with incomplete injuries retain that capacity for psychogenic arousal and vaginal lubrication. Women with spinal cord injury have significantly more difficulty in achieving orgasm than do normal controls.
Hormonal Causes

Dysfunction of the hypothalamic-pituitary axis, surgical or medical oophorectomy, premature ovarian failure, old age, and chronic birth control use are associated with reductions in free testosterone levels and with sexual dysfunction in some women. However, a direct cause-and-effect relationship between hormonal deficiency and sexual dysfunction is difficult to establish because of the coexistence of a number of complex, confounding factors. Estrogen deprivation causes a significant decrease in the clitoral intracavernosal, vaginal, and urethral blood flow. Thus, a decline in circulating estrogen levels can produce significant adverse effects on structure and function of the vaginal and clitoral tissues, including diffuse clitoral fibrosis, thinned vaginal epithelial layers, and decreased vaginal submucosal vasculature; these changes can affect sexual function. Conversely, estrogen administration in postmenopausal women and in women after oophorectomy improves the integrity of vaginal mucosal tissue and has beneficial effects on vaginal sensation, vasocongestion, and secretions, which lead to enhanced arousal.

Although oophorectomized women have often been the subjects in studies of the role of sex steroids in the pathophysiology of female sexual dysfunction, there are several inherent limitations of the use of this patient population. Most of these patients have some underlying pelvic disease that led to oophorectomy in the first place. In most women who have undergone oophorectomy, the uterus has been removed. Hysterectomy can be associated with substantial changes in pelvic and vaginal anatomy that can affect sexual responses. If oophorectomized women do not receive estrogen replacement therapy, they experience vaginal atrophy that can secondarily affect sexual function. It should be recognized that correction of the pelvic pathology by surgical oophorectomy and hysterectomy can, in some women, improve sexual function. For instance, in one retrospective study, a significant proportion of women reported improved sexual function after hysterectomy, although women whose ovaries had been removed were less likely to report improvement than women whose ovaries had not been removed.

The decline in serum testosterone concentrations begins in the fourth decade and is age-related; thus, serum testosterone concentrations in a 60-year old woman, on average, are about half those in a 30-year old woman. In contrast to the rapid decline in estrogen production that occurs at menopause, however, serum testosterone concentrations do not fall abruptly during the menopausal transition. A large epidemiologic study demonstrated that serum testosterone levels either do not change or may even increase in the perimenopausal years. Using a scale that determines well-being and sexual function, these investigators found that 83% of postmenopausal women had scores indicating sexual dysfunction; however, sexual dysfunction did not correlate with androgen levels. Therefore, sexual dysfunction in healthy, postmenopausal women cannot be attributed to androgen deficiency alone.

Because the postmenopausal ovary continues to produce androgens, oophorectomy significantly lowers androgen levels in premenopausal as well as in postmenopausal women.

Estrogen administration reduces free testosterone levels by multiple mechanisms. Estrogen therapy reduces trophic luteinizing hormone (LH) secretion, thus decreasing ovarian androgen production; in addition, estrogens stimulate hepatic sex hormone-binding globulin (SHBG) production, resulting in lower free testosterone levels.
Musculogenic Causes

The pelvic floor muscles, in particular the levator ani and perineal membrane, participate in female sexual function and responsiveness. The perineal membrane, consisting of the bulbocavernosus and ischiocavernosus muscles, when voluntarily contracted, contributes to and intensifies sexual arousal and orgasm. In addition, the bulbospongiosus and ischiocavernosus muscles are responsible for the involuntary rhythmic contractions during orgasm. The levator ani also modulates motor responses during orgasm as well as vaginal receptivity. A hypertonic levator ani can contribute to the development of vaginismus, causing dyspareunia and other sexual pain disorders. When the levator ani is hypotonic, vaginal hypoanesthesia, coital anorgasmia, and urinary incontinence can develop during sexual intercourse or orgasm.
Psychogenic Causes

In women, despite the presence or absence of organic disease, emotional and relational issues significantly affect sexual arousal. Self-esteem, body image, and the quality of the relationship with partner affect a woman’s ability to respond sexually. In addition, depression and other mood disorders are often associated with female sexual dysfunction. Furthermore, the medications commonly used to treat depression can significantly affect the female sexual response.

The most frequently used medications for uncomplicated depression are the serotonin reuptake inhibitors (SSRIs). Women receiving these medications often complain of decreased desire, reduced arousal and genital sensation, and difficulty achieving orgasm. Several studies have documented improvements in SSRI-induced sexual dysfunction in women after administration of sildenafil.
CLINICAL EVALUATION OF THE FEMALE SEXUAL RESPONSE

In the clinical setting, female sexual responses have been difficult to quantify objectively. Female sexual arousal results in a combination of vasocongestive and neuromuscular events, which include increased clitoral, labial, and vaginal wall engorgement as well as increased vaginal luminal diameter and lubrication. Muscle tension, respiratory rate, heart rate, and blood pressure steadily rise during arousal, finally reaching their peak during orgasm. In contrast to the male erectile response, genital changes that occur during female sexual arousal not only are difficult to measure but also are often not readily visible or recognized by the patient. Previously described techniques for evaluating physiologic changes during female sexual arousal have been performed primarily by psychotherapists and physiologists and have focused primarily on estimating vaginal engorgement. The clinical utility of these physiologic measurements remains unknown.

Medical Evaluation

A comprehensive approach to the evaluation of physiologic changes during sexual arousal involves a full history and physical examination, including a pelvic examination and, when clinically indicated, a hormonal profile (i.e., follicle-stimulating hormone [FSH], total testosterone). Increased FSH levels can help document ovarian failure and menopausal status and can aid in differentiating primary ovarian dysfunction from disorders of the hypothalamus and pituitary gland. Measurements of LH levels in women do not add a great deal of information to that obtained from FSH measurements but can provide confirmatory data. Total testosterone levels, measured by an assay that has adequate sensitivity and precision, are needed to determine whether sexual dysfunction is due to androgen deficiency.

Estradiol levels vary greatly in women, depending on the phase of the menstrual cycle; very low levels can confirm ovarian failure but provide little additional information beyond what can be inferred from menstrual history and FSH levels. In women with hypogonadotropic hypogonadism, measurement of serum prolactin and a magnetic resonance imaging (MRI) study of the hypothalamic-pituitary region may be needed to exclude the presence of a space-occupying lesion. In addition, these patients may need evaluation of other pituitary hormones to exclude deficiencies of other trophic hormones.

It is important to exclude systemic illness and medical conditions that disrupt the hypothalamic-pituitary axis such as HIV infection, end-stage renal disease, chronic obstructive lung disease, menopause, prior chemotherapy or bilateral salpingo-oophorectomy. Complete blood counts and a chemistry panel should be part of the general health evaluation. Medications that adversely affect sexual function should be addressed and changed if possible (Table 19-8).

Evaluating the female sexual response in the clinical setting both validates the patient's problem and potentially aids in diagnosis of organic disease, such as vascular insufficiency, a hormonal abnormality, or a neurologic disorder. Current investigations seek to define ranges of normal and the possible relevance to therapeutics for the following parameters:

1. Genital blood flow (clitoral, labial, urethral, and vaginal peak systolic velocities and end-diastolic velocities can be measured with duplex Doppler ultrasound).
2. Vaginal lubrication, as reflected by pH.
3. Vaginal compliance and elasticity (pressure-volume changes).
4. Genital sensation and vibration perception thresholds.
Psychosocial and Psychosexual Assessment

In addition to the medical evaluations, all patients should be evaluated for emotional and relational issues. This includes the context in which the patient experiences her sexuality, her self-esteem and body image, and her ability to communicate her sexual needs with her partner. This is an integral component of the female sexual function evaluation. Emotional and/or relational issues should be addressed before treatment and, certainly, before treatment efficacy is determined. If ongoing therapy is desired or required, it should also be provided.

Several instruments are available for assessment of self-reported sexual function, in particular sexual arousal. These self-report measures can be useful in clinical screening of patients and in clinical trials.

The Brief Index of Sexual Function Inventory (BISF-W), for example, is a validated 21-item self-reported listing of sexual interest, activity, satisfaction, and preference and discriminates between depressed, sexually dysfunctional, and healthy patients. The BISF-W is sensitive to the effects of androgen replacement in oophorectomized women.

The Female Sexual Function Index has also been used in clinical studies and practice. Its self-report measures have been found to have a high degree of reliability and validity. This instrument can discriminate between women with sexual dysfunction and age-matched controls.

### TABLE 19-8 – Medications That Adversely Affect Sexual Function (Desire, Arousal, and Orgasm)

<table>
<thead>
<tr>
<th>1. Antihypertensive agents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clonidine, guanethidine, reserpine, propranolol, prazosin, spironolactone, thiazide diuretics, beta blockers, calcium channel blockers</td>
</tr>
<tr>
<td>2. Antidepressants</td>
</tr>
<tr>
<td>Imipramine, amitryptiline, clomipramine, amoxapine, monoamine oxidase inhibitors, lithium carbonate</td>
</tr>
<tr>
<td>3. Selective serotonin reuptake inhibitors (SSRIs)</td>
</tr>
<tr>
<td>Fluoxetine, sertraline, paroxetine, venlafaxine</td>
</tr>
<tr>
<td>4. Anxiolytics/sedative-hypnotics</td>
</tr>
<tr>
<td>Diazepam, alprazolam</td>
</tr>
<tr>
<td>5. Neuroleptics</td>
</tr>
<tr>
<td>Phenothiazines, butyrophenones, thioridazine, chlorpromazine, chlorprothixene, fluphenazine</td>
</tr>
<tr>
<td>6. Anticonvulsants</td>
</tr>
<tr>
<td>Phenobarbital, tegretol, primidone, phenytoin</td>
</tr>
<tr>
<td>7. Antiluder drugs</td>
</tr>
<tr>
<td>Cimetidine</td>
</tr>
<tr>
<td>8. Anticancer drugs</td>
</tr>
<tr>
<td>Procarbazine, busulfan, chlorambucil, cyclophosphamide, tamoxifen, raloxifen</td>
</tr>
</tbody>
</table>

Daily diaries and sexual event logs have also been used in clinical trials. Subjective sexual response data reflect the personal experience of the patient, an important variable to evaluate because the ultimate goal is to enhance the personal sexual experience of the woman. Intervention is not considered successful unless the woman is able to experience sexual arousal, pleasure, and satisfaction. Thus, it is important to determine whether physiologic changes or improvement in blood flow translate into a better sexual experience.
TREATMENT

Treatment of female sexual dysfunction is gradually evolving as more clinical and basic research studies are targeting this problem. Medical management remains in the early experimental phases. Most treatment modalities being used by patients and physicians are based on empirical observations and are neither supported by rigorous clinical trial data nor approved by the U.S. Food and Drug Administration. Although it is premature to make general evidence-based recommendations for the treatment of female sexual dysfunction, it is important to recognize that not all female sexual complaints are psychological and that there are potential therapeutic options that might be useful in individual patients.

Estrogen Therapy

Estrogen replacement therapy is often indicated in menopausal women (either spontaneous or surgical). Aside from relieving hot flashes and preventing osteoporosis, estrogen therapy probably results in improved clitoral sensitivity and in decreased pain and burning during intercourse. Although cross-sectional epidemiologic studies had previously reported lower prevalence of cardiovascular disease among estrogen users, a recent randomized placebo-controlled trial was unable to demonstrate beneficial effects of estrogen replacement in secondary prevention of heart disease in postmenopausal women.

The role of estrogen replacement in the primary and secondary prevention of heart disease in women remains unclear and is the focus of ongoing investigation. In women with established heart disease, estrogen replacement has not been shown to improve cardiovascular outcomes. In the HERS study, event rates were higher in women given estrogen replacement than in women given placebo. In menopausal women or in oophorectomized women, local or topical estrogen application relieves symptoms of vaginal dryness, burning, and urinary frequency and urgency. A vaginal estradiol ring (Estring) delivers low-dose estrogen locally and may benefit breast cancer patients and other women who are unable to take oral or transdermal estrogen.
Androgen Formulations

Testosterone supplementation is associated with increased well-being, energy, and appetite and with improved somatic and psychological scores in surgically menopausal women. In one study, such women, who were given supraphysiologic doses of testosterone enanthate by intramuscular injection alone or in combination with estrogen, experienced a greater increase in sexual desire, fantasies, and arousal than women who received estrogen alone. In another study, combined administration of testosterone and estradiol implants increased sexual activity, satisfaction, pleasure, and frequency of orgasm more than estrogen implants alone. The dose of testosterone used in each of these studies was supraphysiologic.

A later study evaluated the effects of physiologic testosterone replacement in women who had undergone hysterectomy and oophorectomy and who were receiving stable estrogen replacement (at least 0.625 mg of conjugated equine estrogen daily orally). These women were randomized to receive either placebo patches or testosterone patches designed to achieve nominal delivery of either 150 µg (one active and one placebo patch) or 300 µg (two active patches) of transdermal testosterone daily for 12 weeks each. The highest dose of testosterone resulted in a mean serum free testosterone slightly above the physiologic range and significantly increased scores for frequency of sexual activity, pleasure, and orgasm. It also increased sexual fantasies, masturbation, and positive well-being.

Objectively, testosterone supplementation has been reported to increase vaginal vascongestion, as measured by vaginal plethysmography during exposure to a potent visual stimulus in a small number of women with hypothalamic amenorrhea. Dehydroepiandrosterone (DHEA) replacement of 50 mg/day for 4 months in women with adrenal insufficiency increased sexuality and well-being. It is unclear whether these effects were secondary to direct effect of DHEA on the brain or to an indirect effect via an increase in androgen synthesis. In contrast, a cross-sectional study of perimenopausal women did not show correlation between sexual function and androgen levels. Although pharmacologic doses of testosterone undoubtedly improve overall sexual function, we do not know whether physiologic testosterone replacement can produce clinically meaningful changes in health-related outcomes.

All androgens carry the risk of inducing virilization in women. Early reversible manifestations include acne, oiliness of skin, hirsutism, and menstrual irregularities. Long-term side effects such as male-pattern baldness, hirsutism, voice changes, and hypertrophy of the clitoris are largely irreversible. There is also evidence that testosterone supplementation in supra-physiologic doses decreases high-density-lipoprotein (HDL) cholesterol levels. Women with history of breast cancer should not be prescribed testosterone because testosterone is converted to estrogen by the aromatase enzyme. Surprisingly, the reported prevalence of virilizing side effects, even with the use of supraphysiologic doses of testosterone in surgically menopausal women, has been relatively low.

17-Methyltestosterone is commonly used in combination with estrogen in menopausal women for sexual dysfunction. There are conflicting reports regarding the benefit of methyltestosterone for treatment of inhibited desire or vaginismus in premenopausal women. The suggested dose of methyltestosterone for premenopausal and postmenopausal women ranges from 0.25 to 1.25 mg/day. The dose should be adjusted according to symptoms, free testosterone levels, and cholesterol, triglyceride HDL levels, and liver function tests. The potential side effects include weight gain, clitoral enlargement, increased facial hair, and hypercholesterolemia. The pharmacokinetics and clinical efficacy of this formulation have not been rigorously studied.

Two novel testosterone formulations are currently in early phases of clinical development. A transdermal matrix testosterone delivery system (patch) for women that is applied twice a week is being investigated. Each patch provides a nominal delivery of 150 µg of testosterone daily. Thus, two patches applied simultaneously twice a week can achieve a daily delivery of 300 µg of testosterone, approximating the daily production rates of testosterone in healthy women. The skin tolerability of the patch is excellent. Initial pharmacokinetic studies have demonstrated that each 150-µg patch increases serum total testosterone concentrations by 25 to 30 ng/dL on average in healthy menstruating women and in surgically postmenopausal women. A testosterone gel for women is in initial stages of development. Each milligram of testosterone applied to the nongenital skin can increase serum testosterone concentration by 7 to 8 ng/dL.

Data from clinical trials thus agree that supraphysiologic doses of testosterone improve several aspects of sexual function in oophorectomized women with sexual dysfunction; however, we do not know whether physiologic testosterone that raises serum testosterone levels from low normal to high end of the normal range in women can produce clinically meaningful improvements in sexual function in postmenopausal women.
Sildenafil

Functioning as a selective type 5 (cGMP-specific) PDE inhibitor, this medication decreases the catabolism of cGMP, the second messenger in nitric oxide-mediated relaxation of clitoral and vaginal smooth muscle. Sildenafil may prove useful alone or perhaps in combination with other vasoactive substances for treatment of female sexual dysfunction. Phase II clinical studies assessing safety and efficacy of this medication for use in women are now in progress. Two studies have demonstrated that sildenafil is successful in treating female sexual dysfunction associated with aging and menopause and secondary to SSRI use. [137][166]
Therapeutic Agents under Development

A number of drugs that have previously been used for the treatment of ED in men are now being investigated in women with sexual dysfunction. In the absence of efficacy data from randomized clinical trials, it is premature to make general recommendations about their use at this time.

L-Arginine functions as a precursor to the formation of nitric oxide, which mediates the relaxation of vascular and non-vascular smooth muscle. Although L-arginine has not been tested in clinical trials in women, preliminary studies in men appear promising. A combination of L-arginine and yohimbine (an α2-adrenergic blocker) is currently being investigated for use in women.

An intraurethral formulation of PGE1, absorbed via the mucosa (MUSE), is now available for male patients. A similar application of PGE1, delivered intravaginally, is under investigation for use in women.

Currently available in an oral preparation, phentolamine functions as a nonspecific α-adrenergic blocker and causes vascular smooth muscle relaxation. A pilot study in menopausal women with sexual dysfunction demonstrated enhanced vaginal blood flow and subjective arousal with the medication.

Apomorphine is a short-acting dopamine agonist that facilitates erectile responses in both normal men and men with ED. This drug is being tested in women with sexual dysfunction.
SUMMARY

The ideal approach to female sexual dysfunction is a collaborative effort between therapists and physicians and should include a complete medical and psychosocial evaluation as well as inclusion of the partner or spouse in the evaluation and treatment process. Although there are significant anatomic and embryologic parallels between men and women, the multifaceted nature of female sexual dysfunction is distinct from that of the male. The context in which a woman experiences her sexuality is equally, if not more, important than the physiologic outcome she experiences, and these issues need to be determined before beginning medical therapy or attempting to determine treatment efficacies. Even in women with identifiable organic disease, there are often psychosocial, emotional, or relational factors that contribute to sexual dysfunction. For this reason, a comprehensive approach, addressing both psychological as well as physiologic factors is instrumental to the evaluation of female patients with sexual complaints.

Whether sildenafil or other vasoactive agents are demonstrated to be effective in women with arousal disorder remains to be seen. There is a pressing need for more clinical and basic science research. Until more definitive efficacy data from randomized clinical trials become available, the treatment of women with sexual dysfunction should be individualized and preceded by a discussion of the uncertainty about beneficial effects and the potential for known and unknown adverse effects of drug therapy.
References


Section 6 - Endocrinology and the Life Span

Chapter 20 - Endocrine Changes in Pregnancy

Glenn D. Braunstein

PLACENTAL DEVELOPMENT

Normal placentation requires a coordinated series of events, beginning with fertilization. The fertilization rate following unprotected regular intercourse during a single menstrual cycle is 2350%. However, in approximately one-third of conceptions there is either failure of implantation or clinical or subclinical spontaneous abortion.

For the first 5 days, preimplantation development takes place within the fallopian tube. During this period, the zygote undergoes cleavage division and, at least through the eight-cell stage, the blastomeres remain totipotent. In the 16-cell stage, differentiation of the innermost cells into the inner cell mass and the surrounding cells into the trophectoderm occurs. The inner cell mass develops into the fetus, and the trophectoderm gives rise to the placenta and membranes. On approximately day 5 or 6 after fertilization, the blastocyst enters the uterus, but implantation does not occur for another 1 to 2 days. Implantation occurs after the zona pellucida disappears from around the embryo.

Implantation is a complex process that involves apposition of the microvilli present on the trophectoderm cells with microvilli on the endometrial cells, followed by removal of fluid between the cells through pinocytosis by the endometrial cells, a process stimulated by progesterone. Progesterone synthesis by the corpus luteum is stimulated and sustained during this time and for the first 6 to 7 weeks of pregnancy by secretion of human chorionic gonadotropin (hCG) by the trophoblast cells. The hCG is first detected in the maternal serum 6 to 9 days after conception. Attachment of the embryo is enhanced through the expression of a variety of adhesion molecules, including mucins, integrins, and trophinin, a trophoblast-specific cell membrane adhesion protein.

After the trophoblast attaches to the endometrium, the embryo invades the endometrium through a complex process involving matrix metalloproteinases and differentiation of the trophectoderm into cytotrophoblasts or syncytiotrophoblasts. The syncytiotrophoblasts are multinucleated cells formed by the fusion of cytotrophoblasts. The cytotrophoblasts form a column of cells that invade the endometrium, form anchoring villi, and enter the maternal vasculature, eventually replacing the endothelial layer of the endometrial and myometrial spiral arterioles with a layer of cytotrophoblasts (vascular trophoblasts). This process converts the high-resistance, low-capacity uterine vessels into low-resistance, high-volume vessels, essential for growth of the placenta and fetus. At the site of implantation, the endometrial cells undergo decidualization, enlarging and increasing their metabolic activity with enhanced production of protease inhibitors, extracellular matrix proteins, cytokines, and growth factors that modulate the extent of trophoblast invasion and influence trophoblast function.

The trophoblast cells secrete several angiogenic proteins, including vascular endothelial growth factor, platelet-derived growth factor (PDGF), and basic fibroblast growth factor (bFGF), which stimulate blood vessel development within the villi. The syncytiotrophoblasts form an outer layer of cells in the chorionic villi, between the cytotrophoblast cells and the maternal blood space on the exterior surface. Only three tissues separate the fetal blood from maternal blood: (1) the endothelium of the fetal vessels in the villi, (2) connective tissue, and (3) the trophoblasts; this form of placenta is referred to as hemochorial. Thus, in addition to hCG secretion, which is responsible for maintaining early pregnancy, and progesterone secretion, required for continuation of pregnancy after the luteal-placental shift, the syncytiotrophoblasts provide the major site for transportation of oxygen and nutrients to and removal of waste from the fetus.

Substances are transferred across the placenta through transcellular pathways that include carrier-mediated transport (e.g., immunoglobulin G through the Fc receptor) and simple extracellular diffusion. The degree of transplacental passage of a hormone from the mother to the fetus through diffusion depends on (1) the rate of placental blood flow, (2) the maternal concentration of the free or readily dissociatable hormone, and (3) the molecular mass, lipid solubility, charge, and degree of placental metabolic degradation of the hormone. Maternal-to-fetal transfer occurs for hormones smaller than 700 daltons, but the placenta is not permeable to hormones larger than 1200 daltons (Table 20-1). The syncytiotrophoblasts provide a tight barrier to most substances, although they are able to transport some substances, such as maternal-fetal transfer of hCG, progesterone, and human placental lactogen (HPL). The placenta is highly permeable to substances with a molecular mass of less than 1200 daltons, such as glucose and amino acids.

<table>
<thead>
<tr>
<th>Hormone</th>
<th>Approximate Molecular Size (Daltons)</th>
<th>Transfer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catecholamines</td>
<td>180</td>
<td>Yes</td>
</tr>
<tr>
<td>Melatonin</td>
<td>230</td>
<td>Yes</td>
</tr>
<tr>
<td>Steroid hormones</td>
<td>350</td>
<td>Yes</td>
</tr>
<tr>
<td>Vitamin D</td>
<td>350</td>
<td>Yes</td>
</tr>
<tr>
<td>Thrytropin-releasing hormone</td>
<td>360</td>
<td>Yes</td>
</tr>
<tr>
<td>Thyroid hormones</td>
<td>800</td>
<td>Limited</td>
</tr>
<tr>
<td>Oxytocin</td>
<td>1,000</td>
<td>No</td>
</tr>
<tr>
<td>Vasopressin</td>
<td>1,100</td>
<td>No</td>
</tr>
<tr>
<td>Luteinizing hormone-releasing hormone</td>
<td>1,200</td>
<td>Yes</td>
</tr>
<tr>
<td>Atrial natriuretic hormone</td>
<td>3,080</td>
<td>No</td>
</tr>
<tr>
<td>Calcitonin</td>
<td>3,400</td>
<td>No</td>
</tr>
<tr>
<td>Hormone</td>
<td>Value</td>
<td>Detection</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>-------</td>
<td>-----------</td>
</tr>
<tr>
<td>Glucagon</td>
<td>3,600</td>
<td>No</td>
</tr>
<tr>
<td>Corticotropin</td>
<td>4,500</td>
<td>No</td>
</tr>
<tr>
<td>Corticotropin-releasing hormone</td>
<td>4,800</td>
<td>No</td>
</tr>
<tr>
<td>Insulin</td>
<td>6,000</td>
<td>No</td>
</tr>
<tr>
<td>Parathyroid hormone</td>
<td>9,000</td>
<td>No</td>
</tr>
<tr>
<td>Growth hormone</td>
<td>22,000</td>
<td>No</td>
</tr>
<tr>
<td>Thyrotropin</td>
<td>27,000</td>
<td>No</td>
</tr>
<tr>
<td>Luteinizing hormone</td>
<td>30,000</td>
<td>No</td>
</tr>
<tr>
<td>Erythropoietin</td>
<td>30,400</td>
<td>No</td>
</tr>
<tr>
<td>Renin</td>
<td>40,000</td>
<td>No</td>
</tr>
</tbody>
</table>

hCG during the first trimester but not during the subsequent trimesters. [13]
MATERNAL ADAPTATIONS TO PREGNANCY

Myriad physiologic changes take place beginning shortly after implantation. During early pregnancy, these effects are hormonally mediated. In the second and third trimesters, the uteroplacental vascular system and mechanical factors associated with the enlarging gravid uterus combine with the hormonal milieu to alter the function of every system.

General Adaptations

Weight gain averages 12.5 kg, of which the fetus accounts for about 3.4 kg, the placenta 0.65 kg, amniotic fluid 0.8 kg, uterus 1 kg, breasts 0.4 kg, blood 1.5 kg, extravascular fluid 1.5 kg, and maternal fat stores approximately 3.3 kg. The volume of the uterine cavity increases from about 10 mL to an average of 5 L at term, and blood flow through the uteroplacental circulation reaches 450 to 650 mL/minute.

In order to maintain appropriate perfusion of the mother and fetal-placental unit, blood volume increases throughout pregnancy and at term is 40% to 45% higher than in the nonpregnant state. Both the plasma volume and red cell mass increase. The plasma volume increases about 45% to 50% as a result of aldosterone-stimulated sodium and water retention; the red cell mass increases approximately 20% because of increased production resulting from a twofold to threefold increase in erythropoietin secretion. The net effect is a decrease in the hematocrit by about 15% at term.

The renal blood flow and glomerular filtration rate (GFR) increase rapidly and peak during the second trimester, and there is a 50% increase in creatinine clearance, resulting in a reduction in serum creatinine. Atrial natriuretic peptide (ANP) levels increase during pregnancy and may in part be responsible for the increased renal blood flow, GFR, 24-hour urine volume, and naturesis. An alteration in the osmotic thresholds for the release of vasopressin and activation of the hypothalamic thirst centers, possibly caused by an extragonadal effect of hCG, lead to an approximately 4% reduction in serum osmolality (10 mOsm/kg). Several hemodynamic changes are induced by the low-resistance, high-capacity uteroplacental vessels, which in many respects act like an arteriovenous malformation; the increased blood volume; and the large quantities of estrogens, progesterone, and angiotensin present during pregnancy. The changes include an increase in the heart rate by 10 to 15 beats/minute, a 30% to 50% increase in cardiac output resulting from increased stroke volume in early pregnancy and heart rate during the third trimester, a reduction in diastolic blood pressure with little or no change in systolic pressure, and an approximately 20% reduction in peripheral vascular resistance.

The pulmonary vascular resistance is reduced by about a third. Pregnancy is also associated with increases in pulmonary tidal volume by about 30%, which results in a respiratory alkalosis that is compensated by increased bicarbonate excretion by the kidneys; in minute ventilatory volume by 30% to 40%; and in minute oxygen uptake. There are no changes in respiratory rate, maximum breathing capacity, or forced or timed vital capacity. However, there is an approximately 40% reduction in the expiratory reserve because of the elevation of the diaphragm by the enlarged uterus.

Gastrointestinal tract function is altered during pregnancy. Gastric emptying time is decreased by more than 50% at term and the lower esophageal sphincter tone is reduced, which, together with the displacement of the abdominal contents by the pregnant uterus, results in a marked increase in gastroesophageal reflux disease. Motility of the intestine is also reduced, contributing to the constipation that is common during pregnancy. Decreased motility of the gallbladder leads to an increased gallbladder volume and reduced emptying of bile after meals, which results in a more lithogenic bile and an increase in cholelithiasis during pregnancy.
Maternal Endocrine Alterations

Pituitary Gland

The anterior pituitary gland enlarges by an average of 36% during pregnancy, primarily because of a 10-fold increase in lactotroph size and number. This enlargement results in an increase in height and convexity of the pituitary on magnetic resonance imaging. There are reduced numbers of somatotrophs and gonadotrophs and no changes in corticotrophs or thyrotrophs. The posterior pituitary gland diminishes in size during pregnancy.

The marked increase in estrogen levels during pregnancy enhances prolactin synthesis and secretion, and maternal prolactin serum levels increase in parallel with the enlargement of the lactotrophs. At term, the mean serum prolactin concentration is 207 ng/mL (range 35 to 600 ng/mL), in contrast to a mean of 10 ng/mL in nonpregnant premenopausal women. Prolactin is also present in the amniotic fluid and appears to be primarily of decidual origin because the decidua actively synthesizes prolactin. Amniotic fluid prolactin levels are 10 to 100 times higher than in the maternal circulation in early pregnancy, and maternal bromocriptine ingestion does not reduce the amniotic fluid prolactin levels but does reduce the maternal and fetal serum concentrations. Prolactin levels return to the baseline of nonpregnancy approximately 7 days after delivery in the absence of breast-feeding. With breast-feeding, the basal prolactin levels remain elevated for several months but gradually decrease; however, with suckling, there is a brisk rise in prolactin levels within 30 minutes.

Growth hormone (GH) levels in maternal serum throughout pregnancy are unchanged, although the source of immunoreactive GH during gestation does change. Relaxin, secreted by the corpus luteum of pregnancy, and estrogens stimulate GH secretion during early pregnancy. Pituitary GH messenger ribonucleic acid (mRNA) and GH secretion decrease after the 25th week of pregnancy, and beginning in the fourth month of gestation, the placental syncytiotrophoblast secretes a variant of GH in a nonpulsatile pattern. In concert with the different sources of GH during the first and second halves of pregnancy, the GH response to provocative stimuli differs in each half. Insulin hypoglycemia or arginine infusion results in an enhanced GH response during the first half of gestation, and there is a decreased response during the second half with respect to the response in nonpregnant women.

Maternal serum concentrations of insulin-like growth factor I (IGF-I) are elevated during the second half of pregnancy, probably through the combined effect of the placental GH variant and HPL, which is evolutionarily related to both GH and prolactin. HPL has somatotropic biologic activity, and its serum concentration increases throughout pregnancy, paralleling that of IGF-I. In turn, it is likely that the suppression of pituitary GH synthesis and secretion is due to the high IGF-I concentrations, which in late pregnancy are five times higher than those in nonpregnant women.

Although the placenta synthesizes and secretes biologically active gonadotropin-releasing hormone (GnRH), pituitary gonadotropin production decreases throughout pregnancy, as indicated by a marked reduction in gonadotropin immunoreactivity in the gonadotrophs beginning at 10 weeks of gestation as well as a reduction in serum levels of luteinizing hormone (LH) and follicle-stimulating hormone (FSH). This suppression is probably mediated through the elevated blood levels of ovarian and placental sex steroid hormones along with placental production of inhibin. The suppression is incomplete because administration of exogenous GnRH leads to release of gonadotropins, although the response is blunted in comparison with that of nonpregnant women and does not return to normal until a month after birth.

Mean thyrotropin (hTSH, or thyroid-stimulating hormone) concentrations during the first trimester are significantly lower than in the second and third trimesters or in the nonpregnant state. Much of this early decrease may be accounted for by the intrinsic thyrotropic activity of hCG. The maximal biologic thyrotropic activity in maternal serum corresponds to the peak concentration of hCG at 10 to 12 weeks after the last menstrual period, at a time when there is a reciprocal relationship between the rising hCG levels and falling hTSH concentrations. The only time during pregnancy when the free thyroxine concentration in the maternal serum is elevated corresponds to the time of peak hCG and lowest hTSH, suggesting that the depressed hTSH is the result of feedback suppression by thyroxine. Despite the lower mean hTSH during early pregnancy, the hTSH response to exogenous thyrotropin-releasing hormone (TRH) is normal.

Maternal adrenocorticotrophic hormone (ACTH, or corticotropin) levels rise during pregnancy, increasing fourfold over concentrations in the nonpregnant state between 7 and 10 weeks of gestation. There is a further gradual rise to weeks 33 to 37, at which time a mean fivefold increase over prepregnancy values is found, followed by a 50% drop just before parturition and a marked 15-fold increase during the stress of delivery. The ACTH concentration returns to the prepregnancy level within 24 hours of delivery. Both the pituitary gland and the placenta serve as the source of the circulating ACTH during pregnancy, and exogenous corticotropin-releasing hormone (CRH) stimulates the release of ACTH from both tissues in a dose-dependent manner. Biologically active CRH is synthesized and secreted by the placenta and, to a lesser extent, by the decidual and fetal membranes, but in contradistinction to the inhibitory effect on pituitary CRH, glucocorticoids stimulate the expression of placental CRH.

Of interest, there appears to be a "disconnect" between CRH and ACTH during pregnancy. One would expect biologically active CRH to stimulate ACTH production. However, the qualitative patterns of CRH production, which shows an exponential rise during the sixth month of gestation, and ACTH secretion, which demonstrates a more gradual rise during pregnancy, are quite different. In addition, the lack of a significant correlation between maternal plasma CRH and ACTH during pregnancy suggests that factors such as the elevated levels of free cortisol in the maternal serum may modulate the response to CRH. Both the circadian rhythm and the ability to respond to stress are maintained throughout pregnancy; however, the ACTH response to exogenous CRH during the third trimester is blunted while the responsiveness to vasopressin is maintained, suggesting that the elevation of CRH in the maternal serum down-regulates the responsiveness to CRH.

Arginine vasopressin (AVP) concentrations in the maternal serum are similar to those in nonpregnant women. During pregnancy, however, there is increased synthesis of AVP, which is offset by the increased metabolic clearance of the hormone through destruction by a trophoblast-derived cysteine aminopeptidase (vasopressinase), whose levels rise throughout pregnancy in parallel with the increase in trophoblastic mass. As previously noted, there is a reduced osmolar set-point for thirst and the release of AVP related to the 10 mOsml/kg average decrease in plasma osmolality during pregnancy, possibly reflecting an
extragonadotropic effect of hCG. Taking into account the reduced set-point, the AVP response to dehydration and water loading is normal.

Oxytocin levels progressively increase in the maternal blood and parallel the increase in maternal serum estradiol and progesterone. The levels increase further with cervical dilation and vaginal distention during labor and delivery, stimulating uterine smooth muscles and enhancing fetal ejection.

Uterine oxytocin receptors also increase throughout pregnancy, resulting in a 100-fold increase in oxytocin binding at term in the myometrium.

**Thyroid Gland**

The thyroid gland enlarges by an average of 18% during pregnancy. The enlargement is associated with an increase in the size of the follicles with increased colloid and enhanced blood flow. This enlargement may be a response to the thyrotropic effect of hCG and asialo-hCG, which may also account for some of the increase in serum thyroglobulin concentrations noted during pregnancy. There is enhanced uptake by the maternal thyroid gland, which undoubtedly reflects the combined effects of hCG stimulation and reduction of the blood levels of iodide by enhanced renal iodide clearance.

The rising estrogen concentrations during pregnancy induce increased hepatic synthesis of thyroxine-binding globulin (TBG) as well as enhanced sialylation of TBG, which decreases its metabolic clearance rate. The results are a twofold increase in TBG and increased total thyroxine ($T_4$) and triiodothyronine ($T_3$) levels in maternal serum throughout pregnancy, whereas for the majority of gestation the free $T_4$ and free $T_3$ concentrations remain normal. There are no significant changes in the levels of thyroxine-binding prealbumin, but albumin levels are decreased because of the increase in vascular volume.

**Parathyroid Glands**

During pregnancy, approximately 30 g of calcium is transferred from the maternal compartment to the fetus, with most of the transfer occurring during the last trimester. Maternal total serum calcium levels decrease during pregnancy, with a nadir at 28 to 32 weeks related to the decrease in albumin levels that accompanies the increase in vascular volume. However, the albumin-adjusted total calcium and the ionized calcium concentrations actually rise slightly above the level in the nonpregnant state. The urinary calcium excretion rate increases in parallel with the increased GFR, and intestinal calcium absorption undergoes a twofold increase.

Although some studies have suggested that parathyroid hormone (PTH) levels increase during pregnancy, measurements of intact PTH levels by two-site immunometric assays indicate that they are within the normal, nonpregnancy range throughout pregnancy. In contrast, the circulating concentrations of PTH-related protein (PTHrP) increase throughout pregnancy. Many normal tissues produce this protein and the source of the elevated levels during pregnancy is unclear, although the two most likely sites are the mammary tissue and the placenta. This protein is probably involved in placental calcium transport.

The serum levels of 25-hydroxyvitamin D are unchanged during pregnancy, but the estrogen-induced rise in vitamin D-binding globulin results in a twofold increase in 1,25-dihydroxyvitamin D concentrations in maternal serum. There is also a rise in the biologically active free fraction of 1,25-dihydroxyvitamin D, which may reflect both increased maternal renal 1-hydroxylase activity and the synthesis and secretion of 1,25-dihydroxyvitamin D by the placenta.

This increase in the active metabolite of vitamin D may be responsible in part for the enhanced intestinal calcium absorption.

**Pancreas**

Hyperplasia and hypertrophy of the beta cells in the islets of Langerhans are probably the result of stimulation by estrogen and progesterone. During early pregnancy, the glucose requirements of the fetus lead to enhanced transport of glucose across the placenta by facilitated diffusion, and maternal fasting hypoglycemia may be present. Although basal insulin levels may be normal, there is hypersecretion of insulin in response to a meal. Because the half-life of insulin is not altered during pregnancy, this increase represents an increase in synthesis and secretion. The results are enhanced glycogen storage and decreased hepatic glucose production.

As pregnancy progresses, the levels of hPL rise, as do the levels of glucocorticoids, leading to the insulin resistance found during the last half of pregnancy. Thus, in late pregnancy, glucose ingestion results in higher and more sustained levels of glucose and insulin and a greater degree of glucagon suppression than in the nonpregnant state.

**Adrenal Glands**

As a result of the hyperestrogenemia of pregnancy, hepatic production of cortisol-binding globulin is increased. The increased production results in a doubling of the maternal serum levels cortisol-binding globulin, which in turn results in decreased metabolic clearance of cortisol and a threefold rise in total plasma cortisol by week 26, when the levels reach a plateau until they rise at the onset of labor. The rate of cortisol production is increased, and the plasma free cortisol concentrations are also increased. The enhanced cortisol production is due to an increase in the maternal plasma ACTH concentrations and hyperresponsiveness of the adrenal cortex to ACTH stimulation during pregnancy.

Cortisol secretion follows that of ACTH, and the diurnal rhythm is maintained during pregnancy. Despite the elevated free cortisol levels, pregnant women do not develop the stigma of glucocorticoid excess, possibly because of the antiglucocorticoid activities of the elevated concentrations of progesterone.

Plasma renin substrate levels are increased as a consequence of the effects of estrogen on the liver. Renin levels are also increased, and increased renin activity results in increased levels of angiotensin II, which lead to an eightfold to 10-fold increase in aldosterone production and serum aldosterone levels.

Despite their basal elevations, the various components of the renin-angiotensin-aldosterone system demonstrate normal responses to positional changes, sodium restriction, and sodium loading. The elevated aldosterone levels do not lead to an increase in serum sodium, a decrease in serum potassium, or an increase in blood pressure, which again may reflect the high progesterone concentrations, which are capable of displacing aldosterone from its renal receptors. Another mineralocorticoid, 11-deoxycortisol, shows a sixfold to 10-fold increase in concentration at term.

Elevated levels of this hormone are due to estrogen-induced extraglandular 21-hydroxylase of progesterone produced by the placenta.

Levels of androstenedione and testosterone, whether they are of adrenal or ovarian origin, are elevated because of the estrogen-induced increase in hepatic synthesis of sex hormone-binding globulin. However, the free androgen levels remain normal or low. The adrenal production rates of dehydroepiandrosterone (DHEA) and dehydroepiandrosterone sulfate (DHEAS) are increased twofold, but the maternal serum concentration of DHEAS is reduced to one-third to one-half the nonpregnancy levels because of the enhanced 16-hydroxylation and placental utilization of 16-hydroxydehydroepiandrosterone sulfate in estrogen formation.

Adrenal medullary function remains normal throughout pregnancy. Thus, 24-hour urine catecholamine and plasma epinephrine and norepinephrine levels are similar to concentrations in the nonpregnant state.
PLACENTAL HORMONE PRODUCTION

Steroid Hormones

Placental steroidogenesis takes place in the syncytiotrophoblast, and synthesis and secretion of estrogens and progesterone increase throughout pregnancy in concert with the increase in the trophoblast mass (Fig. 20-2). The placenta has reduced levels of, or lacks several enzymes important for, the de novo production of estrogens and progesterone and thus is dependent on precursors of both maternal and fetal origin. This dependence has led to the concept of the maternal-fetal-placental unit. These interactions are outlined in Figure 20-3.

Progestrone

Although the trophoblast can synthesize cholesterol from acetate, the amount of an essential enzyme needed for cholesterol synthesis, hydroxymethylglutaryl-CoA reductase, in placental microsomes is low because of the inhibitory effects of the high intracellular concentrations of cholesterol that result from progestrone inhibition of cholesterol esterification. Therefore, steroid synthesis by the placenta is dependent on the delivery of low-density lipoproteins (LDLs) and very-low-density lipoproteins (VLDLs) from the maternal circulation. The syncytiotrophoblast contains receptors for LDLs, VLDLs, and high-density lipoproteins (HDLs). Receptor-mediated uptake of LDL cholesterol is stimulated by estrogens, as is the activity of the cholesterol side-chain cleavage enzyme (CYP11A1), which converts cholesterol to pregnenolone.

Progestrone is synthesized in the trophoblast from pregnenolone by a placental isoform of 3-hydroxysteroid dehydrogenase (3-HSD-I). Approximately 90% of the progesterone synthesized is secreted into the maternal compartment, and at term the mean maternal serum concentration is about 150 ng/mL. The fetal compartment is not essential for progesterone production. Hence, progesterone production continues after fetal death.

Progesterone appears to have multiple functions during pregnancy, the most important being preparation of the uterus for implantation and maintenance of the pregnancy. The corpus luteum is the primary source of progesterone during the first 8 to 10 weeks of pregnancy, after which ovarian progesterone production declines and placental synthesis and secretion become the major source of the hormone (the luteal-placental shift).

Luteectomy, or removal of the ovary containing the corpus luteum, within the first 35 days after conception leads to abortion, but the pregnancy generally continues if these procedures are performed 40 or more days after conception. This indicates that placental progesterone production is sufficient to maintain the pregnancy even before the luteal-placental shift. Administration of the progesterone receptor antagonist mifepristone during the first 49 days after conception results in abortion, again demonstrating that progesterone is essential for the maintenance of early pregnancy.

Progesterone also serves as an important substrate for fetal adrenal glucocorticoid and mineralocorticoid synthesis and maintenance of myometrial quiescence, possibly through inhibition of prostaglandin formation. A possible role for the high concentrations of progesterone present at the trophoblast-decidua junction is suppression of cell-mediated rejection of the fetus, which expresses paternal antigens, by maternal T lymphocytes.

Estrogens

The trophoblast lacks 17-hydroxylase and 17,20-lyase (CYP17) activities and therefore cannot directly convert progesterone to estrogen. Pregnenolone produced in the placenta enters the fetal compartment, where it is taken up by the fetal zone of the adrenal cortex, which also synthesizes pregnenolone from fetal LDL cholesterol. Pregnenolone is conjugated with sulfate by fetal steroid sulfotransferase in the fetal liver and adrenals to form pregnenolone sulfate and is converted in the fetal adrenals to 17-hydroxy-pregnenolone sulfate and then DHEAS by 17-hydroxylase and 17,20-lyase (CYP17) activities.

VLDLs, and high-density lipoproteins (HDLs). Receptor-mediated uptake of LDL cholesterol is stimulated by estrogens, as is the activity of the cholesterol side-chain cleavage enzyme (CYP11A1), which converts cholesterol to pregnenolone.

DHEAS enters the fetal circulation and undergoes hydroxylation in the fetal liver to form 16-hydroxy-DHEAS, which is converted to 16-DHEA in the placenta through the action of placental sulfatase. Further metabolism in the trophoblast by 3-HSD-I, 17-hydroxysteroid dehydrogenase (17-HSD), and aromatase (CYP19) leads to the generation of estradiol, which is quantitatively the major estrogen in the maternal circulation during pregnancy. The maternal liver actively conjugates estradiol with glucosiduronate and sulfate, which are excreted into the urine. Approximately 90% of the estradiol present in the maternal serum and urine is derived from fetal precursors, and therefore measurement of estradiol levels in serum or urine serves as an index of fetal well-being.

The DHEAS enters the fetal circulation and undergoes hydroxylation in the fetal liver to form 16-hydroxy-DHEAS, which is converted to 16-DHEA in the placenta through the action of placental sulfatase. Further metabolism in the trophoblast by 3-HSD-I, 17-hydroxysteroid dehydrogenase (17-HSD), and aromatase (CYP19) leads to the generation of estradiol, which is quantitatively the major estrogen in the maternal circulation during pregnancy. The maternal liver actively conjugates estradiol with glucosiduronate and sulfate, which are excreted into the urine. Approximately 90% of the estradiol present in the maternal serum and urine is derived from fetal precursors, and therefore measurement of estradiol levels in serum or urine serves as an index of fetal well-being.

During pregnancy, estrogens have several actions. They accomplish the following:

1. Enhance receptor-mediated uptake of LDL cholesterol, which is important for normal placental steroid production.
2. Increase uteroplacental blood flow.
3. Increase endometrial prostaglandin synthesis.
4. Prepare the breasts for lactation.

However, estrogen action does not appear to be essential in maintaining pregnancy because a fetus with deletion of the gene encoding placental sulfatase cannot...
remove the sulfate moiety from 16-hydroxy-DHEAS and therefore has maternal estrogen levels approaching only about 10% of normal. Similarly, pregnancies complicated by fetal aromatase deficiency may continue to term, again suggesting that the high concentrations of estrogens found in normal pregnancy are not necessary.
Protein Hormones

Human Chorionic Gonadotropin

Chemistry

Human chorionic gonadotropin (hCG) is a glycoprotein composed of two disissimlar subunits, and, which are non-covalently linked through hydrophobic bonding. This molecule shares structural homology with the other glycoprotein hormones, human luteinizing hormone (hLH), hFSH, and hTSH. These hormones have subunits that contain the same sequence of 92 amino acids and differ only in their carbohydrate composition; the subunits differ in both amino acid and carbohydrate structure and are responsible for the biologic and immunologic specificity of the heterodimeric (intact) hormones. The 22,000-dalton subunit of hCG is composed of 145 amino acids. Approximately 80% of the first 115 amino acids are homologous to those in the subunit of hLH. hCG has an additional 24 amino acids on its carboxyl-terminal end that enhance its biologic activity.

Both subunits of hCG contain two oligosaccharide chains attached to asparagine residues through N-glycosidic linkages, and the subunit contains in addition four O-serine-linked oligosaccharide units in the carboxyl-terminal peptide. The carbohydrate composition of hCG contains microheterogeneity and affects hormone clearance and biologic activity. The tertiary structure of hCG is determined by the carbohydrate composition and multiple disulfide bonds within each subunit. The subunit contains five disulfide bonds; the subunit has six. In each of the subunits, three of the disulfide bonds form a cystine knot, similar to that found in PDGF- and transforming growth factor (TGF-).

The free subunit has a 41-minute fast T$_{1/2}$ and a 76-minute slow T$_{1/2}$ of 4 hours, and the free subunit has a 13-minute fast T$_{1/2}$ and a 76-minute slow T$_{1/2}$. After it is secreted, hCG exhibits a biexponential clearance from the circulation with a fast half-time (T$_{1/2}$) of 6 hours and a slow T$_{1/2}$ of close to 36 hours. In contrast, the free subunit has a 41-minute fast T$_{1/2}$ and a slow T$_{1/2}$ of 4 hours, and the free subunit has a 13-minute fast T$_{1/2}$ and a 76-minute slow T$_{1/2}$. Approximately 22% of the intact hormone appears in the urine unchanged; the rest undergoes metabolic degradation (Fig. 20-5). One of the early steps is proteolytic cleavage ("nicking")

The single subunit gene, located on chromosome 6, is actively expressed in both the cytotrophoblast and syncytiotrophoblast. In contrast, the subunit is encoded by a cluster of six genes located on chromosome 19 in proximity to the hLH- gene. Three of the hCG- genes are actively transcribed during pregnancy, primarily in the syncytiotrophoblast, which thus has the ability to synthesize and secrete free subunits and intact hCG. After synthesis of the protein core, each subunit is glycosylated, undergoes further post-translational modification through trimming of the carbohydrate, and then combines to form intact hCG.

Secretion of hCG differs from that of many of the other placental proteins, whose secretory pattern parallels that of the trophoblastic mass. hCG is first detected in maternal serum 6 to 9 days after conception. The levels rise in a logarithmic fashion, peaking 6 to 10 weeks after the last menstrual period, followed by a decline to a nadir at 18 weeks, with subsequent levels remaining constant until delivery. The placenta also secretes free subunits. During the first 13 weeks of pregnancy, relatively more subunit is synthesized than subunit, and throughout the remainder of pregnancy the opposite occurs. In addition, a hyperglycosylated form of subunit (big hCG) that is unable to combine with free subunit is secreted into the maternal serum.

The physiologic factors that regulate hCG secretion in vivo are unknown. Much of the data concerning factors that stimulate or inhibit hCG synthesis and secretion have been derived from in vitro studies and is difficult to extrapolate to the in vivo situation. There is strong circumstantial evidence that GnRH, synthesized in both the cytotrophoblast and syncytiotrophoblast, may be an important factor in hCG secretion. This peptide is identical to hypothalamic GnRH and stimulates placental hCG production both in vitro and in vivo, whereas GnRH antagonists decrease basal hCG secretion.

Immunohistochemical staining for GnRH in placental tissue is highest at 8 weeks of gestation and lower afterward. Roughly paralleling the pattern of hCG production, as do the circulating levels of GnRH measured in maternal serum. In addition, the placenta contains GnRH receptors. Placental GnRH release is stimulated by cyclic adenosine monophosphate (cAMP), prostaglandin E$_2$, prostaglandin F$_2$, epinephrine, epidermal growth factor, insulin, and vasoactive intestinal peptide (VIP), factors also noted to increase hCG secretion in vitro.

Two other peptides synthesized by the cytotrophoblast, activin and inhibin, also modulate GnRH and hCG secretion; activin increases both, and inhibin inhibits the action of GnRH on the syncytiotrophoblast. Increases in hCG production have also been found after trophoblast exposure to FGF, calcium, glucocorticoids, and phorbol esters. Decreased production occurs with TGF-$
\beta$, follistatin, and progesterone. The decidua may also influence hCG production through paracrine mechanisms. Decidual interleukin-1 stimulates hCG secretion in cultured trophoblasts, while decidual prolactin and an 8- to 10-kd decidual protein inhibit hCG production.

Finally, hCG may autoregulate its own production to some extent. hCG receptors are present on the surface of trophoblastic cells, and the addition of hCG to placental cells in culture stimulates cAMP production as well as proliferation and differentiation of the cytotrophoblasts into syncytiotrophoblasts. Both hCG mRNA and hCG production are stimulated by analogues of cAMP or agents that activate adenylyl cyclase, probably through a protein kinase. Thus, the net effect of an increase in syncytiotrophoblast mass and cAMP would be enhancement of hCG secretion.

The placenta is not the only site of hCG synthesis. Immunoreactive hCG has been found by immunocytochemistry or by immunooassay of extracts of a wide variety of normal tissues, including spermatozoa, testes, endometrium, kidney, liver, colon, gastric tissue, lung, spleen, heart, fibroblast, brain, and pituitary gland.

and the hormone has been shown to be synthesized in some fetal tissues. The pituitary gland appears to be the major source of hCG or an hCG-like material present in nonpregnant individuals. Immunoreactive and bioactive hCG has been partially purified from pituitary glands; the material is secreted in vitro by fetal pituitary cells and is shown by immunocytochemistry to be present in gonadotroph-type cells that do not contain hLH or human FSH.

Immunoreactive hCG has been measured in sera from normal, nonpregnant individuals, with the highest concentrations found in postmenopausal women. In postmenopausal women, this material is secreted in a pulsatile fashion in parallel with hLH pulses, and during the normal menstrual cycle the immunoreactive hCG shows a midcycle peak concomitant with the LH peak. In both men and postmenopausal women, GnRH stimulates secretion of the hormone, whereas its secretion is inhibited by oral contraceptives in women and by a GnRH agonist in gonadal men.

Both gestational and nongestational trophoblastic tumors secrete hCG and its free subunits. The sources of hCG secretion in nongestational trophoblastic neoplasms are the syncytiotrophoblastic cells and in seminomas are the trophoblastic giant cells. In many instances, the tumors produce incomplete forms of hCG or its subunits, and differences in carbohydrate content from the hCG in pregnancy have been especially apparent. A wide variety of nontrophoblastic tumors also secrete hCG, although the predominant moiety appears to be the free subunit of hCG.
The presence of hLH-hCG receptors on the corpus luteum. The parallel rise of progesterone and hCG in early pregnancy. The early production of hCG by the implanting trophoblast. Approximately 20% of patients with complete moles develop persistent trophoblastic disease, whereas only 2% to 4% of patients develop persistent disease after partial molar pregnancy. Persistent trophoblastic disease also can occur after a normal term pregnancy as well as pregnancies that end in spontaneous

Nicked hCG is unstable and dissociates into free subunit and nicked free subunit. The latter is further metabolized, primarily in the kidney, to produce the core fragment, which is composed of the subunit amino acids 6 to 40 disulfide bridged to amino acids 55 to 92, trimmed of a portion of carbohydrate, and has a molecular mass of 10,479 daltons. This fragment is the major form of immunoreactive hCG present in the urine in pregnancy. In normal pregnancy, the urine also contains variable quantities of the hyperglycosylated form of subunit, free subunit, free subunit, nicked hCG, nicked free subunit, carboxyl-terminal fragments of the subunit, and fragments of the subunit.

Physiologic Functions

Most, if not all, of the physiologic functions of hCG occur after interaction of the hormone with the hLH-hCG receptor. The receptor gene is located on chromosome 2 and encodes for a G protein-coupled receptor with seven hydrophobic transmembrane domains and a large extracellular amino terminus that binds to hCG (and hLH). The receptor is part of superfamily of receptors, including those for hFSH, hTSH, hPG, VIP, PTH, and receptors for a variety of biogenic amines and neurotransmitters. The hCG-receptor interaction results in increased cAMP production and, in some tissues, increased phosphoinositide turnover.

Because of the close structural homology of the hLH-hCG receptor with the other glycoprotein hormone receptors, hCG may interact with the hFSH and hTSH receptors and thus has weak intrinsic hTSH and hFSH biologic activity. As previously noted, the hTSH-like activity of hCG is clinically manifested during normal pregnancy by the reciprocal decrease in maternal hTSH at the time of the hCG peak between 6 and 12 weeks after the last menstrual period. It is especially important in patients with hydatidiform moles and other forms of trophoblastic disease in which hCG levels may exceed 100,000 IU/L and result in clinical thyrotoxicosis.

One of the major functions of hCG during pregnancy is the "rescue" of the corpus luteum during the conception cycle. During a menstrual cycle without conception, progesterone concentrations in the serum increase for the first 6 to 7 days of the luteal phase, followed by a 3- to 4-day plateau and then a decrease resulting in shedding of the endometrial lining. After conception and implantation, the corpus luteum continues to secrete progesterone and 17-hydroxyprogesterone for another 4 to 6 weeks. The maternal serum progesterone and 17-hydroxyprogesterone concentrations then decrease, indicating a marked diminution in corpus luteum function. The fall in 17-hydroxyprogesterone concentrations continues, but the drop in progesterone levels is only transient. This marks the transition from dependence of ovarian progesterone production to placental progesterone secretion (the luteal-placental shift). As previously noted, luteectomy during the first 50 days after the last menstrual period is associated with a decline in progesterone levels and expulsion of the products of conception. After a therapeutic abortion, progesterone levels also drop rapidly.

Thus, the fetal-placental unit is responsible for the signal to maintain the corpus luteum. The data supporting the idea that hCG is that physiologic signal include the following:

1. The presence of hLH-hCG receptors on the corpus luteum.
2. The early production of hCG by the implanting trophoblast.
3. The dose-dependent increase in cAMP, progesterone, and estradiol from luteal cells cultured in vitro after exposure to hCG.
4. The parallel rise of progesterone and hCG in early pregnancy.
5. The enhanced progesterone secretion and prolongation of the menstrual cycle in nonpregnant women given exogenous hCG during their luteal phase.

The inability of hCG to prolong the life of the corpus luteum of pregnancy beyond the sixth to eighth week of pregnancy appears to be due to homologous desensitization of the adenylate cyclase system and the inhibitory effects of the high estrogen levels on progesterone synthesis through inhibition of 3-hydroxysteroid dehydrogenase and 5α-isomerase in the corpus luteum.

Another physiologic role for hCG is in the differentiation of fetal male genitalia through stimulation of the LH-hCG receptors on the fetal testicular Leydig cells during the period when differentiation of Wolffian duct structures and development of the external genitalia occur. The maximum testosterone production per unit weight of the testes coincides with the maximum binding of 125I-labeled hCG to the fetal testicular receptors at 10 to 12 weeks of development, and fetal Leydig cells produce cAMP and testosterone in vitro after exposure to hCG. The hCG concentrations in fetal serum parallel the fetal testicular testosterone levels at a time when the amount of fetal pituitary LH is not sufficient to stimulate the testosterone production.

There are several other possible actions of hCG during normal pregnancy. In vitro, hCG stimulates the differentiation of cytotoxic lymphoblasts to syncytiotrophoblast and hence may play an important paracrine role in regulating syncytiotrophoblast mass and production of trophoblast hormones.

Additional data supporting this autoregulatory effect of hCG include the in vivo stimulation of placental synthesis of cAMP, activation of glycogen phosphorylase, and incorporation of radiolabeled galactose and leucine into placental proteins upon exposure to hCG. The fetal zone of the adrenal releases DHEAS in response to hCG exposure in vitro, and therefore hCG may have androgenic effects in concert with fetal pituitary ACTH and placental ACTH.

It has also been suggested that hCG plays a role in the immunosuppression that occurs during pregnancy. Many early studies on this topic were hampered by the use of impure preparations of hCG or the presence of preservatives such as phenol that may alter the end-points of the test systems used to define immunosuppression. In addition, the immunosuppressive effects may be due to gonadal steroid secretion in response to the hCG in the in vivo models used in some of the studies. Relaxin secretion from the corpus luteum is stimulated by hCG both in vivo and in vitro.

Finally, the decrease in osmotic threshold for thirst and AVP release during pregnancy is clearly related to hCG. Whether this decrease is due to a direct effect of hCG or an indirect effect through stimulation of gonadal steroids or interaction with hLH-hCG receptors present in vascular smooth muscle is unclear.

Gestational Trophoblastic Disease

Gestational trophoblastic disease (GTD) includes complete and partial hydatidiform moles, choriocarcinoma, and placental-site trophoblastic tumor. Complete molar pregnancy is the most common variety, occurring in 1 to 2 in 1000 pregnancies. Patients usually present with vaginal bleeding, a uterus that is larger than expected for the duration of pregnancy, amenorrhea, and excessive vomiting. Pathologically, trophoblast hyperplasia, marked edema of the choriocarcin, and absence of fetal tissues are observed. In contrast, partial moles demonstrate focal trophoblast hyperplasia and villous swelling and often have fetal tissues with congenital malformations. Approximately 20% of patients with complete moles develop persistent trophoblastic disease, whereas only 2% to 4% of patients develop persistent disease after partial molar pregnancy. Persistent trophoblastic disease also can occur after a normal term pregnancy as well as pregnancies that end in spontaneous

Hormones. Amsterdam, Elsevier Science, 1988, p 33.)
or induced abortion.

Choriocarcinoma is the most aggressive malignant form of persistent trophoblastic disease and may involve complications from local uterine disease, such as bleeding and rupture of the uterus, or from the effects of metastases, especially those involving the liver, lungs, and brain. The least common form of GTD is placental-site trophoblastic tumor, which is derived from the intermediate trophoblast and is often associated with vaginal bleeding and amenorrhea. 161

All of these neoplasms secrete hCG, free subunit, and often additional forms of these molecules. With the exception of placental-site trophoblastic tumor, which secretes relatively low amounts of hCG, the serum and urine concentrations of hCG roughly parallel the tumor burden and also provide prognostic information. Thus, hCG measurements in concert with clinical and radiologic findings, especially vaginal ultrasonography findings, are useful for making the diagnosis of GTD.

Hydatidiform moles are initially treated with uterine dilatation and evacuation with or without adjunctive single-agent chemotherapy with methotrexate or actinomycin D. Approximately 90% of patients with low-risk, persistent trophoblastic disease are cured by single-agent chemotherapy; 75% of patients with high-risk, metastatic disease are cured by multi-agent chemotherapy, including etoposide, methotrexate, actinomycin D, cyclophosphamide, and vincristine. Serial hCG measurements are invaluable for monitoring as they accurately reflect the effect of therapy on the tumor. 162

Human Placental Lactogen

Also called choric somatomammotropin, hPL is a single-chain, nonglycosylated polypeptide composed of 191 amino acid residues and two disulfide bridges, with a molecular mass of 21,600 daltons. 163 It is closely related chemically and biologically to both GH (85% amino acid homology) and prolactin (13% amino acid homology) 164 The hGH-hPL gene cluster is located on the long arm of chromosome 17 and consists of five genes coding for pituitary hGH (hGH-N), one for placental hGH (hGH-V), and three for placental hPL (hPL-L, hPL-A, and hPL-B, of which only hPL-A and hPL-B are transcribed). 165

hPL is synthesized and secreted by the syncytiotrophoblast and is detected in maternal serum between 20 and 40 days of gestation. 166 The maternal serum levels rise rapidly and peak at 34 weeks, followed by a plateau. 167 168 Both the serum concentrations and placental hPL mRNA concentrations are closely correlated with placental weight and syncytiotrophoblastic mass. 169 170 The maternal serum concentrations at term average between 6 and 7 µg/mL; at that time, on the basis of the 9- to 15-minute T½ of disappearance from the circulation, the placental production rate of hPL is in excess of 1 g/day. 171 The fetal serum levels are 1/50 to 1/100 of the maternal levels. 167

The physiologic in vivo regulation of hPL synthesis and secretion, other than the constitutive production related to placental mass, is unknown. Several studies have examined the possible role of nutrients in hPL secretion in pregnant women. Neither acute hyperglycemia nor hypoglycemia appeared to alter the hPL concentrations, although prolonged glucose infusions decreased and prolonged fasting increased the concentrations. 172 173 174 Arginine infusions, dexamethasone administration, and changes in plasma free fatty acid levels did not affect the maternal hPL concentrations. 175 176 Glucose, estrogen, glucocorticoids, progestagens, epinephrine, oxytocin, TRH, GnRH, and l-dopa have been examined in vitro systems and found to be without consistent effects. 177 178 179 180

Angiotensin II, IGF-I, phospholipase A, 181 and arachidonic acid, and epidermal growth factor stimulated hPL release in vitro. 182 183 184 Epidermal growth factor probably enhances production through promotion of cytotrophoblast-to-syncytiotrophoblast differentiation. 185 186 Apolipoprotein AI also stimulated hPL synthesis and release through CAM-dependent and arachidonic acid-dependent pathways. 187 188 189 Because changes in the maternal plasma apolipoprotein AI concentrations parallel those of hPL during pregnancy, it is likely that this apoprotein, alone and as part of circulating HDL, is important in the secretion of hPL. 190

hPL has a number of biologic activities that are qualitatively similar to those of hGH and prolactin and can bind to both the hGH and prolactin receptors. 191 In various bioassay systems, hPL had weak somatotrophic and lactogenic effects 192 193 194 It appears to be a major regulator of IGF-I production, and during pregnancy, hPL concentrations are correlated with those of IGF-I. 195 196 197 hPL also affects the metabolism of maternal nutrients. It stimulates pancreatic islet insulin secretion, both directly and after carbohydrate administration, 198 199 200 and is a diabetogenic factor during pregnancy through its promotion of insulin resistance. It enhances lipolysis, leading to a rise in free fatty acids, which may in part be responsible for the insulin resistance. 201

The various biologic activities of hPL have led to the hypothesis that the role of hPL during pregnancy is to provide the fetus with a constant supply of glucose and amino acids. 202 The hPL-stimulated lipolysis allows the mother to utilize free fatty acids for energy during fasting, allowing glucose, amino acids, and ketone bodies to cross the placenta for use by the fetus. In addition, hPL has actions in the fetus, promoting amino acid uptake by muscle and stimulating protein production, IGF-I production, and glycogen synthesis. 203

Despite the proposed importance of hPL in maternal and fetal metabolic homeostasis during pregnancy, its absence does not appear to impair pregnancy. Deficient or absent hPL production related to gene defects has been described in several women who experienced normal pregnancies and delivered normal infants. 204 205 206

Placental Growth Hormone

Placental GH, hGH-V, is synthesized and secreted by the syncytiotrophoblast. 207 208 Alternate splicing of the hGH-V gene results in two nonglycosylated isoforms with molecular masses of 22 and 26 kD. 209 210 The 22-kD variant may also be glycosylated and circulate as a 26-kD protein. 211 HGH-V is detected in the maternal plasma from 10 weeks of gestation and peaks during the third trimester 212 213 214 (Fig. 20-7). 215

HGH-V has somatotropic activity and stimulates IGF-I production, and the increase in IGF-I concentrations may in turn be responsible for the suppression of maternal pituitary hGH secretion 216 217 218 (see Fig. 20-7). Unlike pituitary hGH, hGH-V is not secreted in a pulsatile fashion, nor is it released from the trophoblast by growth hormonereleasing hormone (GHRH), but it is inhibited by glucose. 219 It has been estimated that at term, 85% of the GH biologic activity in maternal serum is due to hGH-V, 12% to hPL, and only 3% to pituitary hGH. 220 Within 48 hours of delivery, pituitary hGH secretion returns to normal.

Human Chorionic Corticotropin

The syncytiotrophoblast synthesizes an ACTH-like peptide, human chorionic corticotropin (HCC), as well as several proopiomelanocortin-derived peptides, including α-lipotropin, endorphin, and –melanocyte-stimulating hormone. 221 222 223 The maternal serum concentrations of ACTH increase as pregnancy progresses, and the elevation of free cortisol levels during pregnancy may be related in part to both placental hCC and pituitary ACTH production. 224

HCC secretion is stimulated by CRH, which is probably the most important factor regulating the local production of the peptide through paracrine or autocrine mechanisms, or both, because it is also produced by both the cytotrophoblast and the syncytiotrophoblast. 225 Unlike the situation with the pituitary, glucocorticoids and oxytocin also stimulate HCC release from placental cells. 226 227 Indeed, the resistance of maternal plasma ACTH concentrations to suppression after...
glucocorticoid administration may reflect the placental hCC contribution to the total pool of circulating immunoreactive ACTH.

### Hypothalamic Peptides

#### Corticotropin-Releasing Hormone

Both the cytotrophoblast and the syncytiotrophoblast synthesize and secrete GnRH, which has the same chemical structure.

In vitro, GnRH production by placental explants or purified trophoblasts is stimulated by prostaglandins, epinephrine, insulin, epidermal growth factor, VIP, estradiol, and estriol, and secretion is reduced by inhibin, progesterone, and -opioid and -opioid agonists. The syncytiotrophoblast contains low-affinity GnRH receptors, whose concentrations parallel the hCG secretory pattern.

Because GnRH stimulates hCG secretion by placental explants and purified trophoblast cells in vitro, with the response of early to midtrimester placentas being greater than that of term trophoblast, it is reasonable to conclude that GnRH is an important autocrine or paracrine regulator of hCG secretion.

#### Other Peptides

A substance with GHRH-like activity has been found in the placenta. It has unknown physiologic significance because there is no convincing evidence that a chorionic thyrotropin exists, and hCG appears to be the major trophoblastic thyrotropin-like substance.

### Immunoreactive Somatostatin

This pattern has led to the speculation that somatostatin inhibits hPL production and that the loss of inhibition allows the placenta to secrete increasing quantities of hPL. The finding of somatostatin receptors in the placenta adds some indirect support to this hypothesis. However, somatostatin does not inhibit hPL (or hGH-V) production by placental cells exposed to the peptide in vitro.

A substance with GHRH-like activity has been found in the placenta. However, because exposure of placental cells to hypothalamic GHRH does not result in stimulation of hPL or hGH-V secretion, it is unlikely that human placental GHRH is physiologically important.

### Growth Factors

Many growth factors, growth factor-binding proteins, and growth factor receptors have been identified in the placenta. These include IGF-I, IGF-II, relaxin, epidermal growth factor, PDGF, nerve growth factor, FGF, TGF-α, inhibin, activin, and folliculoostatin.

As reviewed earlier, a number of these factors have been implicated in the autocrine or paracrine regulation of placental hormone synthesis and release and placental angiogenesis, and they may have important actions in fetal development. In addition to the placenta, the human endometrium is a rich source of growth factors, cytokines, and vasoactive neuropeptides that are important for uteroplacental function.
References


38. Bigazzi M, Ronga R, Lancranjan I, et al. A pregnancy in an acromegalic woman during bromocriptine treatment: effects on growth hormone and prolactin in the maternal, fetal, and amniotic...

887


References


Chapter 21 - Endocrinology of Fetal Development

Delbert A. Fisher

The unfolding of our understanding of mammalian pregnancy and fetal development represents one of the dramatic chapters of scientific progress during the past half-century, but we are only beginning to understand the complex genetic, growth factor, and hormonal interactions involved in implantation, placentation, embryonic and fetal development, parturition, and fetal adaptation to extrauterine life. An array of homeobox genes program embryogenesis and fetal development in concert with autocrine, paracrine, and endocrine networks of chemicals, hormones, and growth factors that provide the cellular communication coordinating maternal-placental-fetal interactions and fetal maturation. An important concept that has emerged is that the fetal-placental endocrine milieu is unique and the insights from studies of endocrine and growth factor systems of mature species are not directly applicable to the fetal environment.

Unique features of the placental-fetal endocrine environment (Table 21-1) include the growing spectrum of placental hormones and growth factors and a variety of transient fetal endocrine adaptations to the intrauterine environment. The fetal adrenal cortex, the para-aortic chromaffin system including the paired organs of Zuckerkandl, and the intermediate lobe of the pituitary are prominent among these. Vasotocin, the parent neurohypophyseal hormone in submammalian species, is expressed transiently during fetal life, and calcitonin, a largely vestigial hormone in adult mammals, plays a significant role in fetal calcium and bone metabolism.

In addition, the active adrenal glucocorticoid, cortisol, and the major thyroid hormone, thyroxine (T4), are largely inactive during much of fetal life because of the preferential synthesis of inactive moieties. Hormones and growth factors that play prominent roles in the fetus include catecholamines, parathyroid hormone-related protein (PTHrP), antimüllerian hormone, insulin-like growth factor II (IGF-II), transforming growth factor (TGF-), and the neuroregulins. In the perinatal period, cortisol serves to modulate the functional adaptations requisite for extrauterine survival. In addition, hormonal imprinting during the fetal-perinatal period conditions the adult functional characteristics of selected endocrine systems.

This chapter reviews the current status of our understanding of the fetal-placental endocrine and growth factor milieu, maturation of the fetal endocrine systems, and adaptations of the fetal endocrine system to extrauterine life.
PLACENTA

The fetal milieu depends on a functioning placenta, which develops in parallel with the fertilized ovum. By 6 to 7 days after conception, the blastocyst consists of an outer layer of trophoblast cells and an inner cell mass destined to become the embryo. The trophoblast cells have implanted in the endometrium and within 10 days have developed two distinct layers, an inner cytotrophoblast layer and an outer layer of continuous cytoplasm, the syncytiotrophoblast, which forms the early fetal-maternal interface. Pockets of cytotrophoblast cells in the mature placenta serve as a reservoir of stem cells for continuing syncytiotrophoblast development.

The predominantly syncytiotrophoblastic placenta grows progressively throughout gestation. As the placenta develops, the chorionic villi containing the fetal capillaries extend into the maternal lakes of blood within the maternal decidua. Within the villi three layers of fetal tissue separate the fetal circulation from the maternal circulation: the cytotrophoblast-syncytiotrophoblast layer, the fetal mesenchyme layer of extraembryonic connective tissue, and the fetal capillary endothelium. The syncytiotrophoblast is the major site of diffusion between the maternal lakes of blood in the placenta and the fetal capillaries.

![Figure 21-1](image)

**Figure 21-1** Diagrammatic representation of a chorionic villus extending into the maternal blood lake and showing fetal capillaries in the fetal mesenchyme. The villus is sheathed by the syncytiotrophoblast. The residual sparse areas of cytotrophoblast provide cells to renew and maintain the syncytiotrophoblast layer. The villus is surrounded by maternal blood in the maternal intervillous space. The placenta serves as an important endocrine organ. Hormones are produced by cytotrophoblast and syncytiotrophoblast cells. Neuropeptides appear to modulate syncytiotrophoblast production of placental protein hormones, and decidual prostaglandins and cytotrophoblast growth factors may participate in regulation of syncytiotrophoblast steroid hormone production. See text for details.

TABLE 21-1 — Features of the Fetal Endocrine Environment

<table>
<thead>
<tr>
<th>Placental Hormone Production</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estrogens</td>
</tr>
<tr>
<td>Progesterone</td>
</tr>
<tr>
<td>Polypeptide hormones</td>
</tr>
<tr>
<td>Neuropeptides</td>
</tr>
<tr>
<td>Growth factors</td>
</tr>
</tbody>
</table>

**Unique Fetal Endocrine Systems**

- Fetal adrenal cortex
- Para-aortic chromaffin system
- Intermediate lobe of the pituitary

**Prominent Fetal Hormones or Metabolites**

- Vasotocin
- Calcitonin
- Cortisone
- Reverse triiodothyronine (rT₃)
- Sulfated iodothyronines
- Ectopic neuropeptides

**Fetal Endocrine System Adaptations**

- Adrenal-placental interactions
- Developmentally regulated growth factor control of fetal growth
- Neuropeptides and fetal water metabolism
- Parathyroid glands and placental calcium transport
- Catecholamine and vasopressin responses to hypoxia
- Cortisol programming for extrauterine exposure
- Catecholamine and cortisol control of extrauterine adaptation
- Perinatal hormonal imprinting

**Placental Hormone Transfer**

The placenta regulates maternal-fetal molecular exchange, and thin areas of the syncytiotrophoblast adjacent to the fetal capillaries seem to be specialized for this function. However, the fetal endocrine milieu is largely independent of maternal hormones because the placenta is impermeable to most peptide hormones.

There are two major routes for the transfer of molecules across the placenta: an extracellular route through fluid-filled intercellular channels and a transcellular route. The rate of extracellular diffusion is related to the luminal diameter of the intercellular or paracellular channels and to the molecular weight (molecular radius or size) and lipid solubility or hydrophilicity of the transferred molecule. The placenta is more permeable to lipid-soluble molecules, and the permeability for both lipid-soluble and lipid-insoluble molecules decreases with increasing molecular weight. The transfer or diffusion of L-glucose is believed to be accomplished by extracellular diffusion.

The placental transfer of a number of hormones is summarized in Table 21-2. The differences in placental structure among species have a limited influence on
placental hormone transfer, and data derived from some animal and primate species are included. Hormones larger than 0.7 to 1.2 kDa have little or no access to the fetal compartment. The exception is immunoglobulin G, which is actively transported from mother to fetus during the latter half of gestation. 9

Placental cell membranes contain a variety of receptors for polypeptide hormones and growth factors, including insulin, the IGFs, and epidermal growth factor (EGF). These receptors bind and in some instances degrade their respective ligands but do not facilitate placental transfer.

Hormones that traverse the placenta by the transcellular route and are metabolized en route include cortisol, estradiol, thyroid hormones, and catecholamines. The placental cells contain an active 11-hydroxysteroid dehydrogenase (11-HSD) that catalyzes the conversion of most of the cortisol to inactive cortisone. Placental 17-hydroxysteroid dehydrogenase is considered to prevent passage of excessive estrogens to the fetus by catalyzing inactivation of estradiol to estrone. Placental tissue also contains an iodothyronine inner ring monodeiodinase, which deiodinates most of the T4 to inactive reverse triiodothyronine (rT3) and converts active 3,5,3’-triiodothyronine (T3) to inactive diiodothyronine. Catecholamine-degrading enzymes in placental tissue include both monoamine oxidase and catechol-O-methyltransferase, and both metanephrine and dihydroxymandelic acid metabolites of catecholamines are present in placental homogenates.

<table>
<thead>
<tr>
<th>Hormone</th>
<th>Approximate Molecular Size (daltons)</th>
<th>Placental Transfer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catecholamine</td>
<td>180</td>
<td>Yes</td>
</tr>
<tr>
<td>Melatonin</td>
<td>230</td>
<td>Yes</td>
</tr>
<tr>
<td>Steroid hormones</td>
<td>350</td>
<td>Yes</td>
</tr>
<tr>
<td>Vitamin D</td>
<td>350</td>
<td>Yes</td>
</tr>
<tr>
<td>Thyrotropin-releasing hormone (TRH)</td>
<td>360</td>
<td>Yes</td>
</tr>
<tr>
<td>Thyroid hormones</td>
<td>600</td>
<td>Limited</td>
</tr>
<tr>
<td>Oxytocin (OT)</td>
<td>1,000</td>
<td>No</td>
</tr>
<tr>
<td>Vasopressin</td>
<td>1,100</td>
<td>No</td>
</tr>
<tr>
<td>Luteinizing hormonereleasing hormone (LHRH)</td>
<td>1,200</td>
<td>Yes</td>
</tr>
<tr>
<td>Atrial natriuretic hormone</td>
<td>3,080</td>
<td>No</td>
</tr>
<tr>
<td>Calcitonin (CT)</td>
<td>3,400</td>
<td>No</td>
</tr>
<tr>
<td>Glucagon</td>
<td>3,600</td>
<td>No</td>
</tr>
<tr>
<td>Corticotropin</td>
<td>4,500</td>
<td>No</td>
</tr>
<tr>
<td>Corticotropin-releasing hormone (CRH)</td>
<td>4,800</td>
<td>No</td>
</tr>
<tr>
<td>Insulin</td>
<td>6,000</td>
<td>No</td>
</tr>
<tr>
<td>Parathyroid hormone (PTH)</td>
<td>9,000</td>
<td>No</td>
</tr>
<tr>
<td>Growth hormone (GH)</td>
<td>22,000</td>
<td>No</td>
</tr>
<tr>
<td>Thyrotropin</td>
<td>27,000</td>
<td>No</td>
</tr>
<tr>
<td>Luteinizing hormone (LH)</td>
<td>30,000</td>
<td>No</td>
</tr>
<tr>
<td>Erythropoietin</td>
<td>30,400</td>
<td>No</td>
</tr>
<tr>
<td>Renin</td>
<td>40,000</td>
<td>No</td>
</tr>
</tbody>
</table>
Placental Hormone Production

The placenta functions as a major endocrine organ, providing a secondary source of hypothalamic, pituitary, adrenal, and gonadal hormones and growth factors (Table 21-3). The syncytiotrophoblast manufactures steroid and protein hormones; after the eighth week of pregnancy, it is the most active fetal or maternal endocrine organ. The steroid hormones are produced from both fetal and maternal substrates. The protein hormones are synthesized in the rough endoplasmic reticulum of the syncytiotrophoblast from amino acids of maternal origin. Secretion is predominantly into the maternal circulation, but significant amounts reach the fetal compartment. The cytotrophoblast produces a variety of neuropeptides and growth factors, and decidua tissue is a major source of prostaglandins and relaxin.

The network of peptides expressed by the placenta provides an autocrine-paracrine system regulating the fetal-maternal unit. Local placental control systems resemble a miniature hypothalamic-pituitary-target organ network. Maternal and fetal hormones in the placental circulation may also regulate placental hormone production. In addition to the local effects, a number of placental hormones are involved in modulation of maternal and fetal homeostasis during pregnancy (see Fig. 21-1).

Placental Estrogen

The human placenta near term secretes large amounts of estrogens, including estrone, estradiol, and estriol. Excretion rates of these steroids during the latter third of pregnancy are approximately 2, 1, and 30 to 40 mg/day, respectively, whereas total estrogen production in nonpregnant women is less than 1 mg/day. This production is due to the combined effects of the fetal adrenal gland and the placenta, first characterized by Diczfalusy as the human fetoplacental unit. Most of the estrogen is secreted into the maternal circulation, but fetal concentrations and levels in amniotic fluid are high.

The major substrates for placental estrogen synthesis are dehydroepiandrosterone (DHEA) and androstenedione. These relatively inactive adrenal steroids are derived from both fetal and maternal adrenal glands. The fetal zone of the adrenal cortex is deficient in an enzyme with 3-hydroxysteroid dehydrogenase (3-HSD) and isomerase activities but has high steroid sulfokinase activity. Thus, the conversion of pregnenolone to progesterone is limited and the major product of fetal adrenal steroidogenesis is DHEA sulfate (DHEAS), which is transported to the liver for 16-hydroxylation. DHEAS and 16-hydroxy-DHEAS are hydrolyzed in the

---

**TABLE 21-3 -- Hormones and Growth Factors Produced by the Placenta**

<table>
<thead>
<tr>
<th>Cytotrophoblast</th>
<th>Syncytiotrophoblast</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corticotropin-releasing hormone (CRH)</td>
<td>Estrogens</td>
</tr>
<tr>
<td>Growth hormone-releasing hormone (GRH)</td>
<td>Estradiol</td>
</tr>
<tr>
<td>Gonadotropin-releasing hormone (GnRH)</td>
<td>Estradiol</td>
</tr>
<tr>
<td>Thyrotropin-releasing hormone (TRH)</td>
<td>Estriol</td>
</tr>
<tr>
<td>Somatostatin (SRIF)</td>
<td>Estrone</td>
</tr>
<tr>
<td>Inhibin</td>
<td>Progesterone</td>
</tr>
<tr>
<td>Activin</td>
<td>Placental growth hormone</td>
</tr>
<tr>
<td>Neuropeptide Y</td>
<td>Chorionic gonadotropin</td>
</tr>
<tr>
<td>Enkephalins</td>
<td>Chorionic somatomammotropin</td>
</tr>
<tr>
<td>Dynorphins</td>
<td>CRH</td>
</tr>
<tr>
<td>Atrial natriuretic hormone</td>
<td>Leptin</td>
</tr>
<tr>
<td>Parathyroid hormone-related protein (PTHrP)</td>
<td>Vascular endothelial growth Factor (VEGF)</td>
</tr>
<tr>
<td>Leptin</td>
<td>Insulin-like growth factor I (IGF-I)</td>
</tr>
<tr>
<td>IGF-II</td>
<td>Erythropoietin</td>
</tr>
<tr>
<td>Renin</td>
<td>Renin</td>
</tr>
<tr>
<td>Relaxin</td>
<td>Relaxin</td>
</tr>
<tr>
<td><strong>Decidua</strong></td>
<td><strong>IGF binding protein</strong></td>
</tr>
<tr>
<td>Prolactins</td>
<td>SRIF, somatotropin release-inhibiting factor.</td>
</tr>
</tbody>
</table>

| SRIF, somatotropin release-inhibiting factor. |

---

The major substrates for placental estrogen synthesis are dehydroepiandrosterone (DHEA) and androstenedione. These relatively inactive adrenal steroids are derived from both fetal and maternal adrenal glands. The fetal zone of the adrenal cortex is deficient in an enzyme with 3-hydroxysteroid dehydrogenase (3-HSD) and isomerase activities but has high steroid sulfokinase activity. Thus, the conversion of pregnenolone to progesterone is limited and the major product of fetal adrenal steroidogenesis is DHEA sulfate (DHEAS), which is transported to the liver for 16-hydroxylation. DHEAS and 16-hydroxy-DHEAS are hydrolyzed in the
Glucocorticoids, however, have no effect on placental CRH or corticotropin (ACTH) activity. Estrogens also augment maternal glucocorticoid suppression in pregnancy suggest that the placenta may be involved in regulation of the maternal pituitary-adrenal axis during pregnancy.

Compared with corticotropin, -endorphin and -MSH are released from placenta in larger amounts. Thus, control and processing of placental POMC are different than in the anterior pituitary.

The human placenta synthesizes a pro-opiomelanocortin (POMC) and contains the POMC-derived peptides corticotropin (adrenocorticotropic hormone [ACTH] or hGH-like effects mediated by somatomedins during pregnancy.

produce identical hPL molecules. Placental tissue also expresses pituitary PRL and one of the hGH genes. The PRL, hGH, and hPL genes are also closely related. Evolution of pseudogenes.

substrate flow.

progesterone production.

The general control of the synthesis and secretion of these placental hormones is not well understood, and inhibits maternal cell-mediated immune responses to foreign (fetal) antigens. Despite the predominant secretion of progesterone into the maternal circulation, fetal blood progesterone levels are twofold to threefold higher than maternal values because of lower metabolic clearance of progesterone by the fetus. The significance of this progesterone to the fetus is not clear.

The principal substrate for placental progesterone synthesis is circulating maternal LDL and very-low-density lipoprotein (VLDL) cholesterol; de novo placental synthesis of cholesterol from acetate is limited. Placental progesterone production is largely independent of the maternal pituitary or adrenal glands, and fetal death in utero has little acute effect on maternal progesterone levels. Progesterone production is regulated by the number of LDL receptors and thus placental mass.

The major factor in control of placental progesterone production appears to be the expression of steroidogenic enzymes, including CYP11A1 and 3HSD, in cytotrophoblast cells. The type I 3HSD enzyme is expressed in placenta, whereas type II activity is expressed in adrenal and gonadal tissues. Expression of messenger ribonucleic acid (mRNA) for all the placental steroidogenic enzymes is stimulated by cyclic adenosine monophosphate (cAMP), which appears to be produced constitutively in mature cytotrophoblast cells. There is some evidence that endogenous steroids may modulate placental progesterone production.

The production of progesterone is approximately 200 mg/day during the third trimester, a value some 10-fold higher than that during the midluteal phase of the normal menstrual cycle; 90% of this amount is secreted into the maternal circulation. Progesterone acts on the uterine musculature to maintain a state of quiescence and inhibits maternal cell-mediated immune responses to foreign (fetal) antigens. Despite the predominant secretion of progesterone into the maternal circulation, fetal blood progesterone levels are twofold to threefold higher than maternal values because of lower metabolic clearance of progesterone by the fetus. The significance of this progesterone to the fetus is not clear.

Placental Poly peptide Hormone

The placenta produces several pituitary-like hormones. The most abundant are human chorionic gonadotropin (hCG) and human placental lactogen (hPL), also called human chorionic somatomammotropin. HCG is a glycoprotein of 36 to 40 kd with structural, biologic, and immunologic similarities to the pituitary gonadotropins and thyrotropin (also called thyroid-stimulating hormone [TSH]); hCG also has weak thyrotropic hormone-like activity. hPL is a 191-amino-acid protein with 96% homology to human growth hormone (hGH). It has 3% or less of the growth-promoting bioactivity of hGH and equivalent prolactin (PRL)-like effects. hCG is secreted predominantly during the first half of gestation, and hPL is secreted mainly during the second half.

The control of the synthesis and secretion of these placental hormones is not well understood, but hormone secretion is related to placental mass and continues in the absence of the fetus. Luteinizing hormone-releasing hormone (LHRH, also called chorionic gonadotropin-releasing hormone), EGF, activin, and hCG increase hCG synthesis in placental tissue in vitro, whereas inhibin and progesterone suppress synthesis. It is likely that hCG plays a role in the maintenance of the corpus luteum early in pregnancy as well as stimulation of the fetal testes and stimulation of placental progesterone production. hCG has weak thyrotropic-like activity; there is less than 0.5 mU of thyrotropin per unit of hCG, and hCG produces minimal, but sometimes significant, thyroid hyperactivity and hyperthyroidism during normal pregnancy. hPL has weak hGH-like and PRL-like bioactivities and may exert an anti-insulin effect on maternal carbohydrate and lipid metabolism, thereby increasing maternal glucose and amino acid levels and augmenting maternal-to-fetal substrate flow. In addition, hPL appears to be an important stimulus of fetal growth.

The same gene is responsible for the subunit expressed in the placenta for hCG and in the pituitary for production of luteinizing hormone (LH), follicle-stimulating hormone (FSH), and thyrotropin. There is also a single gene for the subunit of LH, whereas there are seven genes or pseudogenes, or both, for the hCG subunit. The hCG and LH subunits have similar structures, and it appears that the hCG gene arose from the LH gene and that the hCG gene family is early in the process of evolution of pseudogenes.

The PRL, hGH, and hPL genes are also closely related. The PRL gene is presumed to be the ancestral gene; hGH evolved nearly 400 million years ago, whereas hPL evolved within the last 10 million years. The hGH gene cluster includes five similar gene loci, two for hGH and three for hPL; these loci have 93% sequence homology in mRNA and probably evolved by repeated duplication over time. Only two of the hPL sequences are expressed in the placenta, and they produce identical hPL molecules. Placental tissue also expresses pituitary PRL and one of the hGH genes (hGHV), and placental hGH may contribute to the maternal hGH-like effects mediated by somatomedins during pregnancy.

The human placenta synthesizes a pro-opiol melanocortin (POMC) and contains the POMC-derived peptides corticotropin (adrenocorticotropic hormone [ACTH] or adrenocorticotropin), lipotropin and lipotropin, -endorphin, -melanocyte-stimulating hormone (-MSH), and three forms of endorphin. Corticotropin-releasing hormone (CRH) is also produced by the placenta, and CRH stimulates corticotropin production from perfused human placental fragments, suggesting a possible paracrine role for placental CRH in modulating corticotropin production in the placenta. Glucocorticoids, however, have no effect on placental CRH or corticotropin release, and otoxin (OT) stimulates placental POMC release but has no effect on pituitary release of corticotropin.

Compared with corticotropin, -endorphin and -MSH are released from placenta in larger amounts. Thus, control and processing of placental POMC are different than in the anterior pituitary. The increased plasma levels of POMC-derived peptides in pregnant women and the resistance of maternal placenta corticotropin to glucocorticoid suppression in pregnancy suggest that the placenta may be involved in regulation of the maternal pituitary-adrenal axis during pregnancy.

Inhibin and activin are produced by placental tissue. The mRNAs for inhibin subunits (α, β, and γ) are present in the placenta, and inhibit A (α) and activin A (α,α)
Activin A is present in cord serum only at term, whereas activin B is present in fetal blood and amniotic fluid before birth; activin B is largely absent from maternal serum. The role of these hormones during pregnancy is not clear. Inhibin production in placenta is stimulated by hCG, FSH, EGF, and prostaglandins. Inhibin suppresses hCG production, whereas activin can stimulate growth hormone-releasing hormone (GHRH), LHRH, progesterone, and prostaglandin production by placental cells in culture. A paracrine role for these hormones in the placenta is suggested.

Activin A and B, two forms of inhibin, are present in placenta. Activin A is produced by decidual cells, while B is produced by syncytiotrophoblast cells. Activin is thought to play a role in the regulation of placental function, including the regulation of angiogenesis, trophoblast invasion, and maternal-fetal immune tolerance. Activin A may also have a role in the regulation of maternal metabolism, including glucose metabolism.

The significance of activin in the placenta is not yet fully understood. Further research is needed to elucidate the specific roles of activin A and B in placental function and to determine how these factors interact with other signaling molecules to regulate placental development and function.
development is not clear.

EGF and TGF- mRNA and proteins have been demonstrated in pools of placental tissue from early, middle, and late gestation. Placental levels of TGF- protein are relatively high throughout gestation (90 to 180 ng/mg protein), whereas EGF values are low (3 to 9 ng/mg protein). The placenta is richly endowed with receptors that bind both EGF and TGF-. TGF- mRNA is localized in the maternal decidua early in gestation in the mouse and is present in fetal tissues. EGF induces differentiation of human trophoblast to syncytiotrophoblast, and this differentiation is associated with increased production of hCG and hPL. These studies have suggested that TGF- or EGF, or both, may influence placental maturation and function.

Transforming growth factor has been purified from human placenta, and precursor mRNA is present in placental tissue. In addition, placental growth factor, PTHrP, platelet-derived growth factor, vascular endothelial growth factor, endothelin-1, tumor necrosis factor, oncomedullin, erythropoietin, and several colony-stimulating factors and receptors have been demonstrated in placental tissue and conditioned media. These factors are also postulated to have autocrine-paracrine roles in placental growth and function.
ECTOPIC FETAL HORMONE PRODUCTION

Ectopic Fetal Polypeptide Hormone

Kidney, liver, and testes from 16- to 20-week-old human fetuses produce immunoreactive and bioactive hCG in vitro. Kidney tissue produces nearly half as much hCG (per milligram of protein) as placenta; liver activity is lower. Corticotropin-like immunoreactivity is present in relatively high concentrations in neonatal rat pancreas and kidney. This material is presumably derived from a POMC parent molecule.
Extraneural Fetal Neuropeptide

Hypothalamic neuropeptides are present in a variety of adult tissues, particularly in the pancreas and gut. In the fetus, hypothalamic neuropeptides are also present in the gut and tissues derived from it. High concentrations of TRH and somatostatin immunoreactivity have been reported in neonatal rat pancreas and gastrointestinal tract tissues, whereas hypothalamic concentrations of these immunoreactive substances are low. These neuropeptides have immunoreactive and chromatographic properties similar to those of the synthetic hypothalamic peptides. Other peptides cleaved from pre-pro-TRH are present in perinatal rat pancreas. In addition, encephalotomy does not alter the circulating TRH levels in the neonatal rat, whereas significant reductions are produced by pancreatectomy. TRH production by monolayer cultures of fetal rat pancreatic cells is stimulated by serotonin and inhibited by carbachol; catecholamines, -aminobutyric acid, and histamine have no effect. Specific neurotransmitter control has been postulated. In the sheep fetus, thyroid hormones modulate pancreatic and gut TRH levels, which suggests thyroid hormone control of extrahypothalamic TRH gene transcription or translation in the fetus.

TRH and somatostatin are present in the human neonatal pancreas and in blood of the human newborn. It seems likely that both hormones are derived mostly from extrahypothalamic sources. The presence of TRH at high concentrations in fetal ovine blood and the control of fetal pancreatic, placental, and blood TRH levels by thyroid hormones suggest a role for extrahypothalamic TRH in the control of fetal pituitary thyrotropin secretion before the near-term maturation of hypothalamic TRH. Infusion of TRH into the fetal sheep also evoked behavioral arousal, caused increased body and eye movements, and stimulated fetal breathing. The role of extraneural somatostatin in the fetus is undefined.

There is a general tendency toward hypersecretion of fetal pituitary hormones during the last half of gestation, and pituitary hormones found at high levels in cord blood from aborted human fetuses and premature human infants include hGH, thyrotropin, corticotropin, -endorphin, -lipotropin, LH, and FSH. Development of hypothalamic-pituitary control is complex, involving maturational events in the cortex and midbrain, the hypothalamus and hypothalamic-pituitary portal vascular system, peripheral endocrine systems, and the placenta itself, including hormone, growth factor, and neuropeptide production. The fetal pituitary hyperfunction appears to be related more to relatively delayed maturation of the central nervous system and hypothalamic control with unrestrained secretion of stimulating hypothalamic hormones than to the action of placental neuropeptides.
FETAL ENDOCRINE SYSTEMS

Anterior Pituitary and Target Organs

Development

The human fetal forebrain is identifiable by 3 weeks of gestation, the diencephalon and telencephalon by 5 weeks. Rathke’s pouch, the buccal precursor of the anterior pituitary gland, separates from the primitive pharyngeal stomodeum by 5 weeks of gestation. [95] The neural components of the transducer system (the hypothalamus, the pituitary stalk, and the posterior pituitary) are largely developed by 7 weeks of gestation, and the bony floor of the sella turcica is present by this time, separating the adenohypophysis from the primitive gut. Capillaries develop within the proliferating anterior pituitary mesenchymal tissue around Rathke’s pouch and the diencephalon by 6 weeks of gestation, and intact hypothalamic-pituitary portal vessels are present by 12 to 17 weeks. [94] Maturation of the portal portal vascular system continues, and the system becomes functionally intact during the period of histologic differentiation of the hypothalamus and development of the portal vascular extension into hypothalamic tissue; this maturation process extends to 30 to 35 weeks of gestation.

The hypothalamic cell condensations, which represent the hypothalamic nuclei, and the interconnecting fiber tracts are demonstrable histologically by 15 to 18 weeks of gestation. [94] Hypothalamic cells and diencephalic fiber tracts for the hypothalamic neuropeptides somatostatin, CRH, GH, and LHRH are also visible by this time. Concentrations of dopamine, TRH, LHRH, and somatostatin are significant in hypothalamic tissue by 10 to 14 weeks of gestation. Specialized anterior pituitary cell types, including lactotropes, somatotropes, corticotropes, thyrotropes, and gonadotropes, can be recognized in the anterior pituitary between 7 and 16 weeks of gestation. Anterior pituitary hormones (including hGH, PRL, thyrotropin, LH, FSH, and corticotropin) are detectable by radioimmunoassay between 10 and 17 weeks of gestation. Thus, the anatomy and biosynthetic mechanisms that make up the hypothalamic-pituitary neuroendocrine transducer appear to be functional by 12 to 17 weeks of gestation in humans.

This embryonic process is regulated by a series of homeodomain proteins or transcription factors that have been characterized by mutation analysis, gene transfection, and gene knockout studies. [95] Mutations of the homeobox genes sonic hedgehog (SHH) and ZIC2 have been identified in patients with familial and sporadic holoprosencephaly. HESX1 homeobox gene mutations have been shown in siblings with septo-optic dysplasia in association with midline brain defects and pituitary hypoplasia. Other genes involved in hypothalamic development include SF1 and LHX4. Early pituitary homeodomain factors include the Rathke pouch homeobox gene (RPX), LHX3, and LHX4. The later factors PROP1 and PIT1 program development and function of the pituitary cells producing GH, TSH, and PRL. Mutations in PROP1 and PIT1 have been described in patients with familial hypopituitarism and are fully addressed in Chapter 8. [102] [95] [97] [94] (Fig. 21-3).

Human Pituitary Growth Hormone and Prolactin

The human fetal pituitary gland can synthesize and secrete hGH by 8 to 10 weeks of gestation. [95] Pituitary hGH content increases from about 1 nmol (20 ng) at 10 weeks to 45 nmol (1000 ng) at 16 weeks of gestation. Fetal plasma hGH levels in cord blood samples are in the range 1 to 4 nmol/L during the first trimester and increase to a mean peak of approximately 6 nmol/L at midgestation. Plasma hGH levels fall progressively during the second half of gestation to a mean value of 1.5 nmol/L at term. [94] Pituitary hGH mRNA and hGH content generally parallel the increase in plasma hGH concentration between 16 and 24 weeks of gestation. [95] This pattern of ontogenesis of plasma hGH reflects a progressive maturation of hypothalamic-pituitary and forebrain function. The responses of plasma hGH to somatostatin and GHRH and to insulin and arginine are mature at term in human infants. [95] Plasma hGH levels are low in anencephalic infants.

The high plasma hGH concentrations at midgestation after the development of the pituitary portal vascular system may reflect unrestrained secretion. [94] Studies of 9- to 16-week-old human fetal pituitary cells in culture have shown a predominant response to GHRH and a limited effect of somatostatin, which suggests that the inhibitory action of somatostatin develops later in gestation. [97] This interpretation has been substantiated by in vivo studies in the sheep fetus, which have shown a failure of somatostatin to inhibit GHRH-stimulated GH release early in the third trimester and maturation of the inhibitory effect of somatostatin near term. [94] Thus, a predominant GHRH enhancement and limited somatostatin inhibition of hGH secretion at midgestation are presumably associated with a limited capacity for inhibition of hGH release by somatotropin feedback. In addition, there may be unrestrained hGH secretion at the pituitary cell level or immaturity of limbic and forebrain inhibitory circuitry that modulates hypothalamic function, or both. Whatever the mechanisms, control of hGH secretion matures progressively during the last half of gestation and the early weeks of postnatal life so that mature responses to sleep, glucose, and -dopa are present by 3 months of age.

The ontogenesis of fetal plasma PRL differs significantly from that of hGH (Fig. 21-4). Levels are low until 25 to 30 weeks of gestation and increase to a mean peak value of approximately 11 nmol/L at term. [95] Pituitary PRL content increases progressively from 12 to 15 weeks, and in vitro fetal pituitary cells from midgestation fetsuses show limited autonomous PRL secretion, although PRL release increases in response to TRH and decreases in response to dopamine. [95] Brain and hypothalamic control of PRL matures late in gestation and during the first months of extrauterine life. [95] Estrogen stimulates PRL synthesis and release by pituitary cells, and the marked increase in fetal plasma PRL concentration in the last trimester parallels the increase in fetal plasma estrogen levels, although lagging by several weeks. [95] Anencephalic fetuses have plasma PRL concentrations in the normal or low-normal range. [95] These data support a role for estrogen in stimulating fetal PRL release. The fetal sheep exhibits a similar pattern of fetal plasma PRL levels, indicating that maturation and integration of brain and hypothalamic mechanisms modulating PRL release develop late in gestation and in the postnatal period, accounting for the delayed postnatal
fall in plasma PRL level in the neonate of this species. [12]

The somatotropins IGF-I and IGF-II are important factors in fetal growth. The mRNA and protein for both factors are present early in gestation in essentially all fetal tissues. [13] IGF-II transcripts are more abundant than those of IGF-I and are predominant in fibroblasts and mesenchymal tissues. [14] Receptors for the IGFs are also widespread in fetal tissues. Studies of transgenic mice with null mutations of the genes encoding IGF-I, IGF-II, or the IGF-I receptor have defined the role of the somatotropins; the birth weight of the embryos lacking IGF-I or IGF-II was 60% of that of the control mice. [15] When both genes were inactive birth weight was reduced another 30%, and mice lacking IGF-I receptor had birth weights averaging 45% of control values. [16]

Postnatally, GH acts through receptors in liver and other tissues to stimulate production of IGF-I and, to a lesser degree, IGF-II. Prenatally, in contrast, GH receptor mRNA levels and receptor binding are low in fetal liver, although receptor mRNA is present in other fetal tissues. [17] The growth of adenocystic fetuses is nearly normal, however, suggesting that factors other than GH stimulate fetal somatomedin production. Nutritional factors are known to play a role. [18] PRL receptors are present in most fetal tissues during the first trimester of gestation, and it is likely that lactogenic hormones have a significant role in organ and tissue development early in gestation. [19] The coordinate increase in fetal adipose tissue and adipose tissue PRL receptors PRL1 and PRL2 suggests that PRL may play a role in growth and maturation of fetal adipose tissue later in gestation. [20] PRL also plays a role in fetal skeletal maturation. [21] Ovine placental lactogen stimulates glycogen synthesis in fetal ovine liver, and hPL stimulates amino acid transport, DNA synthesis, and IGF-I production in human fetal fibroblasts and muscle cells. GH and PRL have little activity in these tissues.

Fetal Pituitary-Adrenal System

The primordium of the adrenal gland can be recognized just cephalad of the developing mesonephros by 3 to 4 weeks of gestation. [22] [23] The fetal adrenal is composed of three functional zones. The zona glomerulosa begins to develop around 8 weeks of gestation, with the zona fasciculata and zona reticularis developing later. [24] Near term the fetal cortisol production rate in blood, per unit body weight, is similar to that in the adult. [25] Ovine placental lactogen stimulates glycogen synthesis in fetal ovine liver, and hPL stimulates amino acid transport, DNA synthesis, and IGF-I production in human fetal fibroblasts and muscle cells. GH and PRL have little activity in these tissues.

Development of the fetal adrenal cortex is under control of several genes and growth factors. The genes include those coding for the orphan nuclear receptors SF-1 (steroidogenic factor-1) and DAX-1 (dosage-sensitive sex reversal, adrenal hypoplasia congenita, X-chromosome factor). [26] [27] These genes show coordinate expression in adrenal cortex, testis, ovary, hypothalamus, and pituitary tissues. SF-1 gene knockout mice manifest adrenal and gonadal agenesis, gonadotropin deficiency, and absence of the hypothalamic ventromedial nucleus. [28] Inactivating DAX-1 gene mutations are associated with adrenal hypoplasia and gonadotropin deficiency. [29] The steroidogenic acute regulatory protein (STAR) is a rate-limiting factor in adrenal steroidogenesis. STAR knockout mice manifest glucocorticoid and mineralocorticoid deficiency and female genitalia in XY animals. [30] [31] In humans, inactivating STAR mutations cause adrenal hyperplasia and adrenal hormone insufficiency. [32]

Fetal adrenal transitional zone development is ACTH-dependent. [33] Proliferation of both the fetal and definitive zones is stimulated by fibroblast growth factor and EGf, and the fetal adrenal expresses high levels of IGF-II mRNA and protein, which are responsive to ACTH. [34] Moreover, IGF-II augments ACTH-stimulated expression of CYP11A1 (p450sc), CYP17, and 3HSD and stimulates cortisol and DHEA-S production in fetal adrenal cortical cells, suggesting a role in adrenal regulation during fetal and postnatal life. [35] The pattern of enzyme maturation in the fetal adrenal suggests that cortisol production does not occur de novo from cholesterol until 30 weeks of gestation. [36]

The major control of fetal adrenal function is mediated by fetal pituitary ACTH. Maternal levels of CRH are elevated during the last trimester of gestation and reach peak values of 0.5 to 1 mU/mL; at term; normal values in nonpregnant women are less than 0.01 mU/mL. [37] This CRH is bioactive and levels correlate with maternal cortisol concentrations, suggesting that CRH plays a role in stimulating maternal corticotropin release. Fetal plasma CRH levels at term, however, are approximately 0.03 mU/mL, and, relative to the presumably high levels in pituitary portal blood, probably have little role in modulating fetal corticotropin release. Midgestation fetal plasma corticotropin concentrations average about 35 pmol/L (250 pg/mL), levels that maximally stimulate fetal adrenal steroidogenesis, and concentrations are higher throughout gestation than in postnatal life, although they fall near term. [38] [39] [40]

Thus, the fetal adrenal cortex is maximally stimulated by pituitary corticotropin and produces large quantities of DHEA and pregnenolone and their sulfate conjugates. Much of the DHEA is converted to 16-Hydrox-DHEAS by the fetal adrenal and fetal liver. As already discussed, DHEA serves as a substrate for placental estrone and estradiol production; 16-Hydroxy-DHEA undergoes metabolism to estradiol in the placenta. In the anencephalic fetus, placental estrogen production is reduced to about 10% of normal. [41] An important factor in fetal adrenal function appears to be substrate inhibition of 3HSD activity by placental estrogens and intracellular adrenal steroids. [42] Near term the fetal cortisol production rate in blood, per unit body weight, is similar to that in the adult. [43] About two thirds of fetal cortisol is derived from the fetal adrenal glands, and one third is derived from placental transfer. [44] Both fetal adrenal cortisol and placental estradiol regulate hepatic synthesis of cholesterol in the fetus.
The corticotropin feedback control system matures progressively during the second half of gestation and the early neonatal period. Dexamethasone can suppress the human fetal parathyroid-adenal axis at term but not at 18 to 20 weeks of gestation. In the fetal sheep, hypothalamic and pituitary glucocorticoid receptors are present at midgestation and corticotropin suppressibility can be demonstrated by the midpoint of the third trimester of gestation. The number of glucocorticoid receptors in the pituitary gland increases at term at the time of increasing glucocorticoid levels, suggesting that some process in the fetus allows the normal autoregulation of glucocorticoid receptors to be overridden at term.

Adrenal hormone receptors, including glucocorticoid receptors (GRs) and mineralocorticoid receptors (MRs), are members of the nuclear receptor superfamily of steroid hormone, thyroid hormone, vitamin D, and retinoid receptors. GRs are present in most body tissues by the second trimester and play an important role in fetal development. Mice lacking GR receptor function manifest enlarged and disorganized adrenal cortices, adrenal medullary atrophy, lung hypoplasia, and defective gluconeogenesis. They appear normal at birth but do not survive without treatment.

Fetal cortisol is converted to cortisone through an 11βHSD in fetal tissues, and levels of circulating cortisone in the fetus at midgestation are fourfold to fivefold higher than cortisol concentrations. Cortisone is a relatively inactive glucocorticoid, and this metabolism protects the anabolic milieu of the fetus because cortisol can retard both placental and fetal growth. GRs are present at birth and are probably present at midgestation in most tissues, including placenta, lung, brain, liver, and gut. As term approaches, selected fetal tissues including liver and lung express 11β-hydroxysteroid dehydrogenase activity that promotes local conversion of cortisone to cortisol. Cortisol serves as an important stimulus to prepare the fetus for extraterine survival. The increase in fetal cortisol conversion occurs during the last 10 weeks of gestation and is the result of increased cortisol secretion and decreased conversion of cortisol to cortisone. This increase in fetal cortisol production has an important role in the maturation of several fetal systems or functions that are critical to extraterine survival (see "Transition to Extraterine Life").

The human fetal adrenal gland is capable of aldosterone secretion near term, and fetal plasma aldosterone concentrations in infants who are born by cesarean section are threefold to fourfold higher than maternal levels. Vaginal delivery and maternal salt restriction increase levels in both mother and infant. The increased aldosterone levels in the fetus are due to increased fetal adrenal secretion and persist during the first year of extraterine life. However, there is a poor correlation between plasma renin activity (PRA) and aldosterone levels in cord blood. Aldosterone secretion is low in the midgestation human fetal adrenal and is unresponsive to the secretagogues that are known to modulate aldosterone production in the adult. In sheep, fetal aldosterone becomes responsive to PRA and angiotensin II in the neonatal period. In this species, in which late fetal aldosterone levels are also high compared with adult levels, furosemide stimulates PRA but not aldosterone during the third trimester; the aldosterone response to furosemide (and PRA) is delayed until the neonatal period. This situation also appears to be the case in the human fetus and neonate.

MRs are present in fetal tissues from 12 to 16 weeks of gestation. MR immunoreactivity is detectable in fetal kidney, skin, hair follicles, trachea and bronchioles, esophagus, stomach, small intestine, colon, and pancreatic exocrine ducts. The role of MRs in these fetal tissues remains unclear. MR knockout mice appear normal at birth but demonstrate defects in mineralocorticoid and renin-angiotensin system functions in the postnatal period.

Angiotensin II levels in the sheep fetus are similar to maternal values, and blockade of fetal production with angiotensin-converting enzyme inhibitors decreases the fetal glomerular filtration rate. Two subtypes of angiotensin receptors, AT1 and AT2, are detectable in various tissues early in fetal development. AT1 receptor mRNA expression in the fetal sheep kidney is low early in gestation, increases in the latter third of pregnancy, and decreases postnatally; AT2 mRNA levels, in contrast, are high at midgestation and decrease during the third trimester. These changes are believed to reflect growth factor-mediated changes in cells that contain AT2 in various tissues. Hormonal factors modulate fetal renal AT1 gene expression in sheep; angiotensin II suppresses both AT1 and AT2, and cortisol increases AT1 gene expression.

The role of the fetal renin-angiotensin system is not clear; rather than modulating renal sodium excretion through aldosterone, it may maintain renal excretion of salt and water into amniotic fluid to prevent oligohydramnios. This renal effect is presumably mediated by modulation of arterial pressure. The mechanism for the high aldosterone levels in the fetal and neonatal periods remains unclear. Because plasma arial natriuretic factor concentrations are high in the fetus, the increased PRA and aldosterone levels are not due to relative arial natriuretic factor deficiency.

Aldosterone affects renal sodium excretion and in premature infants. Despite the fact that the newborn human kidney is relatively unresponsive to exogenous aldosterone, manifestations of mineralocorticoid deficiency in the newborn term infant can occur as a result of aldosterone deficiency or competition for binding to renal MRs by other steroids such as 17-hydroxyprogesterone. Relatively reduced glomerular filtration in the newborn limits sodium loss initially, but by 1 week of age aldosterone deficiency produces the characteristic manifestations of hyponatremia, hyperkalemia, and volume depletion.

**Fetal Pituitary Thyroid System**

The thyroid gland is a derivative of the primitive buccopharyngeal cavity and develops from contributions of two anlagen; a midline thickening of the pharyngeal floor (median anlage) and paired caudal extensions of the fourth pharygobranchial pouches (lateral anlagen). These structures are discernible by 16 to 17 days of gestation, and by 24 days the median anlage develops a thin, flaslike diverticulum extending from the floor of the buccal cavity to the fourth branchial arch. By 50 days of gestation, the median and lateral anlagen have fused and the buccal stalk has ruptured. During this period the thyroid gland migrates caudally to its definitive location in the anterior neck. By 70 days of gestation, colloid is visible histologically and thyroglobulin synthesis and iodide accumulation can be demonstrated within the gland. During the final follicular phase of development, colloid spaces increase in size and there is progressive cell growth and accumulation of thyroid hormones. At 12 weeks of gestation the fetal thyroid gland weighs about 80 mg, and at term it weighs 1 to 1.5 g.

![Figure 21-7](https://example.com/figure21-7.png) **Figure 21-7** Cartoon showing the homeobox genes programming development of the thyroid and parathyroid glands. HOXB3 may be responsible for activation of thyroid transcription factor 1 (TTF1) during early embryogenesis, with TTF2 and PAX8 involved in a synergetic cascade programming thyroid gland embryogenesis. These factors are also involved in thyroid follicul cell function, promoting thyroglobulin (TG), thyroid peroxidase (TPO), and thyroid-stimulating hormone receptor (TSHR) gene transcription. HOX15 gene knockout in mice causes parathyroid gland aplasia. See text for details.

The parathyroid glands develop between 5 and 12 weeks of gestation from the third and fourth pharyngeal pouches. The third pouches encounter the migrating thyroid anlage, and the parathyroid anlagen are carried caudally with the thyroid gland, finally coming to rest at the lower poles of the thyroid lobes as the inferior parathyroid glands. The fourth pouches encounter the thyroid anlage later and come to rest at the upper poles of the thyroid lobes as the superior parathyroid glands. The individual parathyroid glands increase in diameter from less than 0.1 mm at 14 weeks of gestation to 1 mm at birth. The fifth pouches contribute paired ultimobranchial bodies that are incorporated into the developing thyroid gland as the parafollicular or C cells that secrete calcitonin.

Four or more homeobox genes are involved in thyroid and parathyroid gland embryogenesis. These include the genes for thyroid transcription factors 1 and 2 (TTF1, TTF2) and PAX8. TTF2 gene knockout in mice results in thyroid dysgenesis and cleft palate. TTF1 knockout produces pulmonary hypoplasia and thyroid agenesis. Inactivating PAX8 mutations produce thyroid hypoplasia and renal anomalies. TTF1 knockout also produces parafollicular C-cell aplasia. The HOX genes appear to be important in the expression of TTF1 and PAX8. HOX15 gene disruption in mice results in parathyroid gland aplasia. TTF2 and PAX8 gene mutations have been identified in 2% of patients with familial thyroid dysgenesis and congenital hypothyroidism. However, most cases of congenital...
Hypothyroidism occurs sporadically, and the pathogenesis in these cases remains unclear.

During the first half of gestation, before the onset of significant fetal thyroid hormone production, fetal T₄ is derived from maternal-fetal-placental transfer. T₃ is detectable in human coelomic fluid at levels of 0.5 to 2 nmol/L between 6 and 11 weeks of gestation, before the onset of fetal thyroid function. At term, serum T₃ levels in the athyroid fetus range from 30 to 70 nmol/L (2.3 to 5.4 µg/dL). Isotopic equilibrium studies with pregnant rats at term suggest that 15% to 20% of the T₃ in fetal tissues is of maternal origin.

Pituitary and plasma thyrotropin (TSH) concentrations begin to increase during the second trimester in the human fetus, about the time that pituitary portal vascular continuity develops progressively during the last trimester. Metabolites in the fetus are iodothyronine sulfates. After birth indicates that significant amounts of circulating rT₃ are produced by fetal tissues rather than by the placenta.

The type III enzyme is expressed in most fetal tissues and in the placenta early in gestation and is responsible for production of the high levels of fetal plasma rT₃. Type II monodeiodinase activity is present in the brain, pituitary, and brown adipose tissue is a low-Kₗ enzyme insensitive to propylthiouracil and inhibited by thyroid hormone. Type III monodeiodinase in brain, pituitary, and brown adipose tissue is a low-Kₗ enzyme insensitive to propylthiouracil and inhibited by thyroid hormone. Type III monodeiodinase in liver, heart, skin, and placenta is responsible for inner-ring deiodination of T₃ to rT₃ and of T₄ to diiodothyronine. The type I monodeiodinase is largely responsible for production of T₂ that escapes from the cells into the circulation, whereas the type II enzyme is responsible for production of local T₃ in brain, pituitary, and brown adipose tissue. rT₃ also diffuses out of most tissues to appear in plasma.

The metabolism of thyroid hormones occurs through a progressive series of monodeiodinations. Several enzymes act on the iodines in the outer (phenolic) ring or the inner (tyrosyl) ring of the diiodothyronine molecule. Most of the circulating, biologically active T₃ in adults is derived by outer-ring monodeiodination of T₄ in liver and other nonthyroidal tissues; biologically inactive rT₃ derives from inner-ring deiodination of T₄ in peripheral tissues. Three iodothyronine monodeiodinase subtypes have been characterized. Type I, an outer-ring monodeiodinase in liver and kidney, is a high-Michaelis-constant (Kₘ) enzyme inhibited by propylthiouracil and stimulated by thyroid hormone. This enzyme also has innerning deiodinating activity and catalyzes the conversion of T₃ to 3,3,5'-triiodothyronine. Type II outer-ring monodeiodinase in brain, pituitary, and brown adipose tissue is a low-Kₗ enzyme insensitive to propylthiouracil and inhibited by thyroid hormone. Type III monodeiodinase in liver, heart, skin, and placenta is responsible for inner-ring deiodination of T₄ to rT₃, and of T₃ to diiodothyronine. The type I monodeiodinase is largely responsible for production of T₂ that escapes from the cells into the circulation, whereas the type II enzyme is responsible for production of local T₃ in brain, pituitary, and brown adipose tissue. rT₃ also diffuses out of most tissues to appear in plasma.

The type III enzyme is expressed in most fetal tissues and in the placenta early in gestation and is responsible for production of the high levels of fetal plasma rT₃, which peak at midgestation in the range of 3 to 4 nmol/L (200 to 300 ng/dL). The persistence of elevated plasma levels of rT₃ in the neonate for several weeks after birth indicates that significant amounts of circulating rT₃ are produced by fetal tissues rather than by the placenta.

There is little conversion of T₄ to circulating T₂ in the midgestation human fetus; plasma T₂ levels are low (<0.2 nmol/L, <15 ng/dL) until 30 weeks of gestation, after which the mean value increases to 0.7 nmol/L (50 ng/dL) at term (Fig. 21-9). Sulfation is active in fetal tissues, and the predominant thyroid hormone metabolites in the fetus are iodothyronine sulfates. In the last third of gestation in fetal sheep, the mean plasma production rates for T₄ and metabolites are T₄ = 40, T₄ sulfate (T₄-S) = 10, rT₃ = 5, rT₃-S = 12, T₂ = 2, and T₃-S = 2 µg/kg body weight per day. All are biologically inactive except for T₂, and perhaps T₂-S so that 90% of the T₃ metabolites in the fetus are biologically inactive. The sulfated metabolites accumulate in fetal serum as a result of the low type I monodeiodinase activity in fetal tissues and because the sulfated sulfates are not substrates for type III monodeiodinase.

The production rate of T₂ increases progressively between 30 weeks of gestation and term because of maturation of type I monodeiodinase activity in the liver and other tissues and because of decreasing type III monodeiodinase activity in placenta. In the fetal sheep, hepatic type I monodeiodinase activity increases progressively during the last trimester. Type II monodeiodinase activity is present in the brain at

Figure 21-8 Patterns of change of fetal plasma thyroid-stimulating hormone (TSH), thyroid hormone (T₄), triiodothyronine (T₃), and iodothyronine sulfates (T₄-S, rT₃-S, and T₃-S) during gestation and in the neonatal period. The patterns for T₄-S and rT₃-S are based on limited 35-week data. (Data from Fisher DA, Klein AH. N Engl J Med 1991; 324:72712; Santini P, Chiocchio L, Ghieri P, et al. J Clin Endocrinol Metab 1989; 84:493498; Burrow GN, Fisher DA, Larsen PR. N Engl J Med 1994; 331:10721078.)

Figure 21-9 Patterns of change of plasma levels of human chorionic gonadotropin (HCG), luteinizing hormone (LH), testosterone (T₃), and estradiol (E₂) in a male fetus during gestation and in the neonatal period. (Data from Reyes FI, Boroditsky RS, Winter JS, et al. J Clin Endocrinol Metab 1974; 38:612167; Kaplan SL, Grumbach MM, Aubert ML. Recent Prog Horm Res 1995; 51:181206.)
midgestation and helps guarantee adequate brain T<sub>2</sub> in the sheep, a species in which brain maturation depends on thyroid hormone during the second half of gestation.  

Two genes code for the thyroid hormone receptors TR and TR, members of the steroid, retinoid, vitamin D family of nuclear transcription factors. Alternative splicing of expressed mRNA species leads to production of several TR isoforms. The major isoforms, TR<sub>1</sub>, TR<sub>2</sub>, TR<sub>1</sub>, and TR<sub>2</sub>, are developmentally regulated and are present in characteristic concentration ratios in various adult tissues. In human fetal brain, low levels of receptor binding have been detected at 10 weeks of gestation, with higher levels at 16 to 18 weeks. Liver, heart, and lung receptor binding has also been identified at 16 to 18 weeks.  

Human fetal growth is the net result of a complex interplay of genetic, hormonal, and growth factor effects, which are independent of thyroid hormone. Bone maturation levels of the hypothalamic infant, however, is delayed in 50% to 60% of cases and fontanelle closure is often delayed. Serum TSH concentrations are characterized by high levels of variation. Congenitally hypothyroid infants with marked hypothyroxinemia may manifest prolonged jaundice, lethargy, feeding difficulties, umbilical hernia, or macroGLOSSIA, but the classical signs and symptoms of congenital hypothyroidism, including myxedema, metabolic derangements, growth retardation, and irreversible mental and neurologic dysfunction, accrue progressively during the early months and years of life. The relative lack of signs and symptoms in the athyroid fetus and infant is probably related to the effects of transplacentally acquired maternal T<sub>2</sub>, which at term provides an estimated 20% of fetal thyroid hormone turnover. Developmental regulation of T<sub>2</sub> mediated transcriptional events in various organs and tissues and tissue-specific regulation of deiodinase activity are additional mechanisms.  

However, thyroid hormone is essential for normal central nervous system maturation. It regulates a diverse array of processes, including neurogenesis and neural cell migration, neuronal differentiation, dendritic and axonal growth, synaptogenesis, gliogenesis, myelination, and neurotransmitter enzyme synthesis. The most subtle effect of thyroid hormone on fetal brain development is observed in pregnancies associated with maternal hypothyroxinemia. Even subclinical maternal hypothyroidism has been associated with 5- to 10-point I.Q. deficits in the offspring.  

The molecular mechanisms by which thyroid hormone mediates these central nervous system effects remain unclear. Aside from a hearing impairment, TR knockout mice exhibited neither morphologic or functional abnormalities of brain development nor significant differences in mRNA levels of a number of T<sub>2</sub>-dependent cerebellar genes. No gross defects in behavior or in myelin have been reported to date in TR knockout mice. Thus, the major thyroid hormone effects on brain maturation are probably mediated by pathways common to both TR genes.  

Fetal Pituitary-Gonadal Axis  

The mammalian gonad is derived from two tissue anlagen, the primordial germ cells of the yolk sac wall and somatic, stromal cells that migrate from the primitive mesonephros. By 4 to 5 weeks of gestation, the germ cells have begun their migration from the yolk sac and the gonadal ridge has appeared as a derivative of the mesonephros. The germ cells are incorporated into the developing gonadal ridge during the sixth week, when the primitive gonad is composed of a surface epithelium, primitive gonadal cords continuous with the epithelium, and a dense cellular mass referred to as the gonadal blastema.  

Embryogenesis of the gonads is programmed by genes coding for the male sexual determinant SRY as well as SF-1 and DAX-1. SRY is the single critical regulator of male gonadal differentiation (see Fig. 21-5). SF-1 is also required for testicular and ovarian development and mediates müllerian-inhibiting hormone gene expression and gonadotropin production. SF-1 and DAX-1 are orphan receptors of the steroid-thyroid hormone family of nuclear receptors and appear to interact as heterodimers coordinately involved in the regulation of target genes in the adrenal glands and in hypothalamic gonadotropin cells and the ventromedial hypothalamic nucleus. The hierarchical pathway or pathways for these genes' programming events and the full menu of downstream gene targets remain to be defined. The net result, however, is the highly organized pattern of gonadal development and phenotypic sexual differentiation. Fetal pituitary gonadotropins are not required for gonadal development or sexual differentiation; LH or FSH receptor knockout mice are born phenotypically normal.  

Male gonadal differentiation begins at 7 weeks of gestation with organization of the gonadal blastema into interstitium and germ cell-containing testicular cords. The primitive cords lose their connections with the epithelium, primitive Sertoli cells and spermatogonia become visible within the cords, and the epithelium differentiates to form the tunica albuginea. Leydig cells derived from the undifferentiated interstitium are visible by the end of the eighth week of gestation and are capable of androgen synthesis at this time. By 14 weeks of gestation these cells make up as much as 50% of the cell mass, but as the tubules develop they account for a smaller percentage of the tissue. The fetal testes grow from approximately 20 mg at 14 weeks of gestation to 800 mg at birth; at 5 to 6 months they descend into the inguinal canal in association with the epididymis and the ductus deferens. Targeted disruption of the INS13 gene in mice impairs gubernaculum development and leads to bilateral cryptorchidism.  

In females, differentiation of ovaries begins during the seventh week of gestation. The gonadal blastema differentiates into interstitium and medullary cords containing the primitive germ cells now referred to as oogonia. The cords degenerate and cortical layers of surface epithelium, containing individual small oogonia, appear. By 11 to 12 weeks of gestation clusters of dividing oogonia are surrounded by cord cells within the cortex; the medulla at this time consists largely of connective tissue. At 12 weeks of gestation, primitive granulosa cells begin to replicate and many of the large oogonia in the deepest layers of the cortex enter their first meiotic division; other oogonia degenerate and their cords undergo the first meiotic division to become primary oocytes. Primordial follicles are present by 5 months of gestation, and during the seventh month stroma-derived thecal cells develop around the primordial follicles as they mature to primary follicles. This process continues after birth, again progressing toward the superficial layers. Each fetal ovary weighs about 15 mg at 14 weeks of gestation and 300 to 350 mg at birth.  

In the male, the development of Leydig cells leads to an increase in fetal testosterone production between gestational weeks 10 and 20. In vivo studies in the rat have shown that GC binding to fetal testis cells does not down-regulate LH receptors. If this is true in vivo, continuous exposure of the Leydig cell to HCG would not desensitize the fetal testis and would allow the maintenance of augmented testosterone production during development. Fetal LH may contribute to fetal Leydig cell function, but quantitatively HCG is the predominant gonadotropin. Testosterone itself, acting through the androgen receptor, stimulates differentiation of the primitive mesonephric ducts into bilateral ductus deferens, epididymides, seminal vesicles, and ejaculatory ducts. Dihydrotestosterone stimulates male differentiation of the urogenital sinus and external genitalia, including differentiation of the prostate, growth of the genital tubercle to form a phallus, and fusion of the urogenital folds to form the external genitalia. As opposed to the phallus, the urogenital tubercle is masculinized by the 5-reductase-dependent pathway within the urogenital sinus and urogenital tubercle and acts through the same androgen receptor that mediates the action of testosterone in the Wolffian ducts.  

The fetal testes also produces antimüllerian hormone (AMH), which causes dedifferentiation of the müllerian duct system in the male fetus. AMH is a glycoprotein with a monomer molecular size of approximately 72 kd and multimer sizes ranging from 145 to 235 kd. It is produced by testosterone Sertoli cells and reaches the müllerian ducts largely by diffusion; duct regression in vitro requires a 24- to 36-hour exposure to AMH, which is synthesized early in gestation.
most other tissues. The significance of estrogen receptors in fetal development remains unclear. Knockout of the ER gene in mice did not impair fetal development, but adult females were infertile with hypoplastic uteri and polycystic ovaries and adult males manifested decreased fertility. ER knockout mice also developed normally and female adults were fertile with normal sexual behavior; adult males reproduced normally but had prostate and bladder hyperplasia. It is known that estrogens regulate DHEA production in the baboon and human fetal adrenal. Studies of mice with knockout of both ER and ER functions should further clarify the role of these receptors in fetal development.

Gonadal hormones also control gonadotropin production in the brain that results in cyclic ovarian function and normal function of the testes. Testosterone administration to neonatal female rats produces permanent inhibition of cyclic hypothalamic control through local aromatization to estradiol and estrogen receptor binding. In primates and humans, estrogens seem to be more effective in this regard. However, there is no evidence for permanent programming in the primate, and there appear to be no major tissue biochemical differences between the sexes in utero to account for sexual dimorphic behavioral or gonadotropic programming. Thus, the mechanisms for these effects are not yet clear in the primate and human fetus.
Intermediate Lobe of the Pituitary

The intermediate lobe of the pituitary gland is prominent in both the human and the sheep fetus. Intermediate lobe cells begin to disappear near term and are virtually absent in the adult human pituitary, although the intermediate lobe in the adult of some lower species is anatomically and functionally distinct. The major secretory products of the intermediate lobe are -MSH and -endorphin derived from cleavage of the POMC molecule. Cleavage of POMC in the anterior lobe results predominately in corticotropin and -lipotropin formation.

In rhesus monkeys and humans, the fetal pituitary contains high concentrations of compounds resembling -MSH and corticotropin-like intermediate lobe peptide. -MSH levels in the human fetus decrease with increasing fetal age. The circulating levels of both -endorphin and -lipotropin are high in the fetal lamb, and the ratio of -endorphin to -lipotropin increases during hypoxic stimulation of the anterior pituitary. Because hypoxia provokes corticotropin release and -lipotropin production from the anterior pituitary, these data have been interpreted to suggest that basal -endorphin levels in the fetus originate in the intermediate lobe. -MSH and corticotropin-like intermediate lobe peptide may play a role in fetal adrenal activation, and -MSH may play a role in fetal growth. However, these effects are probably minor; the processing of pituitary POMC in the human fetus by the end of the second trimester is similar to that in the adult, but the role of these intermediate lobe peptides in the fetus remains obscure.
Posterior Pituitary

The fetal neurohypophysis is well developed by 10 to 12 weeks of gestation and contains both arginine vasopressin (AVP, also called antidiuretic hormone) and OT. In addition, arginine vasotocin (AVT), the parent neurohypophyseal hormone in submammalian vertebrates, is present in the fetal pituitary and pineal glands and in adult pineal glands from several mammalian species, including humans. AVT is present in the pituitary during fetal life and disappears in the neonatal period. In adult mammals, instillation of AVT into cerebrospinal fluid inhibits gonadotropin and corticotropin release, stimulates PRL release by the anterior pituitary, and induces sleep; however, its physiologic importance in these regards remains unclear. The role of AVT in the fetal pineal gland is unknown.

In the fetal sheep, the baseline fetal plasma AVP concentrations are similar to maternal levels after midgestation. During the last trimester of gestation, fetal hypothalamic and pituitary responsiveness to both volume and osmolar stimuli for AVP secretion are well developed and AVP exerts antidiuretic effects on the fetal kidney. Baseline plasma levels of AVT in fetal sheep during the last trimester approximate values for AVP and OT. Presumably this AVT is derived from the posterior pituitary, but the stimuli for AVT secretion in the fetus are not defined. The neurohypophysial peptides are synthesized as large precursor molecules (neurophysins) and processed to bioactive amidated peptides. Enzymatic processing involves progressive cleavage of carboxyl terminal extended peptides producing sequentially (for OT) OT-glycine-lysine-arginine (OTGKR), OTGK, OTG, and OT. Similar progressive processing yields AVPG and AVP from the AVP neurophysin. Enzymatic processing of neurophysins matures progressively in the fetus so that early in gestation fetal plasma contains relatively large concentrations of the extended peptides. For OT, the ratio of OT extended peptides to OT in fetal sheep serum is approximately 35:1 early in gestation and 3:1 late in gestation.

In the fetus, AVP appears to function as a stress-responsive hormone. Perhaps the major potential stress for the fetus is hypoxia, and the response of AVP to hypoxia is increased compared with the maternal response and with the fetal AVP responses to osmolar stimuli. Plasma AVP concentrations in human cord blood are elevated in association with intrauterine bradycardia and meconium passage. The vasopressor action of AVP may be important in the maintenance of fetal circulatory homeostasis during hemorrhage and hypoxia. AVP has a limited effect on fetoplacental blood flow. Fetal hypoxia is also a major stimulus for catecholamine release. There is little information on interaction between AVP and catecholamines during fetal hypoxia, but both fetal hypoxia and AVP stimulate anterior pituitary function. A role for AVP as a CRH is established in the adult, and the ovine fetal pituitary responds separately and synergistically to AVP and CRH early in the third trimester. The role of AVP in controlling fetal corticotropin release seems to decrease with gestational age. It is not known whether AVT functions as a fetal CRH.

OT receptors have been demonstrated in human fetal membranes at term, and AVP receptors have been found in renal medullary membranes of newborn sheep. Both AVP and AVT evoke antidiuretic actions in the sheep fetus during the last third of gestation, and both hormones act to conserve water for the fetus by inhibiting fluid loss into amniotic fluid through the lungs and kidneys. Aquaporin-1, aquaporin-2, and aquaporin-3 water channel receptors are present in the human fetal and newborn kidney, and the ability of the newborn infant to regulate free water clearance in response to volume and osmolar stimuli has been demonstrated. Whether AVT exerts its effects through AVP receptors or separate fetal AVT receptors is not clear. Maximal concentrating capacity by the fetal kidney is limited to about 600 mmol/L. This limitation is due not to inadequate AVP stimulation but rather to inherent immaturity of the renal tubules.
Fetal Autonomic Nervous System

The primordia of the sympathetic trunk ganglia are visible in the human fetus by 6 to 7 weeks of gestation. The preaortic sympathetic primordia at this time are composed of primitive sympathetic neurons and chromaffin cells, which condense into chains of cell masses along the abdominal aorta. By 10 to 12 weeks of gestation the paired adrenal masses are well developed. In addition, numerous extramedullary paraganglia (derived from preaortic condensations of sympathetic neurons and chromaffin cells) are scattered throughout the abdominal and pelvic sympathetic plexuses. Each of these extramedullary paraganglia may reach a maximal diameter of 2 to 3 mm by 28 to 30 weeks of gestation. The largest of the paraganglia, the organs of Zuckerkandl near the origin of the inferior mesenteric arteries, enlarge to 10 to 15 mm in length at term. After birth, the paraganglia gradually atrophy and disappear by 2 to 3 years of age. With increasing gestational age there is progressive growth of the adrenal medullae, increasing catecholamine content of the adrenal medullae, and progressive maturation of medullary functional capacity. Histologically, the adrenal medullae are somewhat immature at birth, but by the age of 1 they resemble the adult glands.

Both chromaffin and sympathetic nerve cells are derived from common neuroectodermal stem cells, and both respond to NGF. Sympathetic nervous system development is NGF-dependent, and injections of NGF antiserum into neonatal rats lead to degeneration of immature chromaffin cells, sympathetic cells, and pheochromoblasts. Whether NGF and other growth factors are involved in the transient life span and function of the paraganglia in the human fetus and neonate remains to be clarified. The role of placental NGF in maturation of the fetal autonomic nervous system is also unclear.

Catecholamines are present in the para-aortic chromaffin tissue by 10 to 15 weeks of gestation, and concentrations increase until term. The predominant catecholamine is norepinephrine (NE), presumably because of low activity of phenylethanolamine N-methyltransferase in para-aortic chromaffin tissue. This enzyme, which catalyzes the methylation of NE to epinephrine, appears to be activated by the high levels of cortisol that diffuse into the adrenal medulla from the adrenal cortex; in contrast, cortisol levels in extramedullary chromaffin tissue are low.

In fetal mammals, the chromaffin cells of the adrenal medulla can respond directly to asphyxia, long before splanchnic innervation develops, by secreting NE; the noninnervated para-aortic tissue responds similarly. In the fetal sheep, a similar developmental transition occurs between days 120 and 135 of the 150-day gestation. The central nervous system responds to stimuli that evoke sympathetic nervous system responses before the adrenomedullary splanchnic innervation, but the adrenal medulla is relatively unresponsive to such stimuli. The transition is heralded by an adrenomedullary response to hypoglycemia mediated by the central nervous system. This response is present in developing sheep, monkeys, and human fetuses during the third trimester of gestation. Central and adrenal enkephalins are also involved in fetal autonomic nervous system function, and pretreatment with naloxone potentiates and methadone inhibits the catecholamine response to hypoxia.

Basal plasma epinephrine, NE, and dopamine levels during the last third of gestation decrease as term approaches. The metabolic clearance rate of epinephrine increases with gestational age, whereas the production rate remains unchanged, indicating that the decreasing basal catecholamine levels that occur with fetal age are due to maturation of clearance mechanisms. The fetal sheep responds to maternal exercise or hypoxia with increased catecholamine levels. The human neonate responds to parturition with an increase in plasma epinephrine and NE concentrations, and these responses are augmented by hypoxia and acidosis.

Catecholamines are critical for fetal cardiovascular function and fetal survival. Gene knockout studies in mice targeting either tyrosine hydroxylase or dopamine -hydroxylase produced fetal catecholamine deficiency and midgestation fetal death in 90% of the mutant embryos. In addition, fetal catecholamines are the major stress hormones in the fetus. The fetal adrenal and the para-aortic chromaffin masses discharge large amounts of catecholamines directly into the circulation in response to fetal hypoxia. Moreover, the defense against fetal hypoxia involves catecholamine actions mediated through cardiac -receptors that are unique to immature animals. -Adrenergic receptors predominate in immature cardiac tissue and gradually decline in number as -adrenergic receptors increase with maturation. Chromaffin tissue in the fetus is also innervated by opiate receptors and contains relatively large amounts of opiate peptides that appear to be co-secreted with the catecholamines. The extent to which these peptides or pituitary endorphins are involved in modulating fetal catecholamine secretion remains unclear.
Parathyroid Hormone/Calcitonin System

Parathyroid gland development from the third and fourth pharyngeal pouches proceeds in synchrony with thyroid embryogenesis. Disruption of the HOX15 gene in mice resulted in parathyroid gland aplasia, indicating that this gene functions as part of the gene cascade programming normal thyroid-parathyroid gland development. Both endocrine systems are functional during the second and third trimesters.

Studies in fetal sheep and monkey and measurements in human preterm and term infants indicate that high concentrations of fetal calcium (averaging 2.75 to 3 mmol/L in the last trimester) are maintained by active placental transport from maternal blood. The transport of calcium occurs across the syncytiotrophoblast, which contains a calcium-binding protein that buffers intracellular calcium ions as they are transported across the syncytiotrophoblast. Adenosine triphosphatedependent calcium pump transports the calcium across the cell membrane to the fetal circulation. The placental calcium pump is stimulated by a midmolecule portion of the PTH/PTHrP secreted by the fetal parathyroid gland and by the placenta, where it may exert a paracrine effect. The placenta is impermeable to PTH, PTHrP, and calcitonin, but 25-hydroxyvitamin D and 1,25-dihydroxyvitamin D are transported across the placenta and free vitamin D levels in fetal blood are similar to or higher than maternal values.

Thyroparathyroidectomy in the sheep fetus caused a rapid decrease in fetal plasma calcium concentration and a loss of the placental calcium gradient. In mice, knockout of the gene for PTH/PTHrP abolished the maternal-fetal calcium gradient and placental transport of calcium was reduced. Placental calcium transport in these models was restored by the midmolecule fragment of PTH/PTHrP (amino acids 67 to 86) but not by PTH or PTHrP fragments 1 to 34, which activate the PTH/PTHrP receptor.

Other factors are also involved in maintenance of fetal serum calcium levels because knockout of the mouse gene for PTH-PTHrP also results in hypocalcemia in the presence of normal or increased placental calcium transport. PTH and PTHrP, through the PTH-PTHrP receptor, presumably modulate fetal skeletal calcium flux, calcium excretion through the fetal kidney, and perhaps reabsorption of calcium from amniotic fluid. PTHrP has a major role in fetal bone development and metabolism as well as fetal calcium homeostasis. PTHrP knockout mice displayed increased ossification of the basal portion of the skull, long bones, vertebral bodies, and pelvic bones and mineralization of the normally cartilaginous portions of the ribs and sternum; as a result of the cartilaginous mineralization, the animals died of asphyxiation in the early neonatal period.

Fetal nephrectomy also reduces fetal calcium concentrations, and the hypocalcemia can be prevented by administration of 1,25-dihydroxyvitamin D. Moreover, infusion into the sheep fetus of antibody to 1,25(OH)₂D reduced the placental calcium gradient. Thus, fetal PTHrP and PTH appear to stimulate fetal renal 1,25(OH)₂D production, which acts to enhance maternal-fetal transport of calcium by the placenta. The fetal kidney can synthesize 1,25(OH)₂D via 1-hydroxylation of 25-hydroxycholecalciferol, and the placenta contains both 1,25(OH)₂D receptors and a vitamin D-dependent calcium-binding protein. In the sheep fetus, the endogenous production rate of 1,25(OH)₂D during the last third of gestation was six times greater than that in the mother. The metabolic clearance of 1,25(OH)₂D was also higher in the fetus than in the mother.

The fetal parathyroid-placental axis promotes maternal-fetal transfer of bone mineral and accretion of fetal bone mineral. The high blood levels of calcitonin in the fetus, probably resulting from the chronic stimulation by fetal hypercalcemia, are thought to contribute to the fetal bone mineral accreration. A prominent effect of calcitonin is to inhibit bone resorption, and the high fetal serum calcium concentrations coupled with high circulating calcitonin promote bone mineral anabolism. Calcitonin has been called a vestigial hormone because of its limited role in postnatal calcium regulation, but it may have an important role in the fetus. Placental calcitonin production may contribute to the calcitonin in fetal plasma, but the persistence of high plasma levels in neonatal plasma argues for predominant fetal production. Also, 1,25(OH)₂D or 24,25(OH)₂D may play a role in fetal cartilage growth and bone mineral accreration. These concepts are summarized in Figure 21-10.
Endocrine Pancreas: Insulin and Glucagon

Embryogenesis of the pancreas is mediated by a series of homeobox genes, including IDX1, ISL1, PAX4, and PAX6. IDX1 gene knockout in the mouse results in pancreatic agenesis. ISL1 knockout leads to islet cell agenesis, and PAX4 or PAX6 knockout results in beta cell or alpha cell agenesis or hypogenesis, respectively. The 2neuroD gene knockout in the mouse leads to marked beta cell dysplasia and hypoplasia and early death from diabetes. These factors and perhaps others farther downstream program the orderly maturation of pancreatic development and function (Fig. 21-11). The fetal pancreas is identifiable by 4 weeks of gestation, and alpha and beta cells can be recognized by 8 to 9 weeks. Insulin, glucagon, somatostatin, and pancreatic polypeptide are measurable by 8 to 10 weeks of gestation. Alpha cells are more numerous than beta cells in the early fetal pancreas and reach a peak at midgestation; beta cells increase throughout the second half of gestation so that by term the ratio of alpha cells to beta cells is approximately 1.1. The insulin content of the pancreas increases from less than 3.6 pmol/g (0.5 U/g) at 7 to 10 weeks to 30 pmol/g (4 U/g) at 16 to 25 weeks of gestation and 93 pmol/g (13 U/g) near term; the concentration in the adult pancreas is approximately 14 pmol/g (2 U/g). Although the fetal beta cell is functional by 14 to 24 weeks of gestation, secretion of insulin by the fetal pancreas is low. Insulin release from the fetal rat pancreas in vitro in response to glucose or pyruvate is minimal but can be stimulated by leucine, arginine, tolbutamide, or potassium chloride, indicating that parts of the secretory mechanism are functional in the fetus. Insulin secretion in adult islets is mediated by two or more mechanisms, including stimulation of the adenylate cyclase system with production of cAMP and inhibition of potassium efflux, which leads to depolarization of the cell membrane and opening of voltage-dependent calcium channels. The former mechanism, although suppressed in the fetal islets, can be augmented by theophylline, but calcium channel activation does not occur in fetal islets in response to inhibitors of insulin release that cause depolarization of adult islet cells. The infusion of glucose or arginine in pregnant women before hysterotomy fails to provoke fetal insulin secretion at midgestation or near term, and plasma insulin levels in the late human fetus are relatively unresponsive to high glucose concentrations before the onset of labor. Similar observations have been made in the monkey. In this species, neither glucose nor arginine stimulated fetal insulin release near term but glucagon evoked prompt insulin secretion. Late in gestation in the ovine fetus, epinephrine inhibited insulin release through a receptor pathway. In the anencephalic human fetus, the endocrine pancreas develops normally if maternal carbohydrate metabolism is not impaired, but beta cell hypertrophy and hyperplasia do not occur in the anencephalic fetus or in decapitated fetal rabbits exposed to chronic hyperglycemia. This lack of beta cell response to hyperglycemia may be the result of GH deficiency because GH stimulates insulin gene expression and may play a permissive role in beta cell hyperplasia and hypertyrophy. Pancreatic glucagon concentrations are relatively high in fetal plasma and increase progressively with fetal age. The fetal pancreatic glucagon content at midgestation is approximately 6 µg/g, compared with an adult level of 2 µg/g. As is true for insulin, the capacity for glucagon secretion is blunted in the fetus. Hypoglycemia does not suppress fetal plasma glucagon levels in rats, monkeys, or sheep, and acute hypoglycemia does not evoke glucagon secretion in the rat fetus. Amino acids, which are important secretagogues for insulin and glucagon in the adult, probably have little role in modulating insulin and glucagon secretion in the preterm fetus. However, infusion of alanine into women before hysterotomy fails to provoke fetal insulin secretion at midgestation or near term, and plasma insulin levels in the late human fetus are relatively unresponsive to high glucose concentrations before the onset of labor. Thus, the fetal pancreatic islet cells, although histologically mature and capable of hormone synthesis and hyperplasia, are relatively immature functionally at birth with regard to the capacity to secrete both insulin and glucagon. The rapid maturation of responsiveness to glucose in the neonatal period in both premature and mature infants suggests that this blunted state may be a secondary result of the relatively stable fetal serum glucose levels maintained by placental transfer of maternal glucose rather than a primary, temporally fixed maturation process. The blunted capacity for insulin and glucagon secretion has been related to a deficient capacity of the fetal pancreatic islet cells to generate cAMP or a rapid destruction of cAMP by phosphodiesterase, or both. Insulin and glucagon are normally not necessary for substrate metabolism in the fetus. Glucose is obtained by placental transfer through facilitated diffusion. The fetal respiratory quotient is approximately 1, which suggests that glucose is the primary energy substrate for the fetus. Other substrates such as amino acids and lactate may also be utilized in the human as in the sheep fetus. However, at least early in gestation, hepatic metabolism and substrate utilization appear to be independent of insulin and to be modulated in an autoregulatory fashion by glucose. In addition, the constant supply of glucose normally precludes the necessity for endogenous gluconeogenesis, and gluconeogenic enzyme activities are low in the fetal liver. Glycogen storage in the fetus is modulated by fetal glucocorticoids and probably by placental hPL. Fetal insulin plays a role near term, when insulin also has the capacity to increase fetal glucose uptake and lipogenesis. Insulin receptors are present on most fetal cells in higher numbers than on adult cells; moreover, hyperinsulinemia fails to down-regulate fetal insulin receptors. Fetal hepatic glucagon receptors, in contrast, are reduced in number, and fetal liver is relatively resistant to the glycemic effect of glucagon. These conditions tend to potentiate the fetal anabolic milieu during the period of rapid growth in the last trimester of gestation.
NEUTRALIZATION OF HORMONE ACTIONS IN THE FETUS

After the period of embryogenesis, the fetal milieu is programmed to optimize body growth and organ development through an array of generalized and specialized growth factors (see "Fetal Growth"). These function in a stable metabolic environment with substrate supply maintained by the placenta. The endocrine-metabolic systems characterizing the extraterine environment are programmed to maintain metabolic stability in a changing external environment with intermittent substrate provision. Hormonal systems in the fetus are programmed to maintain anabolism with minimal hormonal perturbation. Thus, production of catabolic and thermogenic hormones is limited and the effects of the hormones altering metabolic substrate supply and distribution pathways are muted.

Limitation of Hormone Secretion

The human fetal pancreas is functional during the second trimester, but secretion of insulin in response to glucose or pyruvate is minimal until the neonatal period. Glucagon secretion is also blunted, although fetal blood glucagon levels are relatively high. Fetal islet hyperplasia and increased insulin secretion occur in response to chronic hyperglycemia as in the infant of the diabetic mother, and insulin release can be stimulated by acute fetal infusions of leucine, arginine, or tolbutamide. Moreover, responsiveness of both insulin and glucagon secretion to glucose develops rapidly in the neonatal period. It is not clear whether the limited fetal islet cell responsiveness is due to the relatively stable fetal serum glucose levels or a temporally fixed maturation process (for details, see "Endocrine Pancreas: Insulin and Glucagon").
Production of Inactive Metabolites

Throughout the latter part of gestation, cortisol is metabolized in fetal tissues to inactive cortisone through an 11HSD. The placenta is permeable to steroid hormones including cortisol. During midgestation, placental 11HSD activity is low and cortisol is transferred to the fetus. Placental 11HSD activity increases during the second half of pregnancy under the control of placental estrogens, and enzyme activity near term is high. Thus, maternal-fetal cortisol transfer decreases progressively. In addition, although many adult tissues can convert cortisone to cortisol, conversion is limited during most of fetal life. Consequently, most of the cortisol that crosses the placenta or is produced by the fetus is inactivated to cortisone by the placenta or by fetal tissues.

<table>
<thead>
<tr>
<th>Active Hormone</th>
<th>Inactive Metabolites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cortisol</td>
<td>Cortisone</td>
</tr>
<tr>
<td>Thyroxine (T₄)</td>
<td>rT₃, T₁ S, rT₁ S</td>
</tr>
<tr>
<td>Triiodothyronine (T₃)</td>
<td>T₃ S, T₂</td>
</tr>
</tbody>
</table>

TABLE 21-4 -- Neutralization of Hormone Actions in the Fetus.

**Production of Inactive Metabolites**

**Delayed Expression or Neutralization of Receptors**

<table>
<thead>
<tr>
<th>Active Hormone</th>
<th>Receptor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth hormone (GH)</td>
<td>GHR</td>
</tr>
<tr>
<td>Thyroid hormone</td>
<td>TR, TR</td>
</tr>
<tr>
<td>Catecholamines</td>
<td>AR</td>
</tr>
<tr>
<td>Estrogens</td>
<td>ER</td>
</tr>
<tr>
<td>Glucagon</td>
<td>GR</td>
</tr>
</tbody>
</table>

**Limited Hormone Secretion**

<table>
<thead>
<tr>
<th>Active Hormone</th>
<th>Secretory Cell</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insulin</td>
<td>Islet cell</td>
</tr>
<tr>
<td>Glucagon</td>
<td>Islet cell</td>
</tr>
</tbody>
</table>

RT₃, reverse T₁, T₁ S, T₁ sulfate.

*See text for details.

Levels of cortisone in fetal plasma exceed those of cortisol by threefold to fourfold until after 30 weeks of gestation (see Fig. 21-5). Teleologically, this would help preserve the anabolic and growth-promoting milieu of the fetus and minimize premature maturational and parturitional effects of cortisol. After 30 weeks, the ratio of cortisol to cortisone in fetal tissues and plasma increases as a result of increased fetal secretion and decreased conversion of cortisol to cortisone within the placenta and fetal tissues. Cortisol has important maturational action on several fetal tissues near term (see later under "Transition to Extrauterine Life").

Fetal thyroid hormone metabolism is characterized by conversion of active thyroid hormones to inactive rT₃ and inactive sulfated iodothyronines and by limited receptor and postreceptor responsiveness to thyroid hormone in selected tissues. The placenta contains an iodothyronine inner-ring monodeiodinase that catalyzes conversion of maternal T₄ to rT₃. In addition, the fetal sheep liver and kidney, in contrast to the adult liver and kidney, manifest low levels of iodothyronine type I outer-ring monodeiodinase activity so that conversion of T₃ to active T₁ is limited and large amounts of inactive iodothyronine sulfon conjugates accumulate. As a consequence, plasma T₃ levels in the fetus remain low until the last few weeks of gestation (see Fig. 21-9). Selected fetal tissues (brain, brown adipose tissue) have active iodothyronine, type II, outer-ring monodeiodinase activities that contribute to local tissue T₃ concentrations; local T₃ is important in development, particularly in the hypothyroid fetus. Near term and in the neonatal period in the human fetus, the dramatic increase in plasma T₃ levels, and presumably T₃ production, heralds the onset of thyroid hormone actions on growth and development and on metabolism (Table 21-4).
Neutralization of Receptor Response

Selected ovine fetal tissues seem relatively unresponsive to thyroid hormones. Fetal ovine liver and kidney thermogenesis (as evidenced by oxygen consumption, Na⁺,K⁺-adenosine triphosphatase activity, and mitochondrial-glycerophosphate activity) is unresponsive to exogenous T₃ during the third trimester, and thyroid hormone responsiveness in a number of tissues (cardiac, hepatic, renal, and skin) develops only during the perinatal period. In Adrenergic receptor binding in heart

and lung of the ovine fetus is unresponsive to T₃ late in the third trimester but increases in response to T₃ in the neonatal period. In rodent species, in which development at birth is comparable to human fetal development at midgestation, pituitary GH concentrations become responsive to thyroid hormone only during the first weeks of extrauterine life. Mouse submandibular gland EGF and NGF levels become responsive to thyroid hormone during the second week of life, as do urine and kidney EGF concentrations and hepatic EGF receptor levels. Mouse skin EGF levels and EGF receptors are responsive during the first neonatal week.

Thus, despite the presence of nuclear T₃ receptors in significant concentrations in developing rat and sheep, many thyroid hormone actions in these species are delayed. The mechanism of this delayed thyroid hormone responsiveness is not clear; suppressor nuclear proteins may block gene expression in response to thyroid hormones during fetal development, and the levels of these suppressor proteins may determine the onset and degree of action of thyroid hormones during development.

The effect of the high circulating concentrations of GH in the fetus is also limited. Fetal somatic growth is only partially GH-dependent; indeed, the GH-deficient fetus has little or no growth retardation. The paucity of fetal GH effects is probably due to delayed maturation of GH receptors or postreceptor mechanisms. In animals such as sheep, hepatic GH receptor binding appears only during the neonatal period. Receptor deficiency may also be a factor in the limited PRL bioactivity in the fetus near term.

There is less information on fetal hormone responsiveness in other systems. Adrenergic receptor binding in heart and lung of the sheep fetus is relatively low near term and increases in the neonatal period in response to thyroid hormones. Moreover, premature lambs have an augmented plasma catecholamine surge at birth but have a relatively mild increase in plasma free fatty acid levels, which suggests reduced catecholamine responsiveness. The high levels of progesterone and estrogens in fetal blood also seem to have limited effects in the fetus. Progesterone receptors are present in low concentration in fetal guinea pig kidney, lung, and uterus at midgestation and increase progressively until term. Estrogen receptors appear in neonatal rat uterus, oviduct, cervix, and vagina during the first 10 days of extrauterine life, and both ER and ER mRNAs are present in human fetal tissues during the second trimester. The human neonate often manifests mild breast enlargement at birth, and vaginal estrogenation may be evident in female infants at birth. Estrogen effects otherwise appear limited (see Table 21-4).
FETAL GROWTH

Insulin-like Growth Factors

The somatomedins are involved in regulation of uterine and placental growth during pregnancy and in early embryonic and fetal development. IGF-I, EGF, and estrogens are mitogens for endometrial stromal cells, and the endometrial contents of IGF-I and IGF-I mRNA are high at implantation and during early embryogenesis in the sow. Uterine IGF-I and IGF-I mRNA levels decrease progressively with advancing gestation. Placental tissue also contains IGF-I and IGF-II mRNAs, significant concentrations of the respective proteins, and IGF-I receptors. Autocrine and paracrine roles for the IGFs in uterine and placental tissues are postulated. IGF-I and insulin are produced by embryonic tissues during the prepancreatic stage of mouse development, and both factors stimulate growth of embryonic mouse cells.

Immunoreactive IGF-I is present in most fetal tissues including brain, and fetal growth is regulated by the somatomedins. Transgenic mice with inactivating mutations of IGF-I, IGF-II, or IGF-I receptor have reduced birth weights, organ hypoplasia, and delayed bone development. Animals deficient in IGF-I receptor and some mice deficient in IGF-I die at birth; mice deficient in IGF-II survive and have near-normal postnatal growth, whereas surviving IGF-II-deficient animals have deficient postnatal growth. IGF-I and IGF-II mRNAs are localized in mesenchymal and fibroblast-like cells in interstitial and perivascular connective tissues and surrounding capsular tissues. In addition, immunoreactive IGF-I is produced by in vitro explant cultures of fetal mouse tissues, and fibroblasts cultured from fetal rat lung and skin synthesize both IGFs. These findings are consistent with a predominantly paracrine mode of action for these growth factors in the fetus.

Somatomedin-binding proteins are present as early as 5 weeks of gestation, and prenatally, as postnatally, somatomedins circulate associated with binding proteins. Thus, during fetal and postnatal life, plasma concentrations of somatomedins are relatively high compared with tissue concentrations. In the fetus, IGF-II levels are higher than those of IGF-I, in contrast to these levels in children and adults. Fetal levels of both peptides at term are 30% to 50% of adult levels. In most studies, cord blood IGF-I concentrations correlate with birth size. In spite of the fetal growth-enhancing effects of IGF-II, IGF-II levels are only weakly related to size at birth, largely because of the inhibiting effect of soluble IGF-II receptor (IGF2R). Soluble IGF2R is derived through proteolytic cleavage of the transmembrane region of the receptor in many tissues. Somatomedin receptors have been identified as early as 5 weeks of gestation and are widespread in fetal tissues. IGF-I stimulates glycogenesis in cultured fetal rat hepatocytes and induces formation of myotubes in cultured myoblasts. IGF-II is active in cultured muscle and neonatal rat astroglial cells. Insulin receptors are increased in fetal cells and are resistant to down-regulation; no similar data are available for the IGF-I receptor.

As discussed earlier, GH receptors are relatively deficient and receptors for hPL predominate in fetal tissues. Moreover, hPL stimulates IGF-I production and augments amino acid transport and DNA synthesis in human fetal fibroblasts and muscle cells. In addition, nutrition influences somatomedins in developing mammals. IGF-I levels fell in suckling rats deprived of milk, and IGF-I and IGF-II levels were reduced in fetuses of protein-starved pregnant rats and placently restricted sheep. The low IGF-II levels in the protein-starved rats were reversed by hPL. There is no evidence that thyroid hormones modulate GH or somatomedin levels in the mammalian fetus but, as mentioned earlier, glucocorticoids can inhibit fetal growth, presumably by inhibiting somatomedin action.

These data support the view that the somatomedins are important in embryonic and fetal growth and that in the fetus they are regulated, at least in part, by hPL and by nutritional substrate derived transplacentally. The high levels of IGF-II in fetal rat serum, the high levels of IGF-II mRNA in fetal tissues, and the presence of a truncated form of IGF-I in human fetal brain tissue suggest unique developmental actions of these peptides (Table 21-5).
Insulin

Insulin has been proposed to act as a fetal growth factor. Infants born to women with diabetes mellitus may have hyperinsulinemia associated with increased birth weight. Most of this increased weight is accounted for by body fat; there is little increase in body length, but some organomegaly may occur. Infants with hyperinsulinemia caused by nesidioblastosis or the Beckwith-Wiedemann syndrome may also have increased somatic growth in utero. Conversely, the human fetus with pancreatic agenesis is small and has decreased muscle bulk and little or no adipose tissue. Homozygosity for a null mutation of the insulin receptor gene in fetal mice led to early neonatal death with hyperglycemia and ketonemia; the pups, however, had a normal birth weight. These and other studies suggest that insulin may act as a fetal growth factor by promoting growth or hypertrophy of selected tissues. In clinical conditions associated with fetal hyperinsulinemia, insulin may act through insulin receptors (in adipose and liver tissues) or through type I IGF-I receptors. Insulin may also have a role in regulating IGF-I release.

<table>
<thead>
<tr>
<th>Tissues Affected</th>
<th>Growth Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neural tissue</td>
<td>BNF, NT3, NGF</td>
</tr>
<tr>
<td></td>
<td>Neuregulin, TGF-, EGF</td>
</tr>
<tr>
<td></td>
<td>IGF-I, IGF-II</td>
</tr>
<tr>
<td>Ectodermal derivatives</td>
<td>TGF-, EGF, PDGF</td>
</tr>
<tr>
<td></td>
<td>IGF-II, IGF-I</td>
</tr>
<tr>
<td>Mesodermal derivatives</td>
<td>IGF-II, IGF-I</td>
</tr>
<tr>
<td></td>
<td>TGF-, EGF, FGF family, Insulin, PDGF</td>
</tr>
<tr>
<td>Endodermal derivatives</td>
<td>IGF-II, IGF-I</td>
</tr>
<tr>
<td></td>
<td>TGF-, EGF, FGF</td>
</tr>
<tr>
<td>Hematopoietic tissues</td>
<td>HGF, EGF, PDGF</td>
</tr>
<tr>
<td></td>
<td>Colony-stimulating factors</td>
</tr>
<tr>
<td>Specialized tissues</td>
<td></td>
</tr>
<tr>
<td>Adipose tissue</td>
<td>Insulin</td>
</tr>
<tr>
<td>Skeletal tissue</td>
<td>PTHrP, IGF-I, IGF-II</td>
</tr>
<tr>
<td></td>
<td>Calcitonin</td>
</tr>
</tbody>
</table>

BNF, brain-derived neurotropic factor; EGF, epidermal growth factor; FGF, fibroblast growth factor; HGF, hematopoietic growth factor; IGF-II, insulin-like growth factor; NGF, nerve growth factor; NT3, neurotropin 3; PDGF, platelet-derived growth factor; PTHrP, parathyroid hormonelated protein; TGF-, transforming growth factor.

*This listing will be expanded and more detailed as knockout and other information develops.*
Epidermal Growth Factor-Transforming Growth Factor System

The EGF/TGF- system has been characterized in considerable detail. EGF is a 6-kd peptide product of a large 1207-amino-acid precursor molecule and acts through a 170-kd membrane receptor glycoprotein. This receptor, like the somatodemin receptor, has intrinsic tyrosine kinase activity, and tyrosine kinasesmediated autophosphorylation is a critical event in EGF signal transduction. TGF-, which has 35% amino acid homology with murine EGF and 44% homology with human EGF, also acts through the EGF receptor system. Several additional family members have been characterized, including amphiregulin, heparin-binding EGF, betacellulin, and neuregulins. Three additional receptors are referred to as ErbB2, ErbB3, and ErbB4 in animals; the human receptors are referred to as human EGF receptor (HER) 2, 3, and 4. All were characterized in malignant tissues, where they function as oncogenes, and all are widely distributed in normal mammalian tissues.

EGF is a potent mitogen for ectodermal and mesodermal cells in tissue and organ culture. These cells include keratinocytes derived from skin and conjunctival and pharyngeal tissues, corneal endothelial cells, vascular smooth muscle cells, chondrocytes, fibroblasts, liver cells, thyroid follicular cells, granulosa cells, and mammalian gland cells. In adult humans, EGF is present in highest concentrations in sweat glands, salivary glands, Brunner's (duodenal) glands, stomach, pancreas, bone marrow, prostate, kidney, and endocrine glands (pituitary, adrenal, and thyroid). High concentrations of EGF are also present in urine.

The roles of EGF and TGF- in humans are incompletely understood. In rodents and sheep, EGF provokes precocious eyelid opening and tooth eruption in neonatal animals; stimulates lung maturation; promotes palatal development in organ culture; stimulates gastrointestinal maturation; evokes secretion of pituitary hormones including GH, PRL, and corticotropin; and stimulates secretion of chorionic gonadotropin and placental lactogen by the placenta. Both EGF and TGF- compete for binding to the EGF receptor, and both factors accelerate eye opening and tooth eruption in the neonatal rodent, presumably through interaction with the same "EGF" receptor.

EGF and pre-pro-EGF mRNA are present in most tissues in the postnatal rodent and most adult mouse tissues, but mRNA levels are highest in salivary glands and kidneys. EGF and pre-pro-EGF mRNA levels are absent or low in the fetal rodent, and levels remain low in mouse tissues during the early neonatal period. Nonetheless, the EGF receptor knockout mouse exhibits epithelial immaturity and multilogn failure with early death. Tissue concentrations of both EGF and EGF mRNA increase in the mouse during the first 2 months of postnatal life; indeed, levels of EGF in the salivary glands increase several thousandfold between 3 weeks and 3 months of age. Mouse urinary levels increase 200-fold, and kidney concentrations increase 10-fold between 1 week and 2 months of age. EGF concentrations in mouse ocular tissues increase 100-fold during the first week of life. Liver EGF concentrations increase more slowly, as do serum levels, and there is a high degree of correlation between serum and liver EGF levels in the developing mouse. Thus, the production of EGF in the rodent is accelerated during the early neonatal period, and it is during this time that most hormone-stimulated growth and development occur.

There are few data on tissue TGF- concentrations in developing mammals. Immunoactive TGF- concentrations are measurable at relatively high levels in lung and brain tissues at 20 days of gestation in the rat and show minimal changes through day 50 postnatally. Liver, which also has high TGF- levels at 20 days of gestation, shows a progressive reduction in concentrations postnatally to nadir values in the young adult. Kidney tissue has low concentrations of TGF- in late gestation, and levels increase progressively during the first 2 months of postnatal life. Thus, the ontogenic pattern of TGF- is tissue specific; most late fetal tissues studied contain TGF-, and levels persist or increase in most tissues through the period of growth and development.

EGF plays an important role in pregnancy and fetal development. Maternal salivary gland and plasma EGF concentrations in the mouse increase fourfold to fivefold during pregnancy. Removal of the salivary glands prevents the increase in plasma EGF; moreover, salivary gland removal reduces the number of mice completing term pregnancy (by 50%), decreases the percentage of live pups, and decreases the crownrump length of fetuses delivered. Administration of EGF antiserum to pregnant mice without salivary glands further increases the abortion rate, whereas administration of EGF improves pregnancy outcome. These observations suggest an important role of EGF in pregnancy in the mouse. Because maternal EGF is too large a molecule to traverse the placental barrier, an effect on maternal metabolism and an effect on the placenta are likely. The placenta is richly endowed with EGF receptors, and placental tissue binds and degrades EGF to constituent amino acids.

EGF receptors are present in embryonal and fetal tissues, and EGF stimulates protein synthesis during the morula-blastocyst transition and in postimplantation mouse embryo tissue. In vitro, EGF stimulates differentiation of the inner cell mass during early embryonic development. However, EGF and EGF precursor mRNA levels are absent or present at low levels in selected fetal mouse tissues. Low levels are also present in submandibular gland and kidney during the early neonatal period. Fetal mouse and human tissues have high levels of TGF-, suggesting that this factor may be the ligand for the fetal EGF receptor. TGF- is produced by the maternal decidua during the first half of gestation in rodents, and proTGF- mRNA is present in decidua. Decidual proTGF- mRNA levels peak at 8 days of gestation (term = 21 days) and decline through day 15, when the decidua is being absorbed. EGF receptors are present in decidua, and TGF- may stimulate proliferation of decidual tissue and enhance decidual PRL production.

Inactivation of the gene encoding the EGF receptor in mice led to fetal or neonatal death of homozygous fetuses. The receptor-deficient animals manifested impaired epithelial development in several organs, including skin, lungs, and gastrointestinal tract. Further evidence for a role of EGF in early mammalian development has come from studies of the effect of the administration of EGF antiserum to neonatal mice. EGF antiserum delayed eye and ear opening, delayed tooth eruption, accelerated hair growth, and reduced weight gain during the first 30 days of life.

The factors that control EGF and TGF- production are incompletely understood. The increases in EGF concentration in tissues, blood, and urine of the neonatal rodent correlate with and may be conditioned by the increases in thyroidal and gonadal hormone levels. EGF concentrations in the mouse submandibular gland are increased by thyroid hormones and testosterone. Thyroid hormones increase EGF concentrations in skin, ocular tissue, kidney, and urine in the developing mouse and up-regulate EGF gene expression and the production of pro-EGF in rat kidney; thyroid hormones also increase EGF receptor levels in developing mouse skin and liver. Urinary EGF excretion is highly correlated with serum thyroid hormone concentrations in premature and term human infants. GH increases urinary EGF concentrations in the neonatal mouse, and estrogens increase EGF and EGF mRNA levels in mouse uterus. Testosterone stimulates EGF and EGF mRNA levels in submandibular gland and increases EGF receptor levels in prostatic tissue. Thus, EGF may mediate growth and developmental actions of a variety of hormones in selected tissues (see Table 21-5). There is little information about the regulation of TGF- production postnatally or prenatally. Amphiregulin binds to and stimulates EGF receptor and HER2 in human epithelial cells and has been localized to breast and colonic epithelium.

Considerable evidence suggests a role for the EGF family of growth factors in mammalian central nervous system development. EGF, TGF-, neuregulins, and the EGF receptors are widely distributed in the nervous system. EGF promotes proliferation of astroglial cells, acts as an astroglial differentiation factor, and enhances survival and outgrowth of selected neuronal cells. Transgenic mice with a deficiency of neuregulin, ErbB2, ErbB3, or ErbB4 die in utero with cardiac anomalies and developmental anomalies of the hindbrain, midbrain, and ventral forebrain (see Table 21-5).
Nerve Growth Factor

NGF is a 13-kd protein that is present at high concentrations in mouse salivary gland and at low concentrations in many adult tissues. It is also produced by human placental tissue. It is the original member of an expanding family of neurotropic growth factors that now include brain-derived neurotropic factor, neurotropin 3, and two less well characterized factors and involving two receptors, NGF and NGF2 (or Trk). NGF binds to high-affinity plasma membrane receptors and is internalized and transported to subcellular organelles, including the nucleus, in neurons of the peripheral nervous system. It promotes neurite outgrowth and enhances tyrosine hydroxylase and dopamine-hydroxylase activities in developing sympathetic neurons. NGF acts on undifferentiated sympathetic cell precursors to elicit both hyperplastic and hypertrophic effects and plays a permissive role in stimulating the development of immature autonomic neurons along either a sympathetic or a cholinergic pathway.

The injection of NGF in neonatal mice causes a marked increase in the volume of the superior cervical ganglia and increases in RNA polymerase, ornithine decarboxylase, and tyrosine hydroxylase activities. This growth factor also increases the nerve supply of body organs. Likewise, injection of NGF antiserum during early neonatal life results in a decrease in the size of the superior cervical ganglia, reduction in tyrosine hydroxylase activity, and permanent sympatheticotomy. Maternal NGF autoantibodies in rats and rabbits impair autonomic nervous system development in utero. This impairment affects sympathetic and dorsal root ganglia and autonomic innervation of peripheral organs. NGF is produced by neonatal mouse astroglial cells in tissue culture, is present in developing mouse brain tissue, and, with brain-derived neurotropic factor and neurotropin 3, plays an important role in brain development.

Thyroid hormones and testosterone modulate postnatal NGF levels in the submandibular gland of the mouse. Thyroid hormones increase NGF, neurotropin 3, and brain-derived neurotropic factor mRNA levels in adult rat brain.
Other Factors

Additional growth factors are involved in fetal growth and development, including hematopoietic growth factors, platelet-derived growth factors, fibroblast growth factors, vascular endothelial growth factor, and members of the TGF- family. Hematopoietic growth factors are also active in the fetus during development; erythropoietin in the fetal sheep is produced by the liver rather than the kidney and erythropoietin gene expression in fetal sheep is regulated by glucocorticoids. A switch to kidney production occurs after parturition. Postnatally, thyroid hormones, testosterone, and hypoxia modulate erythropoietin production.

Platelet-derived growth factor (PDGF) represents a family of homodimers and heterodimers of PDGF-A and PDGF-B chains derived from two gene loci. Two PDGF receptors have been characterized, PDGF and PDGF. The genes for PDGF and its receptors are expressed in many tissues. PDGF-A gene inactivation in mice led to defects in lung, skin, intestine, testes, and brain resulting in early postnatal death. PDGF-B gene inactivation led to microvessel disruption and leakage with hemorrhage and edema and intrauterine death.

The fibroblast growth factor (FGF) family of heparin-binding growth factors now includes 17 members with diverse effects on development, angiogenesis, wound healing, and other biologic systems. These effects are mediated by ligand-activated tyrosine protein kinase receptors (FGFRs) transcribed from four related genes. Several receptor isoforms are products of alternative RNA splicing.

Targeted disruptions of FGF and FGFR genes in mice have defined critical roles in development. FGF3-deficient mice show tail and inner ear defects. Knockout of the FGF4 gene is lethal, leading to early death. Knockout of the FGFR1 gene also leads to early fetal death. FGF10 knockout mice die at birth because of pulmonary agenesis. FGF4, FGF8, FGF9, FGF10, or FGF17 deficiency is associated with limb deformities. FGF8 deficiency leads to abnormal left-right axis patterning. In mice, FGFR3 knockout results in chondrocyte hypertrophy and increased bone length.

In humans, a variety of gain-of-function FGFR mutations are associated with chondrodysplasias and craniosynostosis syndromes. FGF, like EGF, stimulates the production of hCG from a choriocarcinoma cell line. These observations and the fact that the placenta contains FGF, NGF, TGF-, TGF-, IGF-I, and IGF-II suggest that the placenta may play an important role in modulating fetal growth.
TRANSITION TO EXTRAUTERINE LIFE

The transition to extrauterine life involves abrupt delivery from the protected intrauterine environment and succor by the placenta to the relatively hostile extrauterine environment. The neonate must initiate air breathing and defend against hypothermia, hypoglycemia, and hypocalcemia as the placental supply of energy and nutritional substrate is removed. Both the adrenal cortex and the autonomic nervous system, including the para-aortic chromaffin system, are essential for extrauterine adaptation. Longer term transition requires adaptation to an environment of intermittent nutrient supply and transient substrate deficiency and requires maturation of the secretory control mechanisms for the PTH-calcitonin system and the endocrine pancreas.

Cortisol Surge

In most mammals, a cortisol surge occurs near term and is mediated by increased cortisol production by the fetal adrenal and a decreased rate of conversion of cortisol to cortisone. Pepe and Albrecht have proposed that the preterm fetal cortisol surge is due to the progressive stimulation by estrogens of placental 11HSD activity and the subsequent increase in placental conversion of cortisol to cortisone. The resulting decrease in maternal-to-fetal cortisol transfer results in stimulation of fetal CRH and corticotropin secretion through the negative-feedback control loop. The concomitant estrogen-stimulated increase in 11HSD activity in fetal tissues potentiates the relative fetal cortisol deficiency and the CRH-corticotropin response. Placental CRH may also potentiate the fetal adrenal activation.

The cortisol surge augments surfactant synthesis in lung tissue; increases lung liquid reabsorption; increases adrenomedullary phenylethanolamine N-methyltransferase activity, which in turn increases methylation of NE to epinephrine; increases hepatic iodothyronine outer-ring monodeiodinase activity and hence increases conversion of T4 to T3; decreases sensitivity of the ductus arteriosus to prostaglandins, which facilitates ductus closure; induces maturation of several enzymes and transport processes of the small intestine; and stimulates maturation of hepatic enzymes (Fig. 21-12). In some cases, these events involve increased synthesis of specific proteins or enzymes. In other instances, such as the action on the ductus arteriosus, the mechanism remains obscure.

Secondary effects of cortisol also promote extrauterine adaptations. The increased T3 levels stimulate -adrenergic receptor binding, potentiate surfactant synthesis in lung tissue, and increase the sensitivity of brown adipose tissue to NE. The significance of prenatal cortisol is demonstrated by the effects of gene-targeted CRH or glucocorticoid receptor deficiency in mice; the progeny of homozygous CRH-deficient or glucocorticoid receptor-deficient animals die in the first 12 hours with lung dysplasia and surfactant deficiency.
Catecholamine Surge

Parurition also evokes a dramatic catecholamine surge in the newborn, resulting in extraordinarily high levels of NE, epinephrine, and dopamine in cord blood. As indicated earlier, plasma NE concentrations exceed epinephrine levels because of peripheral and adrenomedullary and para-aortic catecholamine release. Cord blood NE levels of 15 nmol/L (2500 pg/mL) and epinephrine levels of 2 nmol/L (370 pg/mL) are common after spontaneous delivery of term infants. Levels of 25 nmol/L (4200 pg/mL) of NE and 35 nmol/L (640 pg/mL) of epinephrine are common in cord blood of premature infants. These changes evoke critical cardiovascular adaptations, including increased blood pressure and increased cardiac inotropic effects; increased glucagon secretion; decreased insulin secretion; increased thermogenesis in brown adipose tissue and increased plasma free fatty acid levels; and pulmonary adaptation, including mobilization of pulmonary fluid and increased surfactant release.
Thermogenesis in Neonatal Brown Adipose Tissue

Brown adipose tissue is the major site of thermogenesis in the newborn and is especially prominent in the mammalian fetus. The largest accumulations of brown adipose tissue envelop the kidneys and adrenal glands, and smaller amounts surround the blood vessels of the mediastinum and neck. The mass of brown adipose tissue peaks at the time of birth and gradually decreases during the early weeks of life. Surgical removal of this tissue leads to neonatal hypothermia. NE, through -adrenergic receptors, stimulates thermogenesis by brown adipose tissue, and optimal responsiveness of this tissue to NE is dependent on thyroid hormone.

Brown adipose tissue is rich in mitochondria containing a unique 32-kd protein (thermogenin) that uncouples oxidation and phosphorylation of adenosine diphosphate, reduces adenosine triphosphate production, and consequently enhances thermogenesis. Thermogenin is T3-dependent, and brown adipose tissue contains a 5'-monoiodothyronine deiodinase that synthesizes T3 locally from T4.

Full maturation of catecholamine-stimulated cellular respiration in brown adipose tissue occurs before delivery in the ovine fetus and requires thyroid hormone. Fetal thyroidectomy in this species leads to marked hypothermia, with low plasma free fatty acid levels and increased plasma epinephrine concentrations. In vitro, basal brown adipose tissue thermogenesis and NE-stimulated and dibutyryl cAMP-stimulated thermogenesis are decreased by fetal thyroidectomy.

The rapid onset of thermogenesis in brown adipose tissue is essential for survival in newborn infants. Catecholamine release is the stimulus for brown adipose tissue thermogenesis in the early neonatal period, and responsiveness to catecholamines is markedly increased by cutting of the umbilical cord. Fetal hypoxia and placental inhibitors, including prostaglandin E2 and adenosine, appear to inhibit brown adipose tissue thermogenesis in utero. Cord cutting, neonatal cooling, catecholamine stimulation, and augmented conversion of T4 to T3 in brown adipose tissue in the neonatal period are the essential features that mediate and condition newborn thermogenesis.
Calcium Homeostasis

The neonate must adjust rapidly from a high-calcium environment regulated by PTHrP and calcitonin to a low-calcium environment that requires regulation by PTH and vitamin D. With removal of the placenta in term infants, plasma total calcium concentration falls and reaches a nadir of approximately 2.3 mmol/L (9 mg/dL), and the ionized calcium concentration reaches a low level of about 1.2 mmol/L (4.8 mg/dL) by 24 hours of life. Plasma PTH levels are relatively low in the neonatal period and are minimally responsive to hypocalcemia during the first 2 to 3 days of life.

Calcitonin concentrations are high in cord blood (2000 ng/L), increase further during the early neonatal period, and remain high for several days after birth. The relatively obtunded PTH response and the high calcitonin levels lead to a 2- to 3-day period of transient neonatal hypocalcemia. Inhibition of calcitonin secretion and stimulation of PTH secretion gradually result in increased serum calcium levels in the neonate. The disappearance of PTHrP in the neonatal lamb is approximately coincident with the time of restoration of calcium levels to the adult range. The mechanism of transition from PTHrP to PTH secretion by the neonatal parathyroid glands is not clear.

Calcium homeostasis is also affected in the human newborn period by the low level of glomerular filtration that persists for several days. In addition, renal responsiveness to PTH is reduced in the first few days of life. These factors limit phosphate excretion and predispose the neonate to hyperphosphatemia, particularly if the diet includes high-phosphate milk such as unmodified cow's milk. Premature infants tend to have lower PTH and higher calcitonin levels and more immature kidney function; in these infants, neonatal hypocalcemia may be more marked and prolonged and the incidence of symptomatic hypocalcemia is higher. Birth asphyxia also predisposes the neonate to hypocalcemia. Infants born to mothers with hypercalcemia related to hyperparathyroidism have a high incidence of symptomatic hypocalcemia. These infants have a more marked suppression of parathyroid function and a longer period of transient hypoparathyroidism in the neonatal period. PTH secretion and calcium homeostasis usually return to normal in 1 to 2 weeks in full-term infants and within 2 to 3 weeks in the small premature infant.
Glucose Homeostasis

The abrupt withdrawal of the placental glucose supply leads to a prompt fall in plasma glucose in the term neonate. The low glucose and high catecholamine levels stimulate glucagon secretion, and the plasma glucagon level peaks within 2 hours after birth. Plasma insulin levels are low at birth and tend to fall further with hypoglycemia. The early glucagon response is short-lived; however, levels remain at about 100 ng/L for the first 12 to 24 hours, and the glucagon/insulin ratio is high enough to stabilize glucose levels in the range 2.8 to 4 mmol/L (50 to 70 mg/dL) during this period.

The early glucagon and catecholamine surges deplete hepatic glycogen stores so that the return of plasma glucose levels to normal after 12 to 18 hours requires maturation of hepatic gluconeogenesis under the stimulus of a high plasma glucagon/insulin ratio. Glucagon secretion gradually increases during the early hours after birth, especially with protein feeding, which stimulates gut glucagon release and pancreatic glucagon secretion. Premature infants have more severe and prolonged hypoglycemia because of reduced glycogen stores and impaired hepatic gluconeogenesis. Infants born to diabetic mothers have more severe neonatal hypoglycemia because of relative hyperinsulinism. In the healthy term infant, glucose homeostasis is achieved within 5 to 7 days of life; in premature infants, 1 to 2 weeks may be required.
Other Hormonal Adaptations

Delivery of the placenta results in decreases in fetal blood levels of estrogens, progesterone, hCG, and hPL. The fall in estrogen levels presumably removes the major stimulus to fetal PRL release, and PRL levels decrease within several weeks. The relatively delayed fall may be due to lactotrope hyperplasia in the fetal pituitary or to delayed maturation of hypothalamic dopamine secretion. The gradual fall of hGH levels during the early weeks of life is due to delayed maturation of hypothalamic-pituitary and feedback control of hGH release.\[95\] In the neonatal primate are concomitant decreases in plasma GH levels and GH responsiveness to exogenous GHRH.\[293\] The mechanisms remain unclear. Changes in secretion or in pituitary sensitivity to GHRH or somatostatin, or both, may be involved. Somatomedin levels fall to infantile values within a few days, presumably because of the removal of placental hPL and placental somatomedin production (see Fig. 21-5).\[295\]

In male infants (see Fig. 21-11), after a transient fall in testosterone levels as the hCG stimulus abates, pituitary LH secretion rebounds and there is a secondary surge of plasma testosterone that persists at significant levels for several weeks.\[95\] This surge is mediated by hypothalamic LHRH, and blockade of neonatal activation of the pituitary-testicular axis with an LHRH agonist in neonatal monkeys ablated the neonatal increments in LH and testosterone.\[295\] Such a blockade also resulted in subnormal increments in plasma LH and testosterone levels and subnormal testicular enlargement at puberty in these animals, which suggests that neonatal LHRH release with pituitary-testicular activation may be critical for normal sexual maturation of male primates.\[295\] In females, a transient, secondary surge in FSH may transiently elevate estrogen levels.

Delivery results in a reversal of the high fetal cortisol/cortisol ratio, and plasma cortisol concentrations are higher in the neonate despite relatively lower plasma corticotropin concentrations (see Fig. 21-7). Presumably, this increase is due to decreased inhibition of adrenal 3HSD by estrogen and perhaps to removal of a placental CRH action on fetal pituitary corticotropin release. Plasma DHEAS and DHEA levels fall as the fetal adrenal atrophies.

The increase in serum thyrotropin levels during the early minutes after birth is due to cooling of the neonate in the extrauterine environment.\[136\] The thyrotropin surge peaks at 30 minutes at a concentration of about 70 mU/L (see Fig. 21-9). This peak evokes increased secretion of T4 and T3 by the thyroid gland. In addition, increased conversion of T4 to T3 by liver and other tissues maintains the T3 level in the extrauterine range of 1.6 to 3.4 nmol/L (105 to 220 ng/dL). The reequilibration of thyrotropin levels to the normal extrauterine range is probably due to the readjustment of prevailing serum T3 levels and to maturation of feedback control of thyrotropin by thyroid hormones during the early weeks of life.\[147\] Production of rT3 by fetal and neonatal tissues abates by 3 to 4 weeks of age, at which time serum rT3 reaches adult levels.
IMPRINTING OF DEVELOPING ENDOCRINE SYSTEMS

Data for several mammalian species indicate that hormonal imprinting or programming occurs during a critical fetal or perinatal period of development. In the female rodent, transient neonatal androgen administration masculinized the pattern of hypothalamic control of LHRH secretion and pituitary gonadotropin secretion. Masculinized behavior and adult sexual activity, permanently altered the pattern of GH secretion, increased longitudinal bone growth and body weight, and masculinized the pattern of hepatic steroid metabolism. Prenatal androgens program the timing of neuroendocrine puberty in sheep; the higher the dose of prenatal testosterone, the earlier the initiation of the pubertal LH rise. Estrogen administration to pregnant rats during the last third of gestation produced cryptorchid male offspring and may permanently suppress spermatogenesis in adult males.

Perinatal estrogen administration to the developing female rodent led to growth retardation, delayed puberty, decreased adult pituitary weight, decreased pituitary TRH concentrations, low serum thyrotropin levels, and decreased thyrotropin responsiveness to propylthiouracil challenge. Chronic hyperprolactinemia also occurred, presumably secondary to the low-level continuous estrogen secretion in these anovulatory animals. Prenatal or neonatal diethylstilbestrol in the hamster predisposed to hypoplastic and neoplastic uterine lesions. In these exposed animals uterine levels of c-Jun, c-Fos, c-Myc, BAX, and Bcl-x were markedly increased in luminal and glandular epithelial cells, whereas Bcl-2 levels were decreased. Glucocorticoids are essential for many aspects of normal brain development. Rodents and primates exposed to excess levels of glucocorticoid in utero exhibited hyperactivity of the hypothalamic-pituitary-adrenal axis. Behavioral alterations after glucocorticoid exposure have also been documented in rodent species.

Transient levothyroxine administration to neonatal rodents led to growth retardation, delayed puberty, decreased adult pituitary weight, decreased pituitary TRH concentrations, low serum thyrotropin levels, and decreased thyrotropin responsiveness to propylthiouracil challenge. Adult adrenal function and EGF metabolism were also altered. Administration of insulin or alloxan to neonatal rats produced permanent alteration of glucose tolerance, and a single dose of vasopressin to the neonatal rat permanently enhanced the adult response to vasopressin. Neonatal catecholamine administration altered the response of adult rat vascular tissue to NE. Fetal exposure to high maternal glucocorticoid levels in the rat inhibited fetal growth and led to subsequent hypertension in the offspring.

There is less information about hormonal imprinting in primates and humans. Blockade of LHRH in neonatal monkeys with an LHRH agonist resulted in obtundation of gonadal function and normal with treatment. Congenital hypothyroidism in human infants may be associated with alteration of the set-point for feedback control of thyrotropin release so that serum thyrotropin levels remain inappropriately elevated after return of serum T4 levels to normal with treatment.

A growing literature now supports the concept of fetal programming of cardiovascular disease. It is now clear that intrauterine growth retardation with low birth weight for gestational age is associated with an increased risk of hypertension, insulin resistance and diabetes, and coronary heart disease. The mechanisms are not yet clear, but permanent prenatal changes of fetal endocrine and growth factor homeostasis seem to be involved. These include IGF metabolism, insulin secretion and action, kidney growth and structure, the renin-angiotensin system, and the hypothalamic-pituitary-adrenal axis. There are also permanent changes in the brain. The mechanisms for imprinting remain obscure. Neonatal administration of testosterone or glucocorticoid can have permanent effects on brain structure.

The mechanisms for imprinting remain obscure. Neonatal administration of testosterone or glucocorticoid can have permanent effects on brain structure. The effects in some instances may be transmitted to subsequent generations. A functional overlap of hormone-mediated imprinting may also occur; the administration of both thyrotropin and FSH to the rodent neonate altered the adult response to thyrotropin, and neonatal OT or vasopressin exposure can alter the adult response to vasopressin. Hormonal imprinting is also demonstrable in cell lines and in unicellular organisms, in which a single exposure to a hormone can produce persistent alteration of the hormonal response characteristics.

These observations suggest that hormone imprinting or programming may be due to plasticity of the hormone system during a critical period of maturation. Gene expression and nuclear receptors or plasma membrane receptors, or both, may be involved. Prohormone processing may also be involved in newborn rat intermediate pituitary lobe cells, in vitro treatment with dexamethasone decreased production of -MSH and increased production of corticotropin-related peptides. Whatever the mechanisms, the developing endocrine systems have significant plasticity, and the maturation of endocrine control systems can be influenced by alterations in the prevailing hormone concentrations.
Summary

The foregoing review summarizes current understanding of the intrauterine endocrine milieu and highlights progress in this challenging frontier of medicine. This progress has set the stage for fetal endocrine disease diagnosis, therapy of fetal endocrine and metabolic disorders, management of disorders of fetal growth, and diagnosis and management of perinatal or neonatal endocrine dysfunction. In addition, understanding of developmental endocrinology is increasingly relevant to management strategies for premature infants and infants and children with fetal growth retardation and for our understanding of the pathogenesis of adult endocrine and metabolic diseases.

We are now entering the era of direct access to and management of the intrauterine environment with provision of medical and surgical fetal therapy, entailing both potential advantages and risks. With expansion of the application and scope of amniotic fluid and fetal cell sampling and the advent of fetal visualization and intrauterine fetal blood sampling, direct access for fetal diagnosis is now possible. Women are treated with glucocorticoids to stimulate fetal lung maturation. Intrauterine diagnosis and treatment of fetal adrenal and thyroid disorders have become the standard of care. Intravenous nutritional supplementation of fetal sheep can prevent some forms of growth retardation, and chronic fetal therapy through indwelling pumps is feasible in animal fetuses. These approaches, coupled with increasing availability of synthetic hormones and growth factor agonists and antagonists, facilitate direct fetal endocrine therapy. In addition, intrauterine stem cell transplantation has been successful in the correction of congenital hematologic disease; the fetus in early gestation is a favorable recipient of cellular therapy, and fetal cell transplantation may be applicable to therapy for selected endocrine and metabolic diseases.
References


3. Albrecht ED, Pepe GJ. Central integrative role of oestrogen in modulating the communication between the placenta and fetus that results in primate fetal-placental development. Placenta 1999; 20:129139.


Eng PHK, Cardona GR, Fang SL, et al. Escape from the acute Wolff-Chaikoff effect is associated with a decrease in thyroid sodium/iodide symporter messenger ribonucleic acid and protein.


References


839


Polk DH. Thyroid hormone effects on neonatal thermogenesis. Semin Perinatol 1988; 12:151156.


Chapter 22 - Disorders of Sex Differentiation

Melvin M. Grumbach
Ieuan A. Hughes
Felix A. Conte

Under most circumstances the distinction between male and female is considered absolute, and these terms are often used to epitomize opposites. Usually, the components of an individual's sexual makeup are indeed dominantly of one gender and conform to the chromosomal pattern established in the zygote at the time of fertilization. Most sexual characteristics, however, emerge from identical bipotential precursors in the embryo, and a spectrum of differentiation is possible at each level of sexual organization.

The remarkable accumulation of knowledge over the past five decades and new and continuing insights in the field of sex determination and differentiation represent major landmarks in biomedical science. No aspect of prenatal development is better understood. Advances in embryology, steroid biochemistry, molecular and cell biology, cytogenetics and genetics, endocrinology, immunology and transplantation biology, and the behavioral sciences all have contributed to the understanding of sexual anomalies in humans and to the improved clinical management of these disorders. Major contributions to this understanding have stemmed from studies of patients with abnormalities of sex differentiation. Noteworthy are the advances that have resulted from application of the techniques of molecular genetics.

Failure at any of the sequential stages of sexual development, whether the cause is genetic or environmental, can have a profound effect on the phenotype and lead to complete sex reversal, various degrees of ambisexual development, or less overt abnormalities in sexual function that first become apparent after sexual maturity. For general works on sex determination and differentiation see references. The diversity of sex-determining mechanisms in multicellular organisms is extraordinary and varies from a repertoire of genetic sex-determining models, including the X chromosome/autosome ratio in fruit flies and worms and the evolution of the highly conserved, dominant, testis-determining, Y-linked gene in mammals, to the temperature-dependent sex determination model in many egg-laying reptiles. Of interest is the commonality of sex-specific regulation of transcription factors and cell signaling molecules.
NORMAL SEX DETERMINATION AND SEX DIFFERENTIATION

Sex determination and differentiation are sequential processes that involve successive establishment of chromosomal (and genetic) sex in the zygote at the moment of conception, determination of gonadal (primary) sex by the genetic sex, and regulation by the gonadal sex of the differentiation of the genital apparatus and, hence, the phenotypic sex. At puberty, the development of sex-specific secondary sexual characteristics reinforces and provides more visible phenotypic manifestations of this sexual dimorphism. Sex determination is concerned with control of the development of the primary or gonadal sex, and sex differentiation encompasses the events subsequent to gonadal organogenesis. These processes are regulated by a myriad of different genes located on sex chromosomes and autosomes that act through a variety of mechanisms, including organizing factors, gonadal steroid and peptide hormones, and tissue receptors. Early embryos of both sexes possess indifferent, common primordia that have an inherent tendency to feminize unless there is active interference by masculinizing factors. The indifferent embryonic gonad develops into an ovary unless it is diverted by a testis-organizing gene on the Y chromosome; female differentiation of the somatic sex structures (the internal and external genital tract) is independent of gonadal hormones and takes place in the absence of fetal testes.

### TABLE 22-1 – Ontogeny of Sexual Characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>How Identified</th>
<th>Origin</th>
<th>Factors Determining Differentiation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chromosomal sex</td>
<td>Karyotype analysis</td>
<td>Sex chromosomes of parental germ cell</td>
<td>Normal: chromosomal composition of sperm</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Abnormal: Nondisjunction during meiotic divisions of parental germ cells</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Nondisjunction or anaphase lag in early mitotic divisions of zygote</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Structural errors caused by chromosome breakage</td>
</tr>
<tr>
<td>X chromosome</td>
<td>Buccal smear; neutrophil spreads; smears or sections of other peripheral tissues</td>
<td>Late-replicating (heterochromatinized) X chromosome</td>
<td>Partial inactivation and heterochromatin formation of all X chromosomes in excess of one</td>
</tr>
<tr>
<td>Y body</td>
<td>Same as for X chromosome; also seen in sperm</td>
<td>Y chromosome</td>
<td>Distal heterochromatic segment of long arm of Y</td>
</tr>
<tr>
<td>Gonadal sex</td>
<td>Histologic appearance</td>
<td>Testis</td>
<td>Testis: SRY gene on the Y chromosome just proximal to the pseudoautosomal boundary; upstream and downstream autosomal genes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Ovary</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Ovary: ovary-determining genes on X chromosome(s) and autosomes; germ cells</td>
</tr>
<tr>
<td>Genital ducts</td>
<td>Pelvic examination and imaging; pelvic exploration</td>
<td>Müllerian and Wolffian ducts</td>
<td>Intrinsic tendency to feminize; müllerian involution requires antimüllerian hormone from fetal Sertoli cells, and the antimüllerian hormone receptor testosterone stimulates male duct development</td>
</tr>
<tr>
<td>External genitalia</td>
<td>Inspection; investigation of urogenital sinus by urethroscopy and/or radiographic contrast study</td>
<td>Genital tubercl, urethral folds, labioscrotal folds, and urogenital sinus</td>
<td>Intrinsic tendency to feminize; masculinization requires androgenic stimulation before 12th fetal week</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Normal male: testosterone from fetal testes converted to dihydrotestosterone at end organ</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Virilized female: adrenal hyperplasia (21- and 11-hydroxylase deficiency); maternal androgen; aromatase deficiency</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Incompletely differentiated male: insufficient testosterone secretion by fetal testes; 5-reductase deficiency; end-organ androgen resistance</td>
</tr>
<tr>
<td>Hormonal sex</td>
<td>Secondary sexual characteristics</td>
<td>Hypothalamus and neural centers: LHRH</td>
<td>Hypothalamus and other neural centers: LHRH</td>
</tr>
<tr>
<td></td>
<td>Male: sexual hair pattern; voice; muscularity; phallic size</td>
<td>Pituitary gonadotropin</td>
<td>Pituitary: gonadotropin release governed by pulsatile secretion of hypothalamic LHRR and circulating levels of gonadal steroids and inhibin</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Female: breast development; rounding of contours; growth of reproductive tract; menstruation; ovulation</td>
<td>Secretory cells of testes, ovaries, and adrenals</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Male: testosterone secretion from testes; gonadotropin release</td>
<td>Gonads: differentiation of secretory cells and biosynthetic enzymes; stimulation by pituitary gonadotropins</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Female: cyclic secretion of gonadotropins, estrogen, and progesterone</td>
<td>Hormone expression may be modified by end-organ sensitivity</td>
<td></td>
</tr>
<tr>
<td>Gender identity</td>
<td>Identification of self as either male or female</td>
<td>Appearance of external genitalia, environmental factors, prenatal androgens, genes, pubertal sex characteristics</td>
<td>Prenatal exposure to androgens (testosterone) from the testes can contribute to male gender identity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Psychological environment during early years of critical importance in establishing gender identity:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Attitudes of parents Interactions of both sexes Conformity of genitalia and secondary sexual characteristics at puberty to assigned sex</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Genes, including Y-linked and X-linked</td>
</tr>
</tbody>
</table>
whether ovaries are present or not. Thus, the sexual dimorphism in phenotype in placental mammals is mediated by the fetal testis and its dual hormonal secretions and not by the ovary (Table 22-1). When luteal secretions are present, male differentiation takes place despite a fetal environment in which the concentration of circulating estrogens and progesterone is high.

Abnormalities of sexual development can be classified into two broad categories: (1) disorders of sex determination, which most often are caused by sex chromosome or gene abnormalities that affect gonadogenesis, and (2) disorders of sex differentiation, which usually are caused by a genetic defect and less often by adverse factors in the intrauterine environment. Before the genetic control of sex determination and gonadogenesis are discussed, a description of aspects of cytogenetics that are important to understanding abnormalities of sex determination is presented.

**Chromosomal Sex and X and Y Chromosomes**

A systematized array of metaphase chromosomes from a single cell is known as a karyotype. The meaning of this term is usually extended to imply that the chromosomal pattern in that cell typifies all the diploid cells of that individual or even of that species, although this is by no means always true. The 22 autosomes and two sex chromosomes (i.e., two X chromosomes or an X and a Y) are arranged and serially numbered according to size. The X chromosomes resemble the larger autosomes in the medium-sized group with submedian centromeres (group 6 to 12). The Y chromosome resembles the short acrocentric autosomes in chromosomes 21 and 22 (Fig. 22-1). Each of the pairs of chromosomes can be identified with chromosome banding and painting techniques. The pattern of DNA replication in human chromosomes can be studied by pulse labeling cell cultures with tritiated thymidine and preparing autoradiographs of the chromosomal spreads or by using the bromodeoxyuridine dye technique. One of the two X chromosomes in the female replicates late during DNA synthesis in the cell cycle, and this X chromosome gives rise to the distinctive X chromatin or Barr body seen in female somatic cells (see later discussion).

Chromosome banding techniques differentially stain chromosome segments. Caspersson and associates introduced a fluorescent staining method referred to as the Q staining method, in which substances such as quinacrine mustard or quinacrine hydrochloride are used to give a distinctive banding pattern (Q bands) for each chromosome (Fig. 22-2). The distal portion of the Y chromosome fluoresces intensely. Pardue and Galt subsequently reported a Giemsa staining technique that preferentially stains the centromeric regions of the chromosome; the areas of constitutive (centromeric) heterochromatin are known as C bands. The Giemsa staining technique has been modified by use of a multitude of pretreatment procedures on fixed metaphase chromosomes (e.g., hypertonic saline, sodium hydroxide, variation of pH, temperature, cation concentration, proteolytic enzymes), which produce Giemsa-stained bands that are identical (with minor exceptions) to the Q bands described by Caspersson and associates; this method yields preparations for conventional light microscopy (see Fig. 22-2), and the resulting bands are designated G bands. Reverse (R) banding is a Giemsa staining method that produces a pattern of chromosome banding that is the reverse of either the Q or G bands. The structural components of the chromosome that give rise to the banding patterns are uncertain, but the differential distribution of base composition and the state of condensation of the chromatin appear to be involved. The Q bands result from binding of quinacrine stains to adenine- and thymine-rich (AT-rich) regions of DNA, whereas guanine- and cytosine-rich (G-C-rich) regions of the chromosome quench the fluorescence. The G bands appear to be a consequence of differential binding of dye to nonhistone protein overlaying the AT-rich regions.

High-resolution chromosome banding and painting techniques provide precise methods for identification of each chromosome and analysis of chromosome abnormalities, including chromosome rearrangements. A standard nomenclature for identification and designation of individual chromosomes, chromosome regions and bands, and structurally altered chromosomes is embodied in the report of the 1995 International System for Human Cytogenetics Nomenclature. Table 22-2 summarizes the nomenclature applied to sex chromosome anomalies.

The identification of normal and structurally abnormal chromosomes, of supernumerary chromosomes, and of specific DNA sequences and genes (chromosome microdissection) has been revolutionized by the development of fluorescence in situ hybridization (FISH), the labeling of normal or abnormal DNA with fluorescent dyes, so-called chromosome painting. With the use of special cameras (multiplex FISH) or Fourier spectroscopy (spectral karyotyping), a variety of fluorophores, and computer programs, the karyotype of normal and complex chromosome spreads can be determined rapidly (Fig. 22-3).

### Mechanisms of Chromosome Anomalies

**Aneuploidy**

Aneuploids cells contain a total number of chromosomes different from that characteristic of the species. One mechanism producing aneuploidy is nondisjunction, which can occur during either mitotic or meiotic division. Nondisjunction is characterized by failure of either of a pair of sister chromatids or members of a pair of homologous chromosomes to separate during anaphase. As a result, one daughter cell receives an extra chromosome and the other is one short (Fig. 22-4). Aneuploidy can also be caused by anaphase lag, in which there is a simple loss of a chromosome from one or both of the two daughter cells, presumably because of failure of one chromosome to become properly oriented at the equatorial plate during metaphase. If both chromatids are lost, both daughter cell lines lack this chromosome. If only one member of the chromatid pair is lost, the descendants of one daughter cell are normal, and those of the other are one chromosome short (Fig. 22-4).

**Nondisjunction in the oocyte increases with advanced maternal age**: an abnormality in reciprocal recombination in the fetus manifested as a compromised exchange configuration predisposes to missegregation during the completion of meiosis I (less frequently of meiosis II) immediately before ovulation 10 to 40 years later.
Mosaicism

Mosaicism is the presence in an individual of two or more cell lines that differ in chromosomal constitution but originate from a single zygote. This condition can arise only from errors in mitosis after fertilization has occurred, but embryos with abnormal chromosomal makeup are prone to further errors of replication. Mosaicism is more common than first supposed, and many of the seeming paradoxes between genotype and phenotype are attributable to studies in which mosaicism was not rigorously excluded. 14 The difficulty in detecting or, especially, in excluding sex chromosome mosaicism cytogenetically was formidable in the past, but recombinant DNA techniques provide more specific and more accurate detection. 15 The use of X and Y DNA probes and chromosome- and locus-specific FISH analysis together with the polymerase chain reaction

### TABLE 22-2 -- Nomenclature for Describing Human Karyotypes Pertinent to Designating Sex Chromosome Abnormalities

<table>
<thead>
<tr>
<th>ISCN 1995 (22)</th>
<th>Description</th>
<th>Standard Nomenclature</th>
</tr>
</thead>
<tbody>
<tr>
<td>46.XX</td>
<td>Normal female karyotype</td>
<td>XX</td>
</tr>
<tr>
<td>46.XY</td>
<td>Normal male karyotype</td>
<td>XY</td>
</tr>
<tr>
<td>47.XXY</td>
<td>Karyotype with 47 chromosomes including an extra X chromosome</td>
<td>XXY</td>
</tr>
<tr>
<td>45.X</td>
<td>Monosomy X</td>
<td>XØ</td>
</tr>
<tr>
<td>45.X/46.XY</td>
<td>Mosaic karyotype composed of 45,X and 46,XY cell lines</td>
<td>XØ/XY</td>
</tr>
<tr>
<td>p</td>
<td>Short arm</td>
<td>p</td>
</tr>
<tr>
<td>q</td>
<td>Long arm</td>
<td>q</td>
</tr>
<tr>
<td>46.X,del(X)(q21)</td>
<td>Deletion of short arm of X distal to band Xp21</td>
<td>Xp-</td>
</tr>
<tr>
<td>46.X,del(X)(q21)</td>
<td>Deletion of long arm of X distal to band Xq21</td>
<td>Xq-</td>
</tr>
<tr>
<td>46.X,i(Xq)(q10)</td>
<td>Isochromosome of long arm of X</td>
<td>Xqi</td>
</tr>
<tr>
<td>46.X,r(X)(q22;q25)</td>
<td>Ring X chromosome</td>
<td>Xr</td>
</tr>
<tr>
<td>46.X,del(7)(Y;7)(q11;q13)</td>
<td>Translocation of distal fluorescent portion of Y chromosome to long arm of chromosome 7</td>
<td>46.XY(Yq-;7q+)</td>
</tr>
</tbody>
</table>

*Chimerism*

Chimerism is the existence in an individual of two or more cell lines, each of which has a different genetic origin. In the freemartin, a common form of intersex in cattle, chimerism is derived by admixture of hematopoietic and primordial germ cells between biovular twins of opposite sex through anastomotic placental channels. Although in humans it may be difficult to recognize the presence of chimerism if the separate cell lines have the same sex, the presence of cell lines of different sex is marked by a 46,XX/46,XY karyotype. Chimerism can occur by (1) double fertilization (dispermy) of a binoocyte

![Figure 22-3](https://example.com/figure22-3.png)

*Figure 22-3* FISH for SRY analysis in metaphase and interphase 46,XY cells. FISH images illustrating localization of the Vysis SRY probe on the distal short arm of the Y chromosome (Yp11.3) shown in SpectrumOrange and Vysis CEPX, a probe for the X centromere shown in SpectrumGreen. Note the localization of the X (green) and Y (orange) in an interphase nucleus (right). (Courtesy of Philip Cotter and Helen Jenks.)

Chimerism is derived by admixture of hematopoietic and primordial germ cells between biovular twins of opposite sex through anastomotic placental channels. Although in humans it may be difficult to recognize the presence of chimerism if the separate cell lines have the same sex, the presence of cell lines of different sex is marked by a 46,XX/46,XY karyotype. Chimerism can occur by (1) double fertilization (dispermy) of a binoocyte

*Structural Errors*

With the increased ability to characterize the morphology of human chromosomes by banding, FISH, and molecular biologic techniques, subtle as well as more obvious abnormalities of structure can be recognized. Structural errors result from chromosome breakage or partial deletion, often followed by improper reunion of the fragments (Fig. 22-5). Most structural abnormalities are of sufficient size to be visible by light microscopy are characterized by an abnormally long or short chromosome. Chromosome fragments lacking a centromere or containing an additional functional centromere are usually eliminated from the cell. The following are the more common structural abnormalities (see Table 22-2). 16

Isochromosomes are chromosomes with almost identical arms. They

![Figure 22-4](https://example.com/figure22-4.png)

*Figure 22-4* Daughter cell lines can arise from mitotic nondisjunction or anaphase lag during first mitotic division in the zygote. More complex mosaicism can result if the zygote is aneuploid or if replication errors arise beyond the one cell stage. In females, nondisjunction or anaphase lag may involve either the maternal or paternal X chromosome. Deductions regarding the origin of X chromosomes in aneuploid patients can be made by correlating sex-linked traits with those in parents and by using specific DNA probes for analysis.

were thought to arise by transverse rather than longitudinal division of the chromosome (termed centric fission) (Fig. 22-6). This error involves primarily the X and Y chromosomes and usually results in a chromosome consisting of two long arms (e.g., Xq1 or Yq1). Isochromosomes may have either one or two centromeric bands, and some isochromosomes have subtle differences in banding patterns and size of the two arms. These features and the limited evidence that centric fission can occur in human cells led to the hypothesis that most Xq isochromosomes arise from breaks in the short arm close to the centromere (not by centromeric misdivision), with fusion of the sister chromatids, followed by normal division of the centromere and duplication of the entire chromatid to form an
isochromosome; this view was supported by detailed molecular studies (see Fig. 22-6).

Deletion is characterized by detachment and loss of a portion of a chromosome. The notation q refers to deletion of a portion of the long arm, and p refers to deletion of a portion of the short arm.

Duplication occurs when a deleted segment is incorporated into another chromosome, usually the other member of a homologous pair.

Translocations are characterized by exchanges of chromosome segments between two chromosomes.

Ring chromosomes (e.g., a ring X [Xr]) arise by deletions from the ends (telomeres) of a chromosome, with reunion of the new distal portions to form a ring (see Fig. 22-5).

Biologic Functions of the Y Chromosome

Until the advent of human chromosome analysis, it was widely believed that the Y chromosome (Fig. 22-7) was inert and that male determinants were carried on the autosomes. The finding of a 47,XXY sex chromosome constitution in men with Klinefelter's syndrome and only a single X chromosome in women with the syndrome of gonadal dysgenesis (Turner's syndrome) provided convincing evidence that the Y chromosome carries a male-determining gene that can induce testicular development even in the presence of two or more X chromosomes. The presence of a Y chromosome drives testicular differentiation even in individuals with a 49,XXXXY sex chromosome constitution, whereas testicular differentiation does not occur in 45,X individuals. In addition, the Y chromosome is essential to spermatogenesis. In subsequent work the testis-determining gene, SRY, was localized to the short arm of the Y chromosome.

The human Y chromosome, which is composed of the remnants of a conserved region and, predominantly, of a single autosomal region addition estimated to have occurred 80 to 130 million years ago, represents about 2% of the human genome DNA and is approximately 60 Mb in length. A total of 33 genes on the Y chromosome are known, including 10 genes in pseudoautosomal region 1 (PAR1) and 4 genes in PAR2; the Y chromosome is unique because it contains few active genes and a large heterochromatic segment on the distal long arm. It contains genes that influence stature (GCY, growth controlling gene(s) on Yp) in addition to a gene affecting growth located in the PAR1 of the Y and X chromosomes. XY boys have a mean final height of about 188 cm, more than 13 cm taller than their fathers. Phenotypic females with a 46.X derivative karyotype having three doses of SHOX/PHOG are tall, as are XXY patients. Whether the increased stature in XXY patients is related solely to the extra dose of the pseudoautosomal SHOX/PHOG gene or to additional Y-borne genes is not known.

The length of the human Y chromosome varies as much as threefold in normal men. The length and morphology of the Y are heritable, are relatively constant in first-degree male relatives, and exhibit ethnic variation. Most of this variation is in the length of the long arm and in particular in the length of its distal, heterochromatic, brilliantly fluorescent segment in Q-stained preparations (see Fig. 22-8; see also Fig. 22-7). This polymorphism in the size of the fluorescent portion and even loss of part of the distal nonfluorescent portion of the long arm are consistent with normal male sex differentiation and are not associated with known phenotypic effects; consequently, it is likely that a large segment of the long arm of the Y chromosome is not engaged in gene transcription. The long arm of the Y chromosome contains highly repetitive Y chromosome-specific and non-Y chromosome-specific sequences of DNA. The euchromatic short arm and the proximal portion of the long arm of the human Y chromosome make up about 0.5% of the diploid genome (XY + 44 autosomes).

The euchromatic portion of the Y chromosome contains two regions, a Y-specific segment and regions at the distal ends of the short and long arms that are so-called pseudoautosomal regions (PARs) that are homologous to the distal ends of the short and the long arms, respectively, of the X chromosome. The X and Y chromosomes pair and recombine obligately only along these small segments at the distal ends during meiosis. They form chiasmata, thereby maintaining sequence homology and allowing for the proper distribution of sex chromosomes to the daughter cells. This process is critical to sex determination. Genes on the distal short and long arms of the X and Y chromosomes are paired and are not subject to dosage compensation (i.e., gene inactivation); hence, they are expressed as autosomal genes rather than X-linked genes, leading to the PAR designation (see Fig. 22-7).

At least 10 genes are located on the PAR1 of the short arm of the X and Y chromosomes. The short arm PAR (PAR1) is about 2.6 Mb in length, the boundaries being demarcated distally by the telomeres of the X and Y chromosomes and proximally by the Alu repeat sequence on the Y chromosome. The PARs of the X and Y chromosomes are 99% homologous distal to the Alu sequence. Distally on the short arm PAR is PGFL, which encodes a putative quanosome triphosphosphate (GTP) binding protein; next is SHOX/PHOG, then CSP2RA, which encodes the subunit of the granulocyte colony-stimulating factor receptor. Proximal to CSP2RA is IL3RA (encoding the subunit for the IL3 receptor, followed by ANT3 (adenine nucleotide translocase), ASMTL (ASMT-like), and ASMT (acetyl serotonin methyl transferase), XET (X-escape inactivation, function unknown), TRAMP (sequence homology with transposases suggests involvement in transposition), and MIC2 at the PAR boundary. PBDX is the XG blood group gene, and its homologue on the Y chromosome is an expressed pseudogene of XH.

The congruence between short stature and deletions of either Xp or Yp suggested that a gene for stature was present in the distal 700 kb of the PAR1 because patients with only a single copy of this region have short stature. A gene from this 700-kb distal PAR region of Xp and Yp (PAR1), called PHOG (pseudoautosomal homeobox-containing osteogenic gene), is expressed.
mainly in osteogenic cells and bone marrow stromal fibroblasts and encodes a transcription regulatory factor. Its location in the distal PAR, the nature of its predicted protein, and its expression in developing limbs and the first and second pharyngeal pouches suggested that it is involved in mesomelic short stature as well as the skeletal anomalies of Turner's syndrome. Analysis of 32 Leri-Weill patients from 18 families revealed SHOX/PHOG deletions in only 10 of the 18 families, suggesting multiple genetic causes of Leri-Weill syndrome. The phenotypic variation in patients with SHOX/PHOG heterozygous deletions or mutations (i.e., short normal male Turner syndrome stigmata Leri-Weill syndrome) is not understood.

A second PAR is found on the distal long arms of the X and Y chromosomes. This region is 330 kb long and recombines during meiosis at a 2% rate, much slower than the XpYp pseudoautosomal region combines but six times greater than average for X-specific DNA.

forms by the Y chromosome and its homologue on the long arm of the X chromosome (RS54X). On the short arm of the Y chromosome proximal to RS54Y is a gene that encodes for a protein that has 13 CysCys/HisHis zinc fingers and both an acidic and a basic domain and that has been termed ZFY (zinc finger Y).

The fluorescent end of the human Y chromosome in metaphase, when stained with the fluorochrome quinacrine hydrochloride or its mustard derivative, is visualized by using FISH for Y chromosome and X chromosome satellite DNA analysis the sex chromosomes can be identified in interphase cells.

The association of short stature with SHOX/PHOG haplinsufficiency in patients with Xp-, Yp-, in short otherwise "normal" males, and in patients with Leri-Weill syndrome and its expression in developing limbs and the first and second pharyngeal pouches suggested that it is involved in mesomelic short stature as well as the skeletal anomalies of Turner's syndrome. Analysis of 32 Leri-Weill patients from 18 families revealed SHOX/PHOG deletions in only 10 of the 18 families, suggesting multiple genetic causes of Leri-Weill syndrome. The phenotypic variation in patients with SHOX/PHOG heterozygous deletions or mutations (i.e., short normal male Turner syndrome stigmata Leri-Weill syndrome) is not understood.

A second PAR is found on the distal long arms of the X and Y chromosomes. This region is 330 kb long and recombines during meiosis at a 2% rate, much slower than the XpYp pseudoautosomal region combines but six times greater than average for X-specific DNA. The proximal 295 kb contain two genes SYBLL (synaptobrevin-like 1), a gene that may be involved in synaptic signaling, and HSPRY3 (human sprouty 3), a putative intracellular modulator of fibroblast growth factor (FGF) and FGF receptor tyrosine activity, which antagonizes Fas-dependent mitogen-activated protein (MAP) kinase signaling. Unlike genes on PAR1, these genes are inactivated on both the inactive X and the Y chromosome. The distal 35 kb of PAR2 contains two genes that are not inactivated: IL9R, the interleukin-9 receptor, and CXYorf7, a gene of unknown function, located 5 kb from the Xq telomere.

The euochromatic portion of the Y chromosome, the so-called sex-specific region present only in the male, extends from the proximal boundary of the PAR to the heterochromatic portion of the long arm of the chromosome. Deletion analyses of the Y chromosome in 46.XX males and 46.YX females indicate that the segment just proximal to the PAR on the short arm of the Y chromosome carries a gene or genes critical to testicular organogenesis and subsequent male sex differentiation. A 35-kb region immediately adjacent to the PAR boundary contains a gene termed SRY (sex-determining region Y). This gene encodes a testis-specific transcript that exhibits structural homology to two DNA-binding proteins: Mc, a mating-type protein of the fission yeast Schizosaccharomyces pombe, and HMG1 and HMG2, so-called nuclear high-mobility-group proteins. Proximal to SRY is a gene for ribosomal protein subunit-4 (RPS4Y), one of many housekeeping proteins that are encoded in slightly different

forms by the Y chromosome and its homologue on the long arm of the X chromosome (RS54X). On the short arm of the Y chromosome proximal to RS54Y is a gene that encodes for a protein that has 13 CysCys/HisHis zinc fingers and both an acidic and a basic domain and that has been termed ZFY (zinc finger Y). By analogy to similar zinc finger proteins, such as Xenopus transcription factor IIIA, the protein is thought to bind to DNA in a sequence-specific manner and to regulate transcription. Among other genes on the short arm of the Y chromosome are PRKY (protein kinase Y), TSPY (testis-specific protein Y), and AMELY (amelogenin), the gene that encodes the major extracellular matrix enamel protein in the developing tooth bud.

PRKY has a homologous site on the X chromosome (PRKX), which allows for illegitimate recombination between the X and Y chromosomes and hence the production of XX SRY-positive males. TSPY is present as a multicopy locus on the Y chromosome. TSPY may function as an oncogene responsible for gonadoblastoma formation in dysgenetic gonads. Its repetitive units map to the GSY (gonadalblastoma) critical regions on the Y chromosome.

The euchromatic portion of the long arm of the Y chromosome can be divided into three regions: AZFa, AZFb, and AZFc for azoospermic factors a, b, and c. These three nonoverlapping regions contain genes that when deleted result in infertility. AZFa is the region proximal to the centromere of the Y chromosome and it encompasses four genes: USP9Y, a ubiquitin-specific protease that encodes an H-Y antigen epitope, DBY (dead box Y), UTB (ubiquitous leucineproleopeptide repeat motif Y), and TB4Y (thymosin B4, A isoform). AZFb contains five genes, prominent among which are SMCY (selected mouse CDNA, Y), which encodes two H-Y antigen epitopes and RBMY (RNA-binding motif Y), a gene with a putative role in spermatogenesis. AZFc at the distal end of the euchromatic region has a five-gene cluster at its boundary, the most prominent of which is DAZ (deleted in azoospermia), an RNA-binding protein. Deletions of DAZ and one or more members of the RBMY family as well as USP9Y and DBY are a common cause of spermatogenetic failure. Other genes postulated to reside on the Y chromosome but not yet cloned include genes affecting height, on the long arm pericentromeric region GCY, as well as genes preventing the stigmata of the syndrome of gonadal dysgenesis, on the short arm, and genes affecting spermatogenesis.

### Table 22-3: Sex Chromosome Complement Correlated with X Chromatin and Y Bodies in Somatic Interphase Nuclei

<table>
<thead>
<tr>
<th>Sex Chromosomes</th>
<th>X Bodies</th>
<th>Y Bodies</th>
</tr>
</thead>
<tbody>
<tr>
<td>45,X</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>46,XX</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>46,XY</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>47,XXX</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>47,XXX</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>47,XY</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

Q-stained X chromatin bodies have been observed in cultured
female fibroblasts and certain other female tissues, but the intensity of fluorescence of the X body is less, and the size is three to five times larger than that of the Y body (see Fig. 22-8). In sum, more than 30 genes and gene families are currently identified on the Y chromosome.38 40

They are subclassified as follows:

1. Pseudoautosomal genes with identical sequences on X and Y (e.g., IR and MIC2).
2. Those located on the X-Y homologous regions of the nonrecombining region of Y (NRY); these genes are ubiquitously expressed and include DBY and UTY.
3. Genes that are Y specific and expressed only in the testis (e.g., SRY).

Although SRY is Y specific, it does not fit this classification in view of its different pattern of expression and its single copy nature. Surprisingly, nearly half of all genes involved in early stages of spermatogenesis are X linked, a specialty role of the X chromosome in sperm production that emerged as it evolved from an ancestral autosomal.38 39

Biologic Functions of the X Chromosome

The biologic functions of the X chromosome are more complex than those of the Y chromosome. The X chromosome consists of about 160 Mb; landmark DNA sequences (sequence-tagged sites) have been determined over the entire X chromosome. A large number of genes on one of the two X chromosomes of the female undergo X inactivation, a genesilencing mechanism activated in early embryonic development, to balance the expression of X-linked genes between males and females (see later). Genes on the X chromosome have a critical influence on sex determination in both the female and the male and on the differentiation of the somatic sex structures in the male. More than 300 gene loci unrelated to sex development are known to be X linked.38 39

The organization of the X chromosome resembles that of the Y chromosome in that it has both a pseudoautosomal region (PAR) on its distal short arm (PAR1), homologous to that on the Y chromosome (Xp22.3pter), an X-specific region, and a PAR on the long arm (PAR2) (Fig. 22-9). PAR1 of the Xp is the locus for at least 11 genes (PGFL, C5orf26, SHOX/PHOG, IL3Rα, AN73, ASMT, XE7, TRAMP, MIC2, and PBDX), and a gene deletion results in the neurocognitive defects seen in Turner's syndrome. The locus of the PBDX gene (the XG blood group gene) is unusual in that it appears to span the PAR boundary on the X chromosome.47 Immediately proximal to the boundary of the PAR are the loci for many genes, including those for chorondroplasia punctata and steroid sulfatase (STS), the gene encoding the amelogenin enamel protein in the developing tooth bud, and the locus for the zinc finger gene X (ZF5), which cross-hybridizes with DNA probes of the ZFY gene. Because the ZFX protein has 13 zinc fingers with 97% amino acid sequence homology to ZFY, it appears that both these zinc finger proteins may bind to the same nucleic acid sequences. Genes in the area of the X chromosome immediately proximal to the PAR (i.e., Xg [PBDX], STS, KAL1, ZFX) escape X inactivation. Other genes are postulated to reside in this region, including a gene or genes that prevent many of the somatic abnormalities found in the syndrome of gonadal dysgenesis.47 48 Proximal to this region are genes that are not homologous to sequences on the Y chromosome and are subject to dosage compensation by X inactivation on all X chromosomes in excess of one.47

Several genes that play a role in sex determination and differentiation are present on the short arm of the X chromosome. These include the Kallmann syndrome gene, KAL1, deletion or mutation of which results in anosmia and hypogonadotropic hypogonadism.49 KAL1 maps about 1.5 Mb proximal to STS on Xp22.3 and encodes a protein, anosin, that is critical for the migration of the luteinizing hormonereleasing hormone (LHRH) neurosecretory neurons from the olfactory placode to the hypothalamus.49 Proximal to the genes for Duchenne muscular dystrophy (DMD) and glycerol kinase (GK) in the Xp21 region is a locus that contains two overlapping regions, AHC (adrenal hypoplasia congenital) and DSS (dosage-sensitive sex reversal). A gene, DAX1 (DSS/MHC critical region on the X chromosome), has been cloned from this region.47 Deletions and mutations in the DAX1 gene are associated in the male with adrenal hypoplasia and hypogonadotropic hypogonadism.47 Duplication of the DAX1 gene in the XY human (and rodent) results in testicular dysgenesis and lack of or incomplete masculinization of the internal and external genitalia.47 48 Duplications in 46,XX females, however, do not affect ovarian function.47 48 In contrast, deletions of the DAX1 gene have no effect on testicular determination and differentiation and subsequent in utero masculinization of 46,XY' individuals. More proximal on the short arm, a lymphedema critical region has been proposed to reside in Xp11.4.47 48

Two X chromosomes are required in humans for normal ovarian differentiation and follicular maturation: 45,X individuals have bilateral streak gonads. Studies of patients with deletions of the X chromosome indicate that loci on both the short and long arms of the X chromosome are involved in ovarian differentiation and maturation.47 48

The long arm of the X chromosome contains a large number of genes that are subject to X inactivation and are responsible for a wide variety of X-linked traits. The gene for the androgen receptor protein is located in the paracentromeric region of the long arm of the X.47 The RPS4x gene is also located in this region and is subject to X inactivation. This region also contains the X-inactivation center, XIC, the site around which the X chromosome condenses to form the sex chromatin body and from which X inactivation spreads. (See X Chromatin and Gene Expression.)

X Chromatin (X or Barr Body)

Whereas the Y chromosome is one of the smallest human chromosomes and is mainly concerned with testis organogenesis, the X chromosome is the eighth longest
and contains about 5% of the total DNA content of the haploid genome (X + 22 autosomes). Furthermore, the X chromosome contains genes that encode functions involving every system in the body. Because females have twice as much of this genetic material in their cells as males, the biologic differences between the sexes should be far greater than is the case. Theories proposed to explain this paradox are an outgrowth of Barr's pioneering observations of the X chromatin body in somatic cells of females.

In 1949 Barr and Bertram described the presence of a stainable chromatin mass at the periphery of the nucleus in resting ganglion cells of female but not of male cats. This distinguishing characteristic of the female sex is present in most mammalian cells and can be used as a cytologic means of assessing the number of X chromosomes in humans (reviewed in reference 109, Fig. 22-10; see Table 22-3).

The X chromatin body is usually planarconvex, with the flattened side in apposition to the inner surface of the nuclear membrane; in some nuclei it has a bipartite structure. It is about 1 µm in diameter and stains positively for DNA. In certain tissues, such as amniotic membrane, almost every interphase nucleus is chromatin positive. In buccal mucosal smears (the most commonly used preparation for determining the X chromatin pattern), the proportion of X chromosome-positive nuclei in females may be lower than in other somatic tissues, but in most studies these nuclei are detected in no less than 20% of all nuclei.

In polymorphonuclear leukocytes in females, 1% to 15% of neutrophils (mean, 2.5%) have a drumstick-shaped, dense chromatin accessory nuclear appendage not found in normal males (see Fig. 22-10D). These appendages have the same significance as X chromatin in other somatic tissues.

In patients with more than one X chromosome, the maximal number of X chromatin bodies in any diploid nucleus is one less than the total number of X chromosomes. In 47,XXY females or 48,XXY, males, for example, at most two Barr bodies are present in diploid nuclei, whereas 46,XY and 45,X individuals are X chromatin negative (see Table 22-3). Abnormalities in shape and size of the X chromatin body can often be correlated with structural abnormalities of the X chromosome. The X chromatin body is small in females and one deleted X chromosome (46,XXp) and in those with one ring X chromosome (46,Xp). A large X body is associated with a long arm isochromosome (Xq) of a large X chromosome that gives rise to X chromatin bodies (Fig. 22-11). When an X is structurally abnormal, the aberrant X chromosome replicates late and gives rise to the X chromatin (except when the structurally abnormal X is an X-autosome translocation).

X Chromatin and Gene Expression

In 1959 Ohno and co-workers reported the first evidence that X chromatin (the Barr body) arises from only one of the two X chromosomes in the interphase nuclei of female somatic cells. The staining characteristics of such nuclei arise from the fact that a portion of one X chromosome is highly condensed (heteropyknotic); the other X, like the autosomes, is extended and filamentous.280 This difference in staining quality betokens a striking difference in the functional roles of the two X chromosomes. By studying the sequence of incorporation of tritiated thymidine into replicating chromosomes, Grumbach and colleagues281,282 showed that the X chromosome that gives rise to X chromatin completes DNA synthesis later than any other chromosome and that the maximal number of X chromatin bodies in a single diploid nucleus is equal to the number of late-replicating X chromosomes (Fig. 22-11). These observations and the incisive genetic studies of Lyon, Beutler, and others led to the concept that only one X chromosome in each cell is genetically active during interphase, the other X chromosome being heterochromatinized and genetically inactive for most functions.281,282,283 We do not refer here to constitutive heterochromatin but rather to facultative heterochromatin-euchromatic (active) chromosome regions silenced by transformation into a heterochromatin (or inactive) form.284,285

The change in state (inactivation)286,287 of one X chromosome in each female cell occurs during the late blastocyst stage, between the 12th and the 18th day in the human embryo. X inactivation is a multistep process that leads to stable and epigenetic silencing of genes on all X chromosomes in excess of one X. It involves (1) counting (ascertainment of X-autosome location and inactivation of all X chromosomes in excess of 1); (2) random selection of which X chromosome, either the maternal or paternal X, will inactivate; and (3) silencing of initiation, establishment, and maintenance of X-inactivation.286,287,288 The female germ cells beyond the stage of oogonia are the only cell lines known to be exempted from heterochromatinization and inactivation,289 which is seen as a finding in keeping with the requirement for a second X chromosome for normal ovarian differentiation. Both X chromosomes in mouse oocytes are active and code for the X-linked genes for glucose-6-phosphate dehydrogenase and hypoxanthine-guanine phosphoribosyltransferase.290 This observation has been confirmed in human fetal and postnatal oocytes.289 In each somatic cell, however, either the maternally or the paternally derived X chromosome is usually randomly inactivated. Once this transformation is accomplished, the inactive state of that particular X chromosome is transmitted to all descendants of that cell. This control system functions as an epigenetic mechanism of dosage compensation by which each female somatic cell functions virtually as if it had only one active X chromosome.286,287,289,290 The female, in effect, has only a little more active genetic material than does the male. This hypothesis is commonly referred to as the "Lyon hypothesis" (or the "inactivating X theory," or the "fixed differentiation hypothesis of X chromosome behavior") (Fig. 22-12). Although inactivation of structurally normal X chromosomes in individuals with more than one X chromosome in their genome is usually random, instances of skewed inactivation are well documented289,290 XX individuals heterozygous for X-linked immunodeficiencies and mental retardation disorders, Lesch-Nyhan syndrome, or adrenoleukodystrophy may appear to have nonrandom activation of their X chromosomes owing to post-inactivation selection (i.e., in vivo selection against those cells in which the normal allele is inactivated in tissues where the gene product is required).289,291,292 Skewed inactivation of the X chromosome has been reported in families with X-linked diseases and in monozoic twins discordant for an X-linked disease.293,294 If inactivation occurs normally as a random event in a small number of cells, 10% of "normal females" may show an 80:20 proportion of inactivated X chromosomes from one parent and even manifest symptoms of an X-linked mutant allele. Skewed inactivation also occurs in patients with a structurally abnormal X chromosome: the structurally abnormal X chromosome is inactivated, unless it is a part of an X-autosome translocation, in which case the X-autosome translocation will always be active, probably as the result of cell selection. A skewed pattern of X inactivation also has been described in a multigenerational study, suggesting that this character is controlled in some families by one or more X-linked genes.295,296

The fact that normal females function as genetic mosaics is one of X-linked traits is concerned has been documented in the mouse and in humans. For example, Davidson and associates295 demonstrated two populations of cells in females heterozygous for a mutant form of the X-linked gene for glucose-6-phosphate dehydrogenase (see Fig. 22-12). Inactivation of all X chromosomes in excess of one also explains the relatively minor phenotypic changes in women with more than
two X chromosomes (Fig. 22-13). By contrast, trisomy for an autosome as small as chromosome 21, as in Down's syndrome, is usually associated with profound effects. Biochemical analysis of DNA methylation of active and inactive X chromosomes and studies with 5-azacytidine (which impairs methylation of cytosine) suggest that DNA methylation plays an important role in the maintenance of X chromosome inactivation, late replication, and sex chromatin formation. DNA methylation differs in the two X chromosomes. The double-stranded palindromic cytosome-guanine dinucleotide clusters, the so-called CpG islands, commonly found at the 5' end of genes, are methylated mainly on the inactive X chromosome. The methylated cytosome residues serve to maintain the suppressed transcriptional activity and relative resistance to nuclelease. The chromatin of the inactive X contains more unacyetylated histones (histones H3 and H4) than the chromatin of the active X chromosome does. Histones function as DNA-packaging proteins; the DNA helix is wrapped around core

Figure 22-12 Diagram of the fixed differentiation or Lyon hypothesis of X chromosome behavior in somatic cells of the human female. At the late blastocyst stage (the time when X chromatin can first be identified), one of the two X chromosomes becomes heterochromatinized in each cell and gives rise to an X chromatin body; it is by chance in each cell whether this differentiation involves a maternally derived X (M) or a paternally derived X (P). Once differentiation has occurred, this characteristic is fixed in succeeding generations of somatic cells. Most of the genes on the heterochromatic portion of an X chromosome are suppressed or inactivated, thus serving as a means of "dosage compensation" for the greater number of X-linked genes in the female than in the male. This mechanism has an important bearing on expressivity and penetrance of an X-linked mutant gene in a heterozygous female. In the diagram, the maternally derived X carries a mutant gene (a) that is expressed only in cells in which this X is the Lyon-silenced, euchromatic active X (white X). Although the heterochromatinized X (black X) in this diagram is represented as wholly inactive, some loci on the heterochromatinized X do remain active and exert genetic effects. The female germ cell line beyond the oogonia stage is exempted from heterochromatinization.

Figure 22-13 Diploid somatic cells from a girl with a 49,XXXX karyotype. A. Four X chromatin bodies (arrows) in an interphase nucleus from a culture of skin fibroblasts. B, Autoradiogram of metaphase chromosomes, illustrating four areas of high grain density overlying four of the five X chromosomes. C, An autoradiogram of an interphase nucleus in a culture of skin fibroblasts; four peripheral "hot" areas (arrows) of high grain density overlie four X chromatin bodies and provide direct evidence that each X chromatin body is derived from one late-labeling X chromosome. (Modified from Gribanbach MM, Morishima A, Taylor JA. Human sex chromosome abnormalities in relation to DNA replication and heterochromatinization. Proc Nat Acad Sci USA 1963; 49:581589; and Gribanbach MM. On the significance of sex chromatin. In Second International Congress on Congenital Malformations. New York: International Medical Congress, 1964, pp. 62-67.)

a common requisite for the activation of gene expression, is mediated by histone acetyltransferase A, encoded by the HATA gene.

In contrast to the mouse, human X chromosome inactivation does not involve the entire chromosome. The heteropyknotic X in the human female is only segmentally inactive in terms of transcriptional activity, early studies of the heteropyknotized

X chromosome suggest that about 21% of genes on Xp escape inactivation in contrast to about 3% of genes on Xq. Genes on the PAR of the short and long arm as well as genes scattered along the short and long arms of the heteropyknotic X chromosome escape inactivation. Individuals with a 45.X or 47.XXY constitution, for example, have abnormalities both in gonadal development and in somatic features unrelated to sex. Furthermore, as noted previously, the gene for the red blood cell antigen Xg (PBDXX), the ST3 gene, and the ZFX genes escape inactivation and are active on both X chromosomes in the female; these genes have been mapped to the distal part of the short arm of the X, outside the PAR. Two genes, XIST (X-specific transcripts) and RPS4, which are located on the proximal long arm of the X chromosome, escape inactivation on the heteropyknotic X chromosome. Inactivation of the X chromosome is mediated by a cis-acting region of the X chromosome, the XIC (X-inactivating center), from which inactivation spreads along the X. The XIST gene, discovered by Willard and his colleagues, is an essential component of the XIC, at Xq13.6-14.1 (Fig. 22-13).

XIST is a unique gene in that its product, a noncoding RNA, is expressed in high levels only on the inactive X chromosome. Its expression correlates with the inactivation of the X chromosome in female somatic cells and with meiosis in spermatogonia. The XIST allele is turned off on the active X; hence, its lack of expression on an X chromosome indicates that the gene is transcriptionally active. Human and mouse XIST genes encode a long noncoding RNA transcript that is retained in the nucleus "coating" the inactive X chromosome and initiating the gene silencing process. Knockout of the XIST gene in embryonic stem cells prevents X inactivation in cis, whereas deletions of 65 kb of DNA at the 3' end of Xist (mouse homologue) result in that chromosome always being inactivated. In the mouse the gene-silencing function of Xist is lost with deletion of a conserved repeat sequence at the 5' end, but this mutant Xist continued to exhibit chromatin association and spreading. Additionally, the insertion of ectopic copies of Xist into autosomes of murine stem cells results in the molecular and heterochromatic features of X inactivation, including Xist RNA association in cis, gene inactivation, late replication, and a decrease in histone H4 acetylation. XIST apparently is not required for maintenance of X inactivation. The silent Xist gene on the active X chromosome is fully methylated at its 5' end. In the mouse placenta, X inactivation is imprinted the paternal X chromosome is inactivated. A mouse Xist antisense gene, Tsa, is located 15 kb downstream of Xist and extends across the Xist locus on the opposite strand. Tsa is thought to play a role in the repression of the maternal Xist allele in the mouse. However, the suppression of Xist by cis-acting Tsa, which prevents the initiation of X inactivation, does not account for random selection of the X.

Lee and co-workers searched for a trans-acting factor involved in X-chromosome choice in the mouse. They found a candidate, trans-acting molecule, the ubiquitous chromatin insulator and transcription regulator CTCF, which has 11 zinc fingers. However, Pericic and associates using chemical mutagenesis detected two genetically different autosomal dominant mutations (not involving CTCF) that affect X chromosome choice in the early embryo. They designated these mutations X-inactivation autosomal factors 1 and 2.

Migeon and co-workers have identified the human homologue of Tsa, TSIX on the X chromosome. However, unlike Tsa, TSIX RNA is truncated at the 5' end, does not cover the XIST promoter, and does not have a CpG island, which is critical for the function of Tsa. These differences have led Migeon to question the function of TSIX in X chromosome inactivation in humans. Its apparent lack of functionality may explain the discrepancy in X-chromosome imprinting in the human placenta between the humanc and rodent and creates uncertainty about the role of the human TSIX gene in X chromosome choice. The progress over the past decade in illuminating the epigenetic mechanism of X inactivation is remarkable.

Some patients with a 45,X/46,X "tiny" ring X chromosome karyotype differ in phenotype from other patients with gonadal dysgenesis with a ring X chromosome in their genome. These tiny ring X chromosomes do not express XIST, have histone H4 acetylation at a level consistent with that found on an active X chromosome, and contain genes that are active. These tiny ring X chromosomes that lack the XIST gene do not undergo X inactivation; as a consequence, the phenotype in affected patients results from functional disomy caused by lack of dosage compensation. In general, patients with small ring X chromosomes have a greater incidence of mental retardation than those with other forms of Turner's syndrome. Other congenital malformations such as syndactyly and abnormal facies are common.
Genes and Testicular Organogenesis

The genetic sex of the zygote is established by fertilization of a normal ovum by an X- or Y-bearing sperm, and the mechanisms involved in the translation of genetic sex into a testis or an ovary are understood in broad terms. From the early days of human chromosome analysis, compelling evidence was obtained for the regulation of testicular gonadogenesis by a gene (or genes) on the Y chromosome. Indeed, sex determination is essentially testis determination. The short arm of the Y chromosome contains a gene (SRY) that controls testis differentiation and, hence, maleness. The gene acts in a dominant fashion and leads to differentiation of the bipotential gonad as a testis. Several hypotheses were proposed to explain testicular morphogenesis (Fig. 22-14). H-Y antigen and ZFY were proposed as the sex determinators, but that proposal was discarded.\(^\text{144}\)\(^\text{145}\) Reports beginning in 1988 have provided compelling evidence that the SRY gene is the master gene that controls male sex determination.

H-Y Antigen

In 1955 Eichwald and Stilms discovered in males of a highly inbred strain of mice the H-Y antigen, a male-specific cell membrane component encoded by the Y chromosome that causes rejection by female mice of skin grafts from males of the same strain. Antibodies to H-Y antigen (now identified serologically) in male-grafted female mice by Goldberg and associates in 1971 and were utilized for measurement of H-Y antigen. Initial reports suggested that H-Y antigen was a good candidate for the testicular determining factor (TDF) on the Y chromosome, but the lack of reproducibility of the H-Y antigen assay led to increasing skepticism about its role in testicular determination. Finally, it was demonstrated that the gene for H-Y antigen is located on the long arm of the Y chromosome, separate and distinct from the gene for male sex determination.\(^\text{146}\)

ZFY Gene

In 1987 Page and associates proposed that the sex-determining function of the Y chromosome is located within a 140-kb segment of the short arm of the Y chromosome (see Fig. 22-14). A gene in this region encodes a protein with 13 CytoCysHisHis zinc fingers at the carboxyl terminus (and a basic and acidic region at the amino terminus).\(^\text{147}\) By analogy to Xenopus transcription factor II A, it was suggested that this protein binds to DNA and/or RNA in a sequence-specific manner and regulates transcription and that this zinc finger protein, ZFY, is the primary sex-determining signal on the Y chromosome. A sequence homologous to ZFY is present on the X chromosome (ZFX) in the Xp21.2-p22.1 region.\(^\text{148}\) The latter finding initially suggested that X inactivation (dosage compensation) might play a role in sex determination, but ZFX escapes inactivation, so X-chromosome inactivation cannot play a role in this process.\(^\text{149}\) Convincing evidence has shown that ZFY is not the testis-determining gene. Furthermore, in metatherian species (marsupials), sex determination is dependent even though ZFY-related sequences are not located on the X and Y chromosomes of these animals but rather on autosomes.\(^\text{150}\)

SRY Gene

The long quest for the testes-determining factor (TDF) finally met with success with the identification of the SRY gene. In 1989, Palmer and co-workers described three 46,XX males and one true hermaphrodite, the sibling of one of the 46,XX males.\(^\text{151}\) They were all ZFY negative, in spite of evidence for a Y-to-X chromosome exchange as the mechanism of their XX karyotype in the presence of testes with male sex differentiation.\(^\text{152}\) The fragment of Y chromosome translocated to the X chromosome in these patients involved sequences that were distal to the ZFY locus on the short arm of the Y chromosome. This demonstration of testes in patients with a Y fragment but no ZFY sequences, along with studies in marsupials\(^\text{153}\) and mice, doomed the hypothesis that ZFY was the TDF on the Y chromosome.\(^\text{154}\) The Y-to-X exchange in these four patients involved Y-specific sequences located within 35 kb of the boundary of the PAR on the short arm of the Y chromosome (Fig. 22-15).\(^\text{155}\) A 21-krb clone, pY53.3, was identified in this region within 8 kb of the PAR boundary.\(^\text{156}\) This probe detected male sequences in a wide variety of eutherian mammals.\(^\text{157}\)

Studies in mice established that Sry (the mouse homologue of the human SRY gene) is the TDF. Sry is present in the Sxr,XX mouse, a "male" mouse that has the smallest piece of Y chromosome known to code for testicular determination and differentiation translocated to the X chromosome.\(^\text{158}\) Sry is absent in the XY fertile "female" mouse, which has an 11q deletion involving the testes-determining region.\(^\text{159}\) Further studies in the mouse model demonstrated that Sry is expressed in the embryonic genital ridge for only a brief period, from 10.5 to 12.5 days after coitus and 24 hours before the genital ridge differentiates into a testis.\(^\text{160}\) Sry expression is limited to pre-Sertoli cells in the genital ridge, in contrast, Sry in the adult testes is expressed in the germ cells.\(^\text{161}\) The function of Sry transcripts in the adult mouse testis is unknown, because the transcripts are circular and are not associated with polysomes; they appear not to be translated into a protein.\(^\text{162}\)

Definitive proof that Sry was the TDF came from the demonstration that 46,XX mice with a transgene that contains a 14-kb piece of the Y chromosome including Sry differentiate as males with testes.\(^\text{163}\) In the first series of animals, one 46,XX progeny that expressed the transfact ed Sry gene was a well-differentiated 46,XX male that exhibited appropriate male sexual mating behavior.\(^\text{164}\) Histologic examination of the testes revealed normal somatic elements, absent spermatogenesis, and degenerating germ cells.\(^\text{165}\) The other two mice were 46,XX females; one was able to transmit the Sry transgene to her progeny, resulting in the generation of 46,XX males; 46,XX hermaphrodites; and 46,XX females.\(^\text{166}\) That the Sry transgene produced sex reversal in only 25% of transfected embryos may be attributable to the incorporation of the transgene into a region of the genome where it is either not expressed, expressed at a low level, or expressed late in relation to gonadal determination and differentiation.\(^\text{167}\) Nevertheless, this critical transgene experiment proved conclusively that Sry is the only gene on the Y chromosome necessary for testes determination and differentiation and that Sry is the TDF gene. In addition, these studies indicated that Y-borne genes other than Sry are involved in the regulation of spermatogenesis.

Evidence in the human from sex-reversed 46,XX males and 46,XY females confirms the conclusion that SRY is the TDF. In humans, an aberrant Y-to-X interchange during paternal spermatogenesis can transfer Y-specific loci to the X chromosome, and 80% of 46,XX males have a variable amount of the Y chromosome translocated to the X.\(^\text{168}\) All these patients are SRY positive.\(^\text{169}\) Between 30% to 40% of Xp/Yp interchanges occur between Xp22.3 (PRKX gene) and a region on the Y chromosome (PRKY gene) proximal to the SRY gene.\(^\text{170}\) The PRKX/PRKY genes are 94% homologus, are oriented in the same direction, and encode proteins with an adenosine triphosphate (ATP)-binding domain and a catalytic domain.\(^\text{171}\) The reciprocal translocation (i.e., Xp22.3 Yp31) results in 46,XY females who are SRY negative.\(^\text{172}\) In general, SRY-negative 46,XX males have an increased prevalence of ambiguity of the external genitalia and siblings with true hermaphroditism.\(^\text{173}\) The familial occurrence of 46,XX males and 46,XX true hermaphrodites who are SRY negative suggests the constitutive activation or inactivation of an X-linked or autosomal downstream gene (or genes) in the sex determination and differentiation cascade that is normally regulated by SRY.

Fifteen to 20 percent of females with the complete form of 46,XY gonadal dysgenesis have inactivating mutations in the
single exon SRY gene, an architectural transcription factor. Most of the mutations occur in the DNA-binding domain of the SRY protein (Fig. 22-16), the high-mobility-group (HMG) box that appears to act as a transcription factor and has the capacity of binding to and binding DNA. Rare mutations in the S and Z \( \sigma \) and \( \epsilon \) flanking regions in patients have led to complete and partial gonadal dysgenesis, respectively. The HMG box contains two nuclear localization signals (NLS) that bind calmodulin and importin B. Mutations in these nuclear localization signal domains in the HMG box of SRY result in failure to transport the SRY protein into the nucleus and result in consequent XY gonadal dysgenesis. The evidence in the mouse and in humans with sex reversal strongly support the critical role of SRY in sex determination and male sex differentiation.

The human SRY gene contains no introns and produces a 900-base-pair transcript that encodes a protein of 204 residues with three domains: an amino-terminal domain, a central DNA-binding domain consisting of a single HMG box, and a carboxy-terminal domain. In humans it is expressed in 46,XY gonads coincident with sex cord formation and it persists (unlike the brief expression of Sry in the embryonic mouse Sertoli cell) until at least 18 weeks of gestation in fetal Sertoli cells as well as adult Sertoli and germ cells. Comparison of nucleotide sequences of the SRY gene from different species indicates that only the HMG box is conserved. The HMG box is about an 80-amino acid residue domain that is similar to the DNA-binding domain of over 100 genes. A family of genes referred to as SOX (SRY-like HMG box) genes, exists in which the HMG region exhibits more than 60% sequence homology with that of SRY. The SRY protein binds specifically to the linear consensus DNA sequence 5'-ATGAACT(A/T)-3' and nonspecifically to cruciform (four-way junction) DNA. The SRY/HMG protein is made up of three helices and an extended loop-like domain that encodes a horseshoe shape and presents a concave surface to the DNA for sequence-specific binding. DNA binding occurs in the minor groove of the DNA and results in a bend of 40 to 70 degrees from linearity in the DNA, conforming to the shape of the HMG box. Additional conformational changes in the DNA include helix unwinding and minor groove expansion. The carboxyl terminus of SRY contains a 7-amino acid that can bind PDZ (P, Q, disc large, Zn\(_2\)) domains in the nuclear protein SIP1 (SRY-interacting protein 1). A 41-residue deletion in the carboxyl terminus of SRY resulted in XY gonadal dysgenesis. Furthermore, phosphorylation


SRY has no recognizable trans-activation domain. The control of gene expression by SRY may be mediated by the conformational changes in DNA that result in the approximation of distant regulatory elements of the transcriptional apparatus, thereby allowing them to interact with one another. Analysis of mutations in the HMG box in women with 46,XY gonadal dysgenesis suggests that the spatial rearrangements (bending) in DNA produced by the HMG-domain protein are critical to its activity, as is its bending. Induction of a structural bend in the DNA helix may allow the interaction of other spatially dependent proteins with the DNA. Mutations in the SRY HMG box that affect binding, bending, or nuclear transport of SRY and yet undefined mechanisms can result in the loss of transcriptional activity. The transcriptional activity of SRY has been demonstrated in vitro with the use of FOS-related antigen-1 promoter constructs. However, neither the upstream regulatory genes nor the downstream targets of SRY have been ascertained. The consensus sequence (ATGAACT(A/T)) is ubiquitous in the human genome, occurring at more than 105 sites, which makes it difficult to ascribe specificity to the interaction of SRY with a specific gene based solely on the presence of the consensus sequence in its promoter. The target gene is expressed in pre-Sertoli cells in temporal relationship to SRY protein and is directly transactivated by it. Because of the lack of an apparent strict relation between the absence of SRY and testicular development in some 46,XX males and 46,XX true hermaphrodites, McEreavy and associates proposed that the main function of SRY is to repress a putative gene termed “Z” that itself represses differentiation of the testis. At least four cellular roles for SRY protein have been defined. They include (1) the induction of Sertoli cell differentiation, (2) the migration of mesonephric cells into the genital ridge, (3) the proliferation of cells in the genital ridge, and (4) male-specific vasculature with recruitment of a large number of endothelial cells from the mesonephros. However, as noted previously, in spite of more than 10 years of research on the SRY gene as yet we remain uncertain of its upstream regulators or downstream target genes. The strongest contenders are the downstream autosomal gene SOX9 (see later) and upstream WTX (KTSW) isoform.

In sum, the evolutionary conserved HMG domain of SRY protein seems the only component essential for function. The only candidate downstream gene for SRY, at present, is the activation of SOX9. Finally, Graves has challenged the conventional wisdom that the SRY is the ultimate and all-powerful male determiner. She points out that Sry is an “ephemeral gene” recruited recently from a transcription factor possibly involved in brain development, is younger than the Y chromosome, and whose structure and function in evolutionary terms has evolved relatively rapidly. Graves estimates that the human Y chromosome and its SRY gene might only last another 5 to 10 million years.

Autosomal and X Chromosomal Genes

Other genes on autosomes and the X chromosome participate in the testis determination dosage-sensitive combinatorial network of complex interactions. (Table 22-4). These still incompletely understood, regulatory interactions are more than a linear cascade.

SOX9

Campomelic dysplasia is a severe skeletal malformation syndrome associated with an increased prevalence (75%) of 46,XY gonadal dysgenesis and consequent male to female sex reversal. The campomelic dysplasia and sex reversal locus has been mapped to 17q24.3-25.1, and a gene designated as SOX9 (SOX9-like HMG box) has been cloned from this locus. SOX9 belongs to a family of HMG domain protein transcription factors related to the “testis determining factor” SRY that share very similar HMG box DNA-binding attributes. SOX9, in contrast to SRY, has an intron and a well-defined transcriptional trans-activation domain. The human SOX9 gene from different species indicates that only the HMG box is conserved. Similar to SRY, SOX9 has nuclear localization signal sequences at either end of the HMG box that when mutated can reduce nuclear transport or result in campomelic dysplasia and XY sex reversal. SOX9 is expressed in the kidney, central nervous system, cartilage, chondrogenic precursor cells, and Sertoli cells of the testes. The SOX9 protein HMG box, about 70% homologous to the SRY HMG box, binds to a specific sequence, 5'-AGAACATG-3', in the minor groove of DNA, resulting in unwinding and bending of the DNA in a manner similar to other SOX proteins. Furthermore, SOX9 binds to the 6-base-pair DNA sequences ATGAAT and CACAAT found in the chondrocyte specific enhancer of the first intron of the human type II collagen gene (COL2A1), activating this gene and inducing chondrogenesis. SOX9 is a heterozygous mutation in the SOX9 gene can result in campomelic dysplasia in the absence of sex reversal in 46,XY individuals; chondrogenesis seems more sensitive to SOX9 gene dosage than sex determination. A mouse with haploinsufficiency of SOX9 had defective cartilage development and premature mineralization of cartilage, a phenoctype of the severe skeletal malformations seen in campomelic dysplasia. SOX9 haploinsufficient male mice did not result in XY sex reversal. Notably, lethality in mice with the SOX9 haploinsufficient males were normal. Other features such as micrognathia, cleft palate, respiratory distress, and neonatal death are similar to those in the affected human.

Whereas mutations in the SRY gene occur predominantly in the Y chromosome, they occur throughout the SOX9 gene (see Fig. 22-17). Mutations resulting in campomelic dysplasia and/or XY sex reversal include splice acceptor/donor changes, missense, nonsense, translocation, and frameshift mutations. Two of the major classes of mutations causing campomelic dysplasia and/or XY sex reversal are amino acid substitutions in the SOX9 gene and truncations or frameshifts that affect the carboxy-terminal (trans-activation) domain of SOX9.

SOX9 plays a critical role in the sex-determination pathway in all eutherian vertebrates. It is sexually dimorphically expressed; SOX9 is up-regulated in the testis-determining pathway shortly after SRY is expressed in the mouse as well as human. SOX9 and SRY are both expressed in Sertoli cell precursors and Sertoli cells. Expression of SOX9 is restricted to the nuclei of Sertoli cells after 6.5 weeks of gestation in the human male embryo, whereas its expression remains cytosolic in the XX embryonic gonad. In the XX embryo, SOX9 is not expressed in the mouse ovary; in the human a low level of expression is described...
Haploinsufficiency: XY male pseudohermaphrodite (70%),
Mutations: blepharophimosis/ptosis/epicanthus inversus

Human
Male pseudohermaphrodite, multiple congenital anomalies

Mouse Phenotype
Duplication: XY male pseudohermaphrodite

Transcription
Homozygous missense mutation (ATG ACG) in one patient

Orphan nuclear
Heterozygous mutation in exons encoding zinc finger motifs:

Homozygous missense mutation: XY sex reversal with adrenal insufficiency XX: adrenal insufficiency, normal ovaries

Mutations: XY gonadal dysgenesis

Encodes
Signaling

Deletions and mutations: male pseudohermaphrodite,
10q25-qter

Translocation: to X or autosome XX male

Mutation: XY fertile female

### TABLE 22-4 – Genes Involved in Human Gonadogenesis

<table>
<thead>
<tr>
<th>Gene</th>
<th>Encodes</th>
<th>Human Gene Locus</th>
<th>Human Phenotype</th>
<th>Mouse Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT1</td>
<td>Transcription</td>
<td>11p13</td>
<td>Heterozygous mutation in exons encoding zinc finger motifs: Denys-Drash syndrome (male pseudohermaphrodite)</td>
<td>Null: no kidneys or gonads</td>
</tr>
<tr>
<td></td>
<td>factor</td>
<td></td>
<td>Heterozygous mutation in splice site junction with loss of +KTS isoform: Frasier syndrome (XY gonadal dysgenesis)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Deletion: WAGR syndrome (XY, ambiguous genitalia, aniridia, etc.)</td>
<td></td>
</tr>
<tr>
<td>SF1</td>
<td>Orphan nuclear</td>
<td>9q33</td>
<td>Heterozygous missense mutation: XY sex reversal with adrenal insufficiency XX: adrenal insufficiency, normal ovaries</td>
<td>Null: no adrenals or gonads</td>
</tr>
<tr>
<td></td>
<td>receptor</td>
<td></td>
<td>Homozygous missense mutation: XY sex reversal with adrenal insufficiency; familial heterozygotes not affected</td>
<td></td>
</tr>
<tr>
<td>SRY</td>
<td>Transcription</td>
<td>Yp11.3</td>
<td>Mutation: XY gonadal dysgenesis</td>
<td>Mutation: XY fertile female</td>
</tr>
<tr>
<td></td>
<td>factor</td>
<td></td>
<td>Translocation: to X XX male</td>
<td>Translocation: to X or autosome XX male</td>
</tr>
<tr>
<td>SOX9</td>
<td>Transcription</td>
<td>17q24</td>
<td>Haploinsufficiency: XY male pseudohermaphrodite (70%), camptomelic dysplasia</td>
<td>Conditional (gonad) SOX9 transgenic: XX male</td>
</tr>
<tr>
<td></td>
<td>factor</td>
<td></td>
<td>Insertional mutation: XX male</td>
<td>Null mutation: XY male pseudohermaphrodite, camptomelic dysplasia</td>
</tr>
<tr>
<td>DMRT1</td>
<td>Transcription</td>
<td>9p24.3</td>
<td>Deletion of 9p24.3pter haploinsufficiency: XY male pseudohermaphrodite, anomalies, mental retardation XX females have variable ovarian function</td>
<td>Null: normal male sex differentiation</td>
</tr>
<tr>
<td></td>
<td>factor</td>
<td></td>
<td>Phenotype: postnatal loss of germ cells and Sertoli cells &quot;vanishing testes&quot;</td>
<td></td>
</tr>
<tr>
<td>DAX1</td>
<td>Orphan nuclear</td>
<td>Xp21.3</td>
<td>Mutation: XY (normal testis differentiation), adrenal hypoplasia, hypogonadotropic hypogonadism</td>
<td>XY null: infertile male</td>
</tr>
<tr>
<td></td>
<td>receptor</td>
<td></td>
<td>XX null: ovaries</td>
<td>Transgenic XY: female</td>
</tr>
<tr>
<td></td>
<td>Transcription</td>
<td></td>
<td>DAX1 transgenic: XX male</td>
<td></td>
</tr>
<tr>
<td></td>
<td>repressor</td>
<td></td>
<td>DAX1 transgenic: XX male</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Null mutation: XY male pseudohermaphrodite</td>
<td></td>
</tr>
<tr>
<td>WNT4</td>
<td>Signaling</td>
<td>1p35</td>
<td>Duplication: XY male pseudohermaphrodite</td>
<td>Null: XX mouse lacks müllerian ducts, masculinization of female gonad androgen secretion by &quot;gonads,&quot; results in wolfian ducts but female external genitalia, loss of oocytes</td>
</tr>
<tr>
<td></td>
<td>molecule</td>
<td></td>
<td>Null: XX mouse lacks müllerian ducts, masculinization of female gonad androgen secretion by &quot;gonads,&quot; results in wolfian ducts but female external genitalia, loss of oocytes</td>
<td></td>
</tr>
<tr>
<td>ATRX</td>
<td>(XH2) Helicase</td>
<td>Xq13.3</td>
<td>Deletions and mutations: male pseudohermaphrodite, -thalasssemia, mental retardation</td>
<td>ND.</td>
</tr>
<tr>
<td>10q</td>
<td>?</td>
<td>10q25-pter</td>
<td>Male pseudohermaphrodite, multiple congenital anomalies syndrome</td>
<td>ND.</td>
</tr>
<tr>
<td>FOXL2</td>
<td>Transcription</td>
<td>3q23</td>
<td>Mutations: blepharophimosis/postis/epicanthus inversus (BPES) syndrome</td>
<td>Deletion of 1q43 in the goat (homologous to 3q23 in humans) affects FOXL2 and PIST1 causing the polled intersex syndrome (lack of horns and XX sex reversal)</td>
</tr>
<tr>
<td></td>
<td>factor</td>
<td></td>
<td>BPES type 1: premature ovarian failure</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Affected males are fertile</td>
<td></td>
</tr>
<tr>
<td>DHH</td>
<td>Signaling</td>
<td>12q13.1</td>
<td>Homozygous missense mutation (ATG ACG) in one patient with 46,XY partial gonadal dysgenesis and minifascicular neuropathy</td>
<td>Null mutations, two phenotypes, 7.5% males, 92.5% females</td>
</tr>
<tr>
<td></td>
<td>molecule</td>
<td></td>
<td>Testes from feminized XY mice lacked adult Leydig cells</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>External genitalia female, streak on one side with hemi-tube, hypoplastic testes on other side</td>
<td></td>
</tr>
</tbody>
</table>

*ND. not described in the mouse. Not described in the human but identified in the mouse: Lim1, Lim2, Emx2, Ms3 (all transcription factors), Fgf-9, Vanin, and Nexin.

copies of the DAX1 gene at this dosage-sensitive sex reversal locus (DSS) impairs testicular differentiation despite a normal-functioning SRY gene. The smallest Xp21 duplication found in an XY sex-reversed patient is a 160-kb region of Xp21 that contains the DAX1 gene. Deletions or intragenic mutations in the DAX1 gene cause X-linked cytomegalic congenital adrenal hypoplasia and hypogonadotropic hypogonadism. SOX9 encodes a transcriptional regulator. The gene consists of two exons separated by an intron. The encoded protein has 410 amino acids and is composed of a

46,XY individuals with duplications of the Xp21 region of the short arm of the X chromosome have dysgenetic gonads and male-to-female sex reversal. The presence of two active
SRY, binds to heat shock protein 70 (Hsp70) and has a trans-activation domain at the carboxy-terminal end. Selected mutations causing sex reversal and campomelic dysplasia are indicated by the solid circles, those causing only campomelia are indicated by the open triangles, whereas SRY mutations occur primarily on the HMG box. SOX9 mutations appear to occur throughout the gene.

carboxy-terminal half that is similar to the ligand-binding region of other nuclear receptors. A ligand for this orphan nuclear receptor has not been identified. The DAX1 protein has a unique DNA-binding domain that consists of three repeats of a novel amino acid sequence between amino acids 1 and 199 with another partial repeat extending into the ligand-binding region. Evidence indicates that DAX1 participates in a highly complex transcriptional network; DAX1 appears to act primarily as a transcriptional repressor either by binding to DNA hairpins in promoters through a direct effect on transcription targets or by the recruitment of co-repressor proteins. DAX1 is expressed in temporal conjuction with steroidogenic factor 1 (SF1) in the developing hypothalamus, pituitary, adrenal, and gonads. The evidence that DAX1/SF1 (lower case = mouse homologue) is the DSS gene has been demonstrated by studies in mice. DAX1/SF1 is expressed in Sertoli cells of the developing gonad along with other important testsis-determining factors such as WT1/Wt1, Sf1/Sf1, Sox9/Sox9, and Sry/Sry. A high dosage of Dax1 in transgenic mice retards the development of the embryonic tests in the mouse strain Postivusinus, which has a weak Sry allele, and causes complete female development in the mouse strain Tuxedo. These observations support the hypothesis that (1) the normally developing ovary in normal DAX1 mutation in a human female and in the XX Dax1 knockout mouse, (2) the normal testis determines in males with WT1 mutation, and (3) the XXY sex reversal in patients with a double dose of DAX1 indicate that the DAX1 gene is not critical for either ovarian or testicular differentiation but a double dose of DAX1 acts as an anti-testsis gene and leads to dosage-sensitive sex reversal in the affected genetic XY individual and female external genitalia.

WT1

The Wilms tumor suppressor gene (WT1) was first isolated from chromosome 11p13 because of its association with the contiguous gene deletion syndrome WAGR (Wilms tumor, Aniridia, Genitourinary anomalies, and mental Retardation). The gene WT1 encodes a transcription factor with four carboxy-terminal zinc finger DNA-binding domains and an amino-terminal proline/glutamine-rich regulatory region. WT1 contains 10 exons, including two alternative splice sites in exons 5 and 7 that can give rise to as many as 24 isoforms by means of alternative splicing, alternative translation start sites, and RNA editing. The major isoforms produced exist in a constant ratio; they are the result of alternative splicing of exon 5, with the insertion of a 17-amino acid segment encoded by exon 5 in the middle of the protein and the use of an alternative splice donor site at the end of exon 5, resulting in the insertion of three amino acids, lysine-threonine-serine (+KTS), between zinc finger 3 and 4. This +KTS isoforms affect DNA binding and RNA processing, whereas -KTS isoforms appear to act mainly as regulators of transcription. In vivo and in vitro studies suggest that WT1 acts as both a transcriptional activator and repressor. WT1 (+KTS) up-regulates SRY; two consensus sites for the binding of WT1 are present in the SRY promoter. In addition, there is temporal overlapping in the expression of WT1 and SRY in the human fetal XY gonad. WT1/WT1 is expressed in fetal renal mesenchyme, in the primordial gonads (i.e., the mouse genital ridge at 9 days post coitus), in the Sertoli cells of the developing seminiferous tubules, and in adult Sertoli and granulosa cells. Its critical role in sex determination and differentiation can be seen from studies in mice as well as humans. Knockout of the WT1 gene in mice results in the absence of gonads and kidneys in both male and female mice.

Only 10% to 15% of Wilms' tumors have a WT1 mutation, but two syndromes the Denys-Drash and Frasier are associated with a dysfunctional WT1 gene. The Denys-Drash syndrome, characterized by Wilms tumor, severe renal disease due to mesangial sclerosis, and dysgenetic gonads, which in the affected male leads to ambiguous genitalia. is caused most commonly by a heterozygous point mutation in the exons that encode the zinc finger motifs resulting in the expression of a dominant-negative protein that inhibits the function of the WT1 protein encoded by the normal gene on the opposite allele. The Frasier syndrome, a variant characterized by the XY complete gonadal dysgenesis resulting in sex reversal, late-onset glomerulopathy, foot deformities, and prenatal growth retardation, is also associated with the expression of a dominant-negative WT1. WT1 is a mediator of the testicular differentiation at the donor splice site of intron 9 of the WT1 gene, which results in a decrease in the ratio of +KTS/−KTS isoforms. The observations suggest that Denys-Drash and Frasier syndromes are part of a continuous spectrum rather than separate entities. Patients with isolated genital anomalies alone rarely have WT1 mutations.

SF1

Steroidogenic factor 1 (SF1), an orphan nuclear receptor, was first identified as a transcriptional regulatory element that interacted with the promoter elements in the proximal regions of several steroid (cytochrome P450) hydroxylases. Because it was initially discovered in steroidogenic cells, it was named steroidogenic factor 1 (SF1) or adrenal 4-binding protein (Ad4BP). SF1 is a homologue of the Drosophila fushi tarazu factor 1 (FTZ-F1). The SF1 gene is located on chromosome 9p33. SF1 has two zinc finger motifs with a DNA-binding domain in its amino-terminal region and a lig-and-binding domain at its carboxy-terminal end (Fig. 1-18). At the carboxy-terminal end of the DNA-binding domain an 8-amino acid residues (the A box) recognizes specific nucleotides 5' to the DNA response element AGGTCA of the target gene. SF1 has a carboxy-terminal AF-2 trans-activation domain similar to other ligand-inducible nuclear receptors. SF1 is expressed early in the development of the adrenals, gonads, hypothalamus, and pituitary gonadotropes. Its critical role in gonadal and adrenal development was assessed from the SF1 "knockout" mouse, which is born without gonads, resulting in female differentiation in XY males and adrenal aplasia. SF1 "knockout" mice also have impaired pituitary expression of gonadotropins and agenesis of the ventromedial hypothalamic nuclei. SF1 regulates gonadal and adrenal gonadotropinogenesis and the genes encoding the adrenocorticotropic hormone (ACTH) receptor, the steroidogenic acute regulatory protein ( STAR), the luteinizing hormone (LH)-subunit, the -subunit of glycoproteins, the high-density lipoprotein receptor SR-B1, and AMH. Furthermore, SF1/SF1 is a mediator of the development of the adrenals and gonads as well as male sexual differentiation (fetal testicular androgens and AMH) and steroid synthesis in the ovary.

SF1 is expressed at 32 days post ovulation in the gonadal ridge of the human embryo and by 26 days in the embryo. At this stage of the indifferent gonad, expression of SF1 is not sexually dimorphic. After differentiation of the definitive gonad, SF1 expression remains diffusely distributed in the human ovary (in contrast to its striking down-regulation in the mouse ovary). Similar to the mouse, SF1 is widely expressed in the human developing testis and is localized mainly to the Sertoli cells of the developing sex cords of the gonad. During embryonic gestation, SF1 expression is greater over the testis than the sex cord. In the human as well as the mouse, SF1 plays a critical role in gonadal development and function.

SF1 regulates the transcription of a variety of target genes, but unlike many nuclear receptors that bind to DNA as homodimers or heterodimers, SF1 binds specific gene promoters as a monomer. For example, WT1 and SF1 can associate to increase the expression of genes driven by SF1. SF1 bound to a consensus-binding site on Dax1 appears to be necessary for Dax1 expression in the mouse embryonic gonad. Several factors have been identified that regulate SF1 expression, including ubiquitously expressed E-box proteins (upstream regulatory factors 1 and 2) and Sox9. A specific Sox-binding site (AAGACG) is present in the proximal SF1 promoter. Mutations of this site resulted in down-regulation of SF1 promoter activity in embryonic and postnatal Sertoli cells. The interaction of SOX9 and SF1 appears to have a role in the greater expression of SF1 in the fetal testes than in the ovaries in the human and mouse. The complex interactions of these genes proteins, co-localized in Sertoli cells and temporally expressed in relationship to one another, have an important but incompletely understood role in the sex determination and differentiation cascade.

SF1 mutations are rare; three examples are known. The first SF1 mutation was reported in an XY sex-reversed "female" who had primary adrenal insufficiency, persistent müllerian structures, and streak gonads owing to a heterozygous loss of function missense mutation (Gly35Glu) in an amino acid in the first zinc finger of the DNA-binding domain of SF1 (see Fig. 23-18). The features of this patient as predicted in the ninth edition of this chapter, were similar to those of the SF1 "knockout" mouse. A prepubertal XX female with adrenal insufficiency and "putatively" normal ovarian differentiation had a heterozygous mutation, a GT transversion in exon 4 that resulted in Arg255Leu substitution in the ligand-binding region of SF1 and a transcriptionally inactive SF1 protein. A third patient, an infant, had XY sex reversal and adrenal insufficiency owing to a homozygous missense mutation (Arg52Gln) in the A-box of SF1, a region that functions as a secondary response element. In contrast to the other two patients, family members heterozygous for the mutation were asymptomatic. Functional assay expression of this mutation showed partial loss of binding and trans-activation, a less severe than first described heterozygous mutation. This case suggests that,
as with other genes in the sex determination and differentiation cascade, gonadal differentiation and function is highly gene dosage dependent.

**Chromosome 9p and DMRT1**

The human 9p- syndrome is characterized by varying degrees of XY sex reversal, mental retardation, and craniofacial abnormalities. Quite likely it is a contiguous gene syndrome because small terminal deletions of 9p are associated only with XY sex reversal. The proximal boundary of the sex reversal region is 9p24.3, 30 kb upstream from the start of a testis-specific gene DMRT1 (double sex, mab3, related transcription factor 1). DMRT1 encodes a protein homologous to the double sex gene in *Drosophila* and the mab3 gene of *Caenorhabditis elegans*; the latter genes are involved in the sex determination cascade in these organisms. Two other DMRT genes, DMRT3 and DMRT2, have been found proximal to DMRT1 on the short arm of chromosome 9.

DMRT1, 120 kb long, has five exons and encodes a protein of 373 amino acids that is expressed in the human embryonic gonad at 5 to 8 weeks' gestation. A male-specific pattern of expression is present in the genital ridge and the developing Sertoli cells by 7 weeks' gestation. The fact that DMRT1 is conserved across the phylogeny in sex-determining cascades, that it has a male-specific expression in the mouse and human sex determination, and that it is in close proximity to the minimal deleted 9p region, which causes XY sex reversal, make it a strong candidate gene in the human sex determination cascade. Gene dosage appears to be critical: haploinsufficiency of DMRT1 results in sex reversal. No point mutations in this gene have been described as yet in the human. It is of note that gene disruption in the mouse results in normal male development with absent testes postnatally similar to the vanishing testes syndrome in the human, but it had no effect on ovarian development and function. Nonetheless, the role of DMRT1 in ovarian development and function in the human is uncertain. In two 46,XX females with a distal deletion of 9p, Ogata and co-workers found evidence of mild gonadal dysgenesis. Accordingly, DMRT1 has a role in the formation of the indifferent gonad and haploinsufficiency can result in a spectrum of gonadal dysfunction in both males and females.

**Chromosome 10q**

Terminal deletions of chromosome 10q distal to 10q25 are frequently associated with urogenital anomalies, including renal hypoplasia, cryptorchidism, micropenis, hypospadias, hypoplastic labia major, and, rarely, complete XY sex reversal. A candidate gene on distal 10q involved in gonadogenesis, although suspected clinically, has not been identified.

**ATRX**

The -thalassemia mental retardation syndrome is an X-linked disorder characterized by the association of -thalassemia, severe mental retardation, and a variety of developmental anomalies, including genital abnormalities in 80% of affected 46,XY individuals. The ATRX gene (synonyms: XH2, XNF) is located at Xq13.3, spans about 300 kb, contains 35 exons, and undergoes X inactivation. It encodes a calcium-binding protein that is expressed in the early fetal mouse testes. The amino-terminal domain of the ATRX protein contains a zinc finger motif, the central portion has the ATPase and helicase motifs, and in the carboxyl-terminal domain a "P" box involved in transcriptional regulation and a "Q" box is thought to be involved in protein interaction. The spectrum of genital anomalies ranges from undescended testes to hypospadias, micropenis, and varying degrees of XY sex reversal. Mutations in affected males that lead to truncation of ATRX with loss of the carboxyl terminal region (the helicase domain) usually result in genital abnormalities. The spectrum of genital anomalies ranges from undescended testes to hypospadias, micropenis, and varying degrees of XY sex reversal. Mutations in affected males that lead to truncation of ATRX with loss of the carboxyl terminal region (the helicase domain) usually result in genital abnormalities. A mutation in ATRX is associated with XY gonadal dysgenesis, dysmorphic features, and mental retardation similar to those seen in the -thalassemia mental retardation syndrome but with no evidence of -thalassemia.

**Other Genes Involved in Gonadal Determination and Differentiation**

Lhx (LIM1) and 9 are members of a subfamily of homeobox genes involved in developmental processes. Lim homeobox proteins are characterized by two specialized zinc fingers located in the amino-terminal region of the protein followed by a DNA-binding homeobox domain. Lim1, the mouse homologue of human LIM1, is involved in the early stages of gonadal development. It is initially expressed in the intermediate mesoderm and nephrogenic cords. Mice homozygous for targeted deletions of Lim1 fail to develop kidneys, gonads, and anterior head structures. LIM1 mutations have not yet been described in humans but may represent a cause of gonadal agenesis or dysgenesis associated with cranial abnormalities.

Targeted disruption of the Lhx9 gene, which is expressed in the brain, limb buds, and urogenital ridge, leads to gonadal agenesis and no other abnormalities. Because SRY expression is reduced in the sex determination cascade of Lhx9-deficient mice, it has been suggested that Lhx9 may function upstream of SRY in the cascade. The human homologue of Lhx9, maps to chromosome region 1q25-q31 and is expressed at the time of gonad formation. A mutation in LHX9 was not detected in 41 patients with gonadal dysgenesis or dysgenesis including two "sisters." 

Emx2 is a mouse gene homologous to the *Drosophila* empty spiracles (*Em's*) gene involved in head morphogenesis. It is expressed early in the developing urogenital system and sex determination cascade of the mouse. Emx2 null mice have developmental abnormalities of the brain as well as absent kidneys, ureters, gonads, and genital tracts. Emx2 maps to chromosome region 10q26.1, and de novo heterozygous mutations have been found in humans with schizophrenia. However, no genital abnormalities have been observed.

m33, a mouse homologue of the polycomb genes in *Drosophila* may act by modifying higher-order chromatin structure to repress developmentally regulated genes. XY mice with null mutations of this gene exhibit male-to-female sex reversal. Abnormalities in the gonads of m33 mutant XY and XX mice were noted, suggesting that this gene functions early in the sex determination network upstream of SRY.

Desert hedgehog (DHH), a member of the hedgehog family of signaling proteins, which includes sonic hedgehog and Indian hedgehog, is the only known mammalian hedgehog protein expressed in the mouse embryonic gonad in Sertoli cells and the interstitium. DHH and Patched 1, the receptor for DHH, signaling have a critical role in the differentiation of Leydig cells by up-regulating SFR1, which in turn activates cholesterol side-chain cleavage. In mice, in which mutations in mice resulted in feminized 46,XY mice with decreased spermatogenesis. The human homologue, DHH, maps to human chromosome bands 12q12-q13.1. A homoygous mutation in DHH is reported in a 46,XY patient with partial gonadal dysgenesis and a polynuropathy; his "normal" father was heterozygous for this mutation.

Multiple developmental processes are regulated by fibroblast growth factors. The wide expression of Fgf9 prompted targeted deletion experiments in mice. An unexpected finding was sex reversal in homozygous Fgf9 null males, which usually but not always was complete. Fgf9 acts downstream of Sry and is involved in migration of the mesonephric cells into the embryonic testis and in Sertoli cell development. The mutation is lethal owing to severe pulmonary hypoplasia. A defect in Fgf9 signaling in the human is a potential cause of XY gonadal dysgenesis and sex reversal.

The felsalinC gene encodes a calcium-binding protein that is expressed in the early fetal mouse testes. Testatin, a cystatin-related gene, is expressed in fetal gonads and adult testes in a pattern similar to Sox9. A role, if any, of these two genes in human gonadogenesis is not known.
Genes and Ovarian Organogenesis

As early as 1958, it was suggested that two intact X chromosomes are required in the human for differentiation of the indifferent gonad into a normal functional ovary, in contrast to the mouse and some other mammals, in which an XO sex chromosome constitution does not prevent the development of a fertile ovary (although it leads to accelerated atresia of ovarian follicles). In 45,X individuals and in those with deletions of the short (Xp) or long (Xq) arm of the X chromosome, ovarian development commences in utero; however, oocytes usually do not survive meiosis, and folliculogenesis fails to occur or is defective. This results in loss of germ cells, oocyte degeneration, and, secondarily, gonadal dysgenesis (streak gonads). Both X chromosomes appear to be active in the primordial germ cell and oocyte from the onset of meiosis to oulation. In the human the functional integrity of genes on both X chromosomes is necessary for maintenance but not organogenesis of the ovary. The occurrence of familial 46,XX gonadal dysgenesis, which is transmitted as an autosomal recessive trait, suggests that autosomal genes, expressed through direct or indirect actions on the germ cell, are essential for ovarian organogenesis. A homozygous inactivating mutation in the gene encoding the follicle-stimulating hormone (FSH) receptor leads to ovarian dysgenesis and hypergonadotropic hypogonadism in family cohorts, as reported in Finland [see Chap. 24]. Other possible causes of familial 46,XX gonadal dysgenesis include a mutant autosomal gene that leads to a defect in development of the rete ovarii or in the synthesis or action of the putative meiosis stimulating factor.

In contrast to testis determination, there is a paucity of knowledge about confirmed and candidate genes involved in ovarian determination. Whereas DAX1 is an anti-testes gene, it does not have a critical role in ovarian differentiation. The WNT7A family of genes has been implicated in the development of ovaries. Wnt4, Wnt7a, and Wnt5a

Wnt4 is a member of a large family of WNT signaling glycoprotein molecules (WNT stands for wingless-type mouse mammary tumor virus integration site); their expression pattern during embryogenesis suggests that they are involved in determination of cell fates, cell-cell communication, and cellular proliferation. Wnt4 expression is maintained in the ovary, whereas Wnt7a is expressed in the genital primordium of the mouse from 10.5 to 14.5 days post conception and thereafter in the mesenchyme of the genital tract. Wnt5a is expressed in the genital primordium of the mouse from 10.5 to 14.5 days post conception and thereafter in the mesenchyme of the genital tract. Wnt5a mutant mice have a defective genital tubercle and lack external genitalia. Humans with mutations in WNT7A or WNT5A have not been described as yet. However, these genes play an important role in müllerian duct and clitoral/penile development in the mouse. Another candidate gene for ovarian determination and function, FOXL2, the winged helix/forkhead transcription factor maps to chromosome 3q23. Patients with blepharophimosis/ptosis/epicanthus inversus (BPE) syndrome have ovarian failure, which varies from streak gonads with no function to premature ovarian failure, and have a mutation that results in haploinsufficiency of FOXL2 expression. The polled intersex syndrome (PIS) in goats consists of a null mutant of Sry which leads to testis development in an XY female. In both sexes mutants fail to develop müllerian ducts and this occurs before AMH is active in regression of müllerian ducts in the male. The Wolffian ducts are stabilized in the female null mutant because of differentiation of Leydig-like interstitial cells in the ovary; the external genitalia, however, are not masculinized. Wnt4 is required for initial müllerian duct development in both sexes, and in the female to suppress Leydig cell differentiation in the ovary.

Another member of the WNT family, Wnt7a is required for complete differentiation of the müllerian ducts into the uterus, fallopian tubes, and upper part of the vagina, possibly by signaling through epithelial-mesenchymal interactions. Mutant male mice have persistent müllerian ducts and are infertile; mutant female mice develop functional ovaries but are sterile because of abnormal development of the oviducts and uterus. The abnormal development of the müllerian structures has similarities to that observed in the offspring of women treated with diethylstilbestrol during pregnancy.

Wnt5a is expressed in the genital primordium of the mouse from 10.5 to 14.5 days post conception and thereafter in the mesenchyme of the genital tract. Wnt5a mutant mice have a defective genital tubercle and lack external genitalia. Humans with mutations in WNT7A or WNT5A have not been described as yet. However, these genes play an important role in müllerian duct and clitoral/penile development in the mouse. Another candidate gene for ovarian determination and function, FOXL2, the winged helix/forkhead transcription factor maps to chromosome 3q23. Patients with blepharophimosis/ptosis/epicanthus inversus (BPE) syndrome have ovarian failure, which varies from streak gonads with no function to premature ovarian failure, and have a mutation that results in haploinsufficiency of FOXL2 expression. The polled intersex syndrome (PIS) in goats consists of a null mutant of Sry which leads to testis development in an XY female. In both sexes mutants fail to develop müllerian ducts and this occurs before AMH is active in regression of müllerian ducts in the male. The Wolffian ducts are stabilized in the female null mutant because of differentiation of Leydig-like interstitial cells in the ovary; the external genitalia, however, are not masculinized. Wnt4 is required for initial müllerian duct development in both sexes, and in the female to suppress Leydig cell differentiation in the ovary.

Wnt4 is an essential signal in ovarian development and function.

Another member of the WNT family, Wnt7a is required for complete differentiation of the müllerian ducts into the uterus, fallopian tubes, and upper part of the vagina, possibly by signaling through epithelial-mesenchymal interactions. Mutant male mice have persistent müllerian ducts and are infertile; mutant female mice develop functional ovaries but are sterile because of abnormal development of the oviducts and uterus. The abnormal development of the müllerian structures has similarities to that observed in the offspring of women treated with diethylstilbestrol during pregnancy.

Wnt5a is expressed in the genital primordium of the mouse from 10.5 to 14.5 days post conception and thereafter in the mesenchyme of the genital tract. Wnt5a mutant mice have a defective genital tubercle and lack external genitalia. Humans with mutations in WNT7A or WNT5A have not been described as yet. However, these genes play an important role in müllerian duct and clitoral/penile development in the mouse. Another candidate gene for ovarian determination and function, FOXL2, the winged helix/forkhead transcription factor maps to chromosome 3q23. Patients with blepharophimosis/ptosis/epicanthus inversus (BPE) syndrome have ovarian failure, which varies from streak gonads with no function to premature ovarian failure, and have a mutation that results in haploinsufficiency of FOXL2 expression. The polled intersex syndrome (PIS) in goats consists of a null mutant of Sry which leads to testis development in an XY female. In both sexes mutants fail to develop müllerian ducts and this occurs before AMH is active in regression of müllerian ducts in the male. The Wolffian ducts are stabilized in the female null mutant because of differentiation of Leydig-like interstitial cells in the ovary; the external genitalia, however, are not masculinized. Wnt4 is required for initial müllerian duct development in both sexes, and in the female to suppress Leydig cell differentiation in the ovary.

The expression of Wnt4 is similar to Dax1, both persisting in the ovary during fetal development. Overexpression of Dax1 in some human sex-reversed XY females with a Xp21 duplication suggests that Wnt4 may function in a similar manner as an anti-testis gene. The human WNT4 gene is in chromosome 1p35. An XY female had duplication of 1p31-p35 and over-expression of WNT7A. The mechanism of sex reversal due, at least in part, to WNT4 induced up-regulation of DAX1. The sexually dimorphic expression of WNT7A in the human, as well as the occurrence in the Wnt4 null mice of structurally and functionally masculinized ovaries, suggests that WNT4 is an essential signal in ovarian development and function.

Another member of the WNT family, Wnt7a is required for complete differentiation of the müllerian ducts into the uterus, fallopian tubes, and upper part of the vagina, possibly by signaling through epithelial-mesenchymal interactions. Mutant male mice have persistent müllerian ducts and are infertile; mutant female mice develop functional ovaries but are sterile because of abnormal development of the oviducts and uterus. The abnormal development of the müllerian structures has similarities to that observed in the offspring of women treated with diethylstilbestrol during pregnancy.

Wnt5a is expressed in the genital primordium of the mouse from 10.5 to 14.5 days post conception and thereafter in the mesenchyme of the genital tract. Wnt5a mutant mice have a defective genital tubercle and lack external genitalia. Humans with mutations in WNT7A or WNT5A have not been described as yet. However, these genes play an important role in müllerian duct and clitoral/penile development in the mouse. Another candidate gene for ovarian determination and function, FOXL2, the winged helix/forkhead transcription factor maps to chromosome 3q23. Patients with blepharophimosis/ptosis/epicanthus inversus (BPE) syndrome have ovarian failure, which varies from streak gonads with no function to premature ovarian failure, and have a mutation that results in haploinsufficiency of FOXL2 expression. The polled intersex syndrome (PIS) in goats consists of a null mutant of Sry which leads to testis development in an XY female. In both sexes mutants fail to develop müllerian ducts and this occurs before AMH is active in regression of müllerian ducts in the male. The Wolffian ducts are stabilized in the female null mutant because of differentiation of Leydig-like interstitial cells in the ovary; the external genitalia, however, are not masculinized. Wnt4 is required for initial müllerian duct development in both sexes, and in the female to suppress Leydig cell differentiation in the ovary.

Wnt4 is an essential signal in ovarian development and function.
Gametogenesis

Origin of Primordial Germ Cells

In humans as well as mice, primordial germ cells, the progenitors of the male and female gametes, originate from pluripotential cells in the epiblast. The specification of mammalian germ cells from proximal epiblast cells involves the action of at least three "bone morphogenetic factors": Bmp2, Bmp4, and Bmp8b and Oct4 and Cdx2. These "bone morphogenetic factors" are members of the transforming growth factor- superfamily, which act through the transcription factors Smad-a and Smad-b, whose targets are not yet known. In the 24-day-old human embryo, the differentiated germ cells recognized by their histologic appearance and the presence of alkaline phosphatase are located in the dorsal endoderm of the yolk sac close to the allantoic evagination. From this site, the cells increasing in number by mitosis, migrate during the fourth and fifth weeks of gestation in contact with each other through cell-cell and cell-matrix adhesive molecules to the hindgut and then through the dorsal mesentery to the genital ridge. In the absence of primordial germ cells, the gonad ridges in the female remain undeveloped. In contrast, germ cells are not required for the differentiation of a testis. Germ cells that fail to reach the gonads by the time of sex differentiation usually disappear, although they may persist outside the gonad and give rise to germ cell neoplasms.

Spermatogenesis

During early testicular differentiation the primordial germ cells become distributed throughout the primitive seminiferous tubules as progenitors of spermatagonia. A series of mitotic divisions occurs, followed thereafter by inhibition of entry into meiotic prophase presumably by a putative local factor secreted by Sertoli cells until the transient postnatal increase in mitosis, and later the initiation of full spermatogenesis late in the prepubescent period. Furthermore, it appears that both male and female germ cells are programmed to enter meiosis unless they are inhibited by a putative signal from the testicular somatic cells. With the onset of puberty, the basement membrane of the spermatogenic tubule becomes lined by proliferating spermatogonia that arise by the mitotic division of the spermatogonia. A glial cell line neutrophilic factor (GDNF) secreted by Sertoli cells appears to play a role in the proliferation and differentiation of spermatogonia.

Figure 22-20 Types of cell division. A female somatic cell undergoing mitosis is represented. At the metaphase plate are two X chromosomes and two homologous autosomes of group 21 to 22. Division occurs through the centromere, giving rise to two daughter cells of identical chromosomal composition. Replication of each arm into two chromatids takes place while the chromosomes are extended and before the next metaphase. The first meiotic division involves pairing of homologous chromosomes. The centromere does not divide in this cell division. It is by chance whether the maternal (X0) or paternal (X0) member of each pair goes to the respective daughter cells. During the complex prophase of first meiotic division, multiple exchanges are formed between the chromosomes of each pair, facilitating exchanges of chromosomal segments (crossing-over) between them. During the second meiotic division, the centromere again divides, giving rise to daughter cells identical to the parent cell. This division more nearly resembles mitosis than the first meiotic division. Nondisjunction can take place in mitosis or in the first or second meiotic division; representative examples are illustrated.

adult male are continually being renewed and undergoing maturation. Heller and Clermont estimated that the complete cycle in adult males from spermatagonium to mature sperm requires 74 ± 5 days.

The X chromosome undergoes inactivation in pachytene spermatocytes during meiosis. The condensed sex chromosomes form a sex vesicle or XY body. XY pairing must occur during normal meiosis. X inactivation may be involved in the heterochromatinization of those regions of the X chromosome that are similar to but not homologous to regions on the Y chromosome.

Oogenesis

Female germ cells pursue a different course. During ovarian differentiation, the primary germ cells undergo vigorous mitotic replication and successive differentiation
into oogonia. When mitotic division ceases and the cells enter meiosis, they are then termed oocytes. Meiotic oocytes are critical for differentiation of pre-follicular cells into follicular cells. The absence or loss of meiotic oocytes leads to degeneration of the prefollicular cells. The period of oogonial proliferation results in a peak population of about 6 million to 7 million germ cells in the two ovaries at 5 months' gestational age, including oogonia, oocytes in various stages of prophase, and degenerating oocytes. Oocytes degenerate at different stages of meiosis. Only 5% of the peak number of germ cells in the fetal ovary reach the diplotene stage. Formation of oogonia from primary germ cells ceases by the seventh month of gestation. Some oocytes remain in undifferentiated nests, whereas others form primordial follicles. A primordial follicle is formed when presumptive granulosa cells surround the diplotene (meiotic) oocyte and an intact basal lamina encloses this unit. If the oocyte is not enclosed in a follicle, it degenerates. The number of primordial follicles in the ovary is maximal at birth, and the number thereafter diminishes. In the germ cells that survive, the oocyte is arrested at late prophase of the first meiotic division (diplotene state). Oocytes remain in the prophase of the first meiotic division from fetal life until puberty when some unknown stimulus allows them to progress and eventually ovulation occurs. Before ovulation, the first polar body is extruded, thus completing the first meiotic division. The haploid secondary oocyte immediately begins a second meiotic division but remains in metaphase and does not extrude the second polar body until the ovum is penetrated by a sperm. The triploidy that is common in spontaneously aborted fetuses may be caused either by failure of extrusion of the second polar body (polygyny) or by double fertilization (polyspermy). The long life span of female germ cells, in contrast to those of the male, may explain the increased prevalence of certain chromosomal anomalies with advanced maternal age (see section on aneuploidy).
The gonads of both sexes develop from anlagen located on the mesodermal border of the urogenital ridge, adjacent to the kidney and the primitive adrenal. Until the 12-mm stage (approximately 42 days of gestation), the gonads of the male and female are indistinguishable on morphologic grounds and could potentially differentiate either as testes or as ovaries. The close ontogenic and anatomic relation between gonadal and adrenal cells at this early stage is noteworthy, because, as differentiation proceeds, nests of adrenal cells frequently separate with the gonad and are found as adrenal rests in the hilum of the mature ovary or testis. Such rests may become a problem in patients with long-standing untreated congenital adrenal hyperplasia (CAH). Adrenal cell rests in testes, for example, may enlarge under persistent corticotropin (ACTH, adrenocorticotropic) stimulation and be mistaken for testicular tumors or true testicular enlargement (see Chap. 24).

The primitive undifferentiated gonad is made up of four cell lineages, germ cells, connective tissue cells, steroid-producing cells, and supporting cells. They are derived from proliferation of the mesodermal coelomic epithelium, the mesenchymal cell mass in the urogenital ridge, mesonephric elements, and the large alkanine phosphatase-containing primordial germ cells that have migrated from the posterior endoderm of the yolk sac through the mesenchyme of the mesentery to the gonad. According to Witschi and co-workers, the number of migrating germ cells in the human embryo is 700 to 1300, and by the eighth week of embryogenesis about 600,000 germ cells are present, which later become either oogonia or spermatogonia. Lack of germ cells is incompatible with ovarian differentiation but does not prevent testicular morphogenesis. However, in the mouse, genes such as the (S1) gene, which encodes a peptide growth factor (SCF, stem cell factor), the proto-oncogene c-kit (also known as white spotting [W]), which encodes a tyrosine kinase receptor in the plasma cell membrane (a receptor for S1), and caderins affect the proliferation, mobility, and migration of primordial germ cells to the urogenital ridge. SCF, in its transmembrane second form, is expressed in human Sertoli cells and acts as an adhesion protein, binding germ cells. The precursor of the Sertoli cell of the testis and its counterpart in the ovary, the granulosa cell, originate from the coelomic epithelium.

There is a striking sexual dimorphism in the timing of gonadal differentiation. Under the influence of the testis-determining genes, testis organization begins at about 45 days of gestation (6 to 7 wk); the testis develops more rapidly than the ovary. The ovary does not emerge from the indifferent stage until 3 months of gestation, when the earliest sign appears: the beginning of meiosis, as evidenced by the maturation of oogonia into oocytes.

In the past it was believed that the testis is derived primarily from the medullary portion of the primitive gonad and the ovary from the cortical portion. According to this concept, the testis and ovary are not strictly homologous. Witschi and co-workers suggested that in genetic males the medullary portion secretes an inductor substance that stimulates development of seminiferous tubules and inhibits cortical development; conversely, the cortex of genetic females, the coelomic epithelium, was thought to secrete an inductor substance that inhibits testicular development and results in ovarian dominance. The proposal was that the differentiation of the primordial gonad was regionalized. Jost and co-workers, among others, called into question the histologic descriptions of gonadal differentiation that served as the basis for these theories. After careful examination of early embryos, Jost and Jirasek concluded that it is not possible to identify primary sex cords as such before the 15-mm stage and that the primitive gonad is truly bipotential. At about 45 days morphologic sex dimorphism is evident in the intermediate mesoderm of the genital ridge when epithelial cords derived from the coelomic epithelium, the gonadal blastema including stromal cells of mesonephric origin, and the primordial germ cells, antecedents of the seminiferous tubules, are apparent in the male. The onset of testicular differentiation is marked by the EN/dependent differentiation of the Sertoli cell, the first cell type to differentiate, and by the subsequent incorporation of the germ cells into the primitive seminiferous cords, when the germ cells of the gonad are suppressed and differentiation is arrested at the primitive spermatogenital stage. The XY gonadal somatic cells essentially block the development of primordial germ cell (pre-spermatogonia) from advancing to meiotic prophase (a cardinal feature of oocyte differentiation) and promote their spermatogonial differentiation.

While the somatic cells of the primordial gonad are bipotential, SRY expression directs the delaminated and proliferating somatic cells derived from the coelomic epithelium to differentiate into Sertoli cells that first cell to differentiate (whether Sertoli cells are derived as well from other progenitors) is not yet resolved. The differentiation of Sertoli cells drives development of the testis. This directed process includes proliferation of the somatic cells, which is more rapid and extensive in the XY gonad than the XX gonad; migration of stromal mesonephric cells and endothelial cells from the adjacent mesonephros into the gonad, which occurs only in XY gonads and is dependent on the expression of SRY; testis cord development and Sertoli cell differentiation, the initiation of which in the mouse, appears to be dependent on the migration of mesonephric cells. The latter include the progenitors of Leydig cells and peritubular myoid cells; and the differentiation of a testis-specific vasculature. The stromal cells migrating from the mesonephros have the capacity to differentiate into peritubular myoid cells and surround the developing testis cords to contribute along with the coelomic epithelium to the differentiation of Leydig cells and the testis-specific architecture. In sum, the developing testis is composed of (1) germ cells, the progenitors of future spermatagonia that migrated into the genital ridge from an extragonadal source; (2) Sertoli cells derived from the coelomic epithelium, which with their envelopment of the primordial germ cells leads to the formation of testicular cords containing (3) peritubular myoid cells of mesonephric origin that surround the testicular cords; and (4) the interstitial Leydig cells, which originate from the mesonephros and coelomic epithelium. This construct is largely derived from lineage tracing, studies in the mouse embryo by Capel and associates that included the use of differentiation and proliferation markers, and in the human, from ultrastructural studies of the early fetal gonad.

Burgoyne and associates proposed from studies of XX:XY chimeric mice that the Sry gene acts autonomously to induce Sertoli cell differentiation, which then mediates further testicular differentiation. After testicular differentiation occurs (at 43 to 50 days' gestational age), the male fetus can also be recognized by beginning regression of the primordial müllerian ducts (30-mm stage, about 60 days' gestation) and by differentiation of male external genitalia (45-mm stage, 65 to 77 days' gestation). An early endocrine function of the testis is the secretion by the Sertoli cells of AMH, a homodimeric glycoprotein that functions as a paracrine secretion; it passes by diffusion to the paired müllerian ducts and induces their dissolution by apoptosis. The versatile Sertoli cell also secretes inhibin, nurtures the germ cells, expresses stem cell factor, synthesizes an androgen-binding protein, and prevents meiosis. Both fetal Sertoli and germ cells exhibit apoptosis as well as proliferation during gestation.

Leydig cells are first found in 32- to 35-mm fetuses (about 60 days' gestation). After differentiation of the primitive testicular cords, they rapidly proliferate during the third month and the first half of the fourth month; during this period the interstitial spaces between the seminiferous tubules are crowded with Leydig cells. The onset of testosterone biosynthesis occurs at about 9 weeks. Human chorionic gonadotropin/luteinizing hormone (hCG/LH) receptors are present.
critical period of male sex differentiation, but whether testosterone synthesis at its onset is HCG dependent is unclear. The question of whether HCG is required to initiate testosterone secretion in the human is complicated further by the presence of HCG-like material in the fetal testis. The pattern of testosterone secretion early in gestation follows that of HCG. The number of Leydig cells decreases after 18 weeks, by apoptosis and by dedifferentiation, and few cells show Leydig cell characteristics in the interstitium of the testis at birth. However, a low level of testosterone secretion is maintained after 15 weeks of gestation under the control of fetal pituitary LH and HCG. Fetal pituitary LH is necessary for the normal growth of the differentiated penis and scrotum during the latter half of gestation and for the descent of the testes. The male fetus with anencephaly or congenital hypopituitarism often has hypoplastic male external genitalia and undescended testes containing a decreased number of Leydig cells. Fetal Leydig cells differ from adult Leydig cells in their morphology, their regulatory mechanisms, and their lack of desensitization to high doses of HCG and LH.

Figure 22-22 correlates the pattern of testosterone, HCG, and fetal pituitary FSH and LH concentrations during gestation with the histologic changes in the fetal testis. In surm, organogenesis of the testis involves successive differentiation of the Sertoli cell and the seminiferous cords with envelopment of the extragonadally derived germ cells by Sertoli cells, development of the tunica albuginea, appearance of Leydig cells, and differentiation of the mesonephric tubules into the ductuli efferentes, which connect the seminiferous tubules and rete network with the epididymis to provide the pathway for sperm transport into the ejaculatory duct system.

Ovary

In the absence of testis-determining genes, the gonadal primordium has an inherent tendency to develop as an ovary, provided that germ cells are present and survive. The indifferent stage persists in the female fetus weeks after testis organogenesis begins. There is, however, continued proliferation of the coelomic epithelium and primordial germ cells, which gradually enlarge and become oogonia. Despite the discordance in the histologic appearance of the primordial testis and ovary, George and Wilson noted the simultaneous development at 8 weeks of gestation of the capacity of the fetal testis to synthesize testosterone and of the fetal “ovary” to synthesize estradiol when incubated with C19-steroid precursors. In contrast to the fetal adrenal gland, the gonads of both male and female fetuses have 3-hydroxysteroid dehydrogenase 2 (3-HSD) activity at this stage; however, the activity of this enzyme is more than 50-fold greater in the fetal testis. Testosterone is synthesized by the fetal Leydig cell, but the site of the meager synthesis of estradiol in the primordial ovary is not known. Gongos and Hobet identified interstitial cells in the ovarian primordium at about 12 weeks of gestation that have the ultrastructural characteristics of steroidogenic cells and that may be a site of estrogen synthesis. The human fetus is bathed in estrogens of placental origin, but the fetal ovary does not contribute significantly to circulating estrogens, which in the fetus are almost exclusively of placental origin. The ovary has no documented role in differentiation of the female genital tract.

During the ninth week the rete ovarii arise from the hilar mesonephric tubules and infract the gonad as a syncytium of tubules and cords. At about the 11th to 12th week (80-mm stage), long after differentiation of the testis in the male fetus, germ cells in the ovary begin to enter meiotic prophase, which characterizes the transition of oogonia into oocytes and marks the onset of ovarian differentiation. The oogonia in the central part of the ovary are the first to come in contact with the rete ovarii and the first to enter meiosis but not beyond the diplotene phase. According to Byuskov, the rete secretes a meiosis-inducing substance; meiosis activity steroid (MAS), a precursor of cholesterol, is suggested as the factor initiating meiosis in the fetal ovary. The initial stages of ovarian organogenesis and the formation of primordial follicles do not involve placental or fetal pituitary gonadotropins or their receptors (see reference ). After the diplotene stage, the first meiotic division stops and the chromosomes enter a resting stage until meiosis is completed just before ovulation in the adult. The formation of primordial follicles (in which the oocyte is enveloped by a single layer of flat granulosa cells that share a common lineage with Sertoli cells) reaches a maximum during the 20th to the 25th week of gestation. In contrast to the fetal testis, FSH and LH/HCG receptors are not detected in the human fetal ovary between 8 and 16 weeks’ gestation; the fetal ovary becomes responsive to FSH stimulation later in gestation (see reference ).

The growth, development, and maintenance of follicles appear to be regulated in late gestation by fetal pituitary gonadotropins, mainly FSH. The sequence and timing of events in gonadal organogenesis and their relation to the differentiation of male and female somatic sexual characteristics are shown in Figure 22-24.
Differentiation of the Genital Ducts

At the seventh week of intrauterine life, the fetus is equipped with both male and female genital ducts derived from the mesonephros. The müllerian ducts serve as the analog of the uterus and fallopian tubes, whereas the mesonephric or wolffian ducts have the potential to differentiate further into the epididymis, vas deferens (ejaculatory ducts of the male), and seminal vesicles. During the third fetal month, either the müllerian or wolffian ducts complete their development, and involution occurs simultaneously in the opposite structures (Fig. 22-25).

Jost and Josso and associates demonstrated that secretions from the fetal testis play a decisive role in determining the direction of genital duct development. In the presence of functional testes, the müllerian structures involute and undergo programmed cell death and the wolffian ducts complete their development; in the absence of testes, the wolffian ducts do not develop and the müllerian structures differentiate (Fig. 22-26). The regression of the müllerian ducts and the stabilization and differentiation of the wolffian ducts are mediated by different secretions of the fetal testes: the glycoprotein AMH secreted by the fetal Sertoli cells, and the steroid testosterone synthesized by the fetal Leydig cells.

Female development is not contingent on the presence of an ovary, because development of the uterus and tubes occurs if no gonad is present. However, the müllerian duct (paramesonephric duct) fails to differentiate in the absence of the mesonephric ducts, which serve as the anlage for both the male urogenital tract and the metanephros (primordial kidney); and therefore renal aplasia is commonly associated with hypoplasia of the fallopian tubes and uterus and vaginal agenesis.

The influence of the fetal testis on duct development is exerted locally and unilaterally; if one testis is removed at an early stage of development, the oviduct develops normally on that side but müllerian regression occurs on the side of the intact testis.

Systemic administration of androgen to an early embryo does not cause regression of müllerian structures. Even when large amounts of androgen are implanted locally in the gonadal region of female fetuses, the müllerian ducts do not atrophy, although the differentiation of the wolffian ducts is stimulated. On the other hand, if a testis is grafted onto an ovary, müllerian regression and wolffian stimulation occur on that side (see Fig. 22-26). For these reasons, Jost proposed that the fetal testis secretes a müllerian duct inhibiting substance that is distinct from ordinary androgens.

Jost and co-workers and Josso and associates studied the influence of the fetal testis on müllerian duct inhibition in organ culture. Direct contact between the testis and the müllerian anlage was not necessary to bring about this inhibition. By separating the testis from the müllerian ducts with dialysis membranes, they concluded that the material secreted from the testis was a protein and not a steroid. They also demonstrated that the human fetal testis, regardless of age, inhibits the müllerian ducts of 14.5-day-old fetal rats in similar organ culture studies and that AMH activity is present in human testes until 8 to 10 years of age.

Using bovine fetal testes in which tubules and interstitial tissue were isolated and assayed separately, they showed that AMH activity is derived from the Sertoli cell, with peak levels occurring at the time of müllerian duct regression (9 to 12 weeks). Thereafter, the levels remain high until birth, after which a steady decline occurs until the pubertal period.

The decline (but not disappearance) in AMH levels observed at puberty in males has been attributed to various factors that include testosterone secretion, meiotic entry, and terminal maturation of the Sertoli cells. AMH is present in the ovarian follicle and is synthesized and secreted by the granulosa cells, but only after birth; whereas plasma concentrations are low in childhood, they increase at puberty. Elevated serum levels of AMH occur in patients with granulosa cell tumors.

AMH secretion by the postnatal ovary does not affect the fallopian tubes and the uterus, because they are apparently insensitive to AMH after 9 to 12 weeks of gestation, the period during which müllerian duct regression usually occurs.

In the freemartin the fetal ovary and the müllerian structures are exposed to AMH before the refractory period. AMH secreted by the fetal Sertoli cells of the male twin passes by means of placental vascular anastomoses to the female twin and results in müllerian regression, ovarian inhibition with loss of germ cells, tunica albuginea formation, and development of seminiferous tubule-like cords.

Studies show that transgenic female mice that persistently express the human AMH gene resemble the bovine freemartin. The AMH transgenic female mice lack müllerian derivatives, and at birth the ovaries have fewer germ cells than normal. During the first 2 weeks of life germ cells are lost, and the somatic cells become organized into seminiferous tubule-like structures that do not persist to adulthood.

In the transgenic male mice, sex differentiation is usually normal, although some males, those that express the highest AMH levels, have incomplete virilization of the external genitalia, incomplete wolffian duct development, and undescended testes secondary to the effect of AMH on Leydig cell maturation and steroid synthesis.

The relevance of these studies to normal sex differentiation is unclear, because the levels of AMH and its continuous secretion are different from those in the normal mouse fetus. Gene knockout of Amh produced normal male mice with normal müllerian duct derivatives that were fertile, indicating that AMH is not required for normal ovarian function, at least in mice.
In Amh-deficient males, as expected, had persistence of the müllerian derivatives. Ninety percent of the Amh-deficient mice were infertile because of interference by the müllerian duct derivatives with the passage of sperm from the epididymis and vas deferens to the urethra. The testes showed normal spermatogenesis but marked Leydig cell hypoplasia, suggesting that AMH may affect Leydig cell proliferation.

The mechanism of action of AMH includes proteinase activated programmed cell death. It causes a gradient of cranial...
During differentiation of the Wolffian ducts to form the epididymides, vasa deferentia, and seminal vesicles, the Wolffian ducts lack the enzyme 5-reductase, which converts testosterone to dihydrotestosterone (DHT). Experimental data and studies in humans with steroid 5-reductase type-2 deficiency provide additional evidence that testosterone (not DHT) mediates the differentiation of the Wolffian ducts. This is in striking contrast to the urogenital sinus and genital tubercle, which express steroid 5-reductase-2 even before the testis has developed the capacity to synthesize testosterone. DHT mediates the masculinization of the urogenital sinus (including formation of the prostate) and external genitalia. Despite this difference in the action of testosterone and DHT on the primordia of the genital tract, the androgen receptor, at least in the rabbit fetus, is the same in the Wolffian duct and in the anlage of the urogenital sinus and external genitalia.

In patients with ambiguous genitalia, male genital ducts are well differentiated only in those who have functional testes and androgen receptors. Females with CAH do not display Wolffian duct differentiation even though their external genitalia may be highly virilized in utero. Patients with asymmetric gonadal differentiation likewise have asymmetric male duct development that correlates with the degree of testicular differentiation on that side.

If the critical role of the testis in male duct development is to provide a high local concentration of testosterone, male duct development would be expected to be deficient, even though testes are present, in patients with severe defects in steroid biosynthesis (e.g., deficiency of steroidogenic acute regulatory protein [STARD]) and in XY patients whose tissues are unresponsive to testosterone (complete androgen resistance syndrome). The epididymides and vasa deferentia of these patients are generally underdeveloped. However, Wolffian duct remnants (and sometimes fully developed duct derivatives) are observed with 17-hydroxysteroid dehydrogenase 3 deficiency, Leydig cell hypoplasia (hCG/LH unresponsiveness) and, indeed, in some patients with complete androgen resistance. During sex differentiation, testosterone and AMH effect their morphogenetic actions on the underlying mesenchymal cells rather than directly affecting the epithelial cells. Action of the hormone-stimulated mesenchyme on the epithelial cells mediates the morphogenesis of the male ducts and regression of the müllerian ducts.

Furthermore, from animal studies, EGF and GH may also play a role in stabilizing Wolffian duct development.
Differentiation of the External Genitalia and Urogenital Sinus

Origin of the External Genitalia

At the eighth fetal week the external genitalia of both sexes are identical and have the capacity to differentiate in either direction. They consist of a urogenital slit bounded by paired urethral folds and, more laterally, by labioscrotal swellings. The urogenital slit is surmounted by a genital tubercle consisting of corpora cavernosa and glans. The mucosa-lined urethral folds may remain separate, in which case they are called labia minora, or they may fuse to form a corpus spongiosum enclosing a phallic urethra. The fleshy labioscrotal swellings may remain separate to form labia majora or fuse in the midline to form a scrotum and the ventral epididymal covering of the penis. The distinction between a clitoris and a penis is based primarily on size and whether the labia minora fuse to form a corpus spongiosum. The clitoris has two corporeal bodies that are analogous to but smaller than those of the penis. Baskin and colleagues describe in detail the neural innervation and the extensive, pudendal nerve-derived neurovascular network of the fetal clitoris through three-dimensional computer reconstructions. This study has an important bearing on the preservation of sensation and function in the design of clitoral reconstructive surgery.

By the 50-mm crown-rump stage, male and female fetuses can be distinguished by inspection of the external genitalia; in the male, the urethral folds have fused completely in the mid-line to form the cavernous urethra and corpus spongiosum by 12 to 14 weeks of gestation. The penis in the male increases linearly, at about 0.7 mm/week, from 10 weeks to normal term; a 12-fold increase occurs from 0.3 cm at 10 weeks to 3.5 cm at term, a rate of growth about 3.5 times that of the clitoris. The mean stretched length of the penis in full-term infants is 3.5 ± 0.7 cm (SD) and the diameter is 1.1 ± 0.2 cm (SD). The clitoral size in full-term infants is similar in several studies. The mean clitoral length is 0.40 ± 0.012 cm (SD), and the width is 0.033 ± 0.078 cm (SD).

 Origin of the Vagina

The urogenital sinus separates from a common cloaca in early fetal life. There is disagreement about the relative contribution of the müllerian duct and the urogenital sinus to the vagina, but the contact and interaction of the fused müllerian ducts with the urogenital sinus is essential for normal development of the vagina. In normal female development, proliferation of the vesicovaginal septum pushes the vaginal orifice posteriorly so that it acquires a separate external opening; thus no urogenital sinus, as such, is preserved. The lower vagina probably originates from the urogenital sinus. Uroplakins are membrane proteins specific to the urothelial plaque. Examination of human female fetuses of 9 to 18 weeks of gestation showed expression of uroplakins in the urothelium of the urogenital sinus, opening; thus no urogenital sinus, as such, is preserved. The lower vagina probably originates from the urogenital sinus. Uroplakins are membrane proteins specific to the urothelial plaque. Examination of human female fetuses of 9 to 18 weeks of gestation showed expression of uroplakins in the urothelium of the urogenital sinus, including the part that evaginates to form the sinovaginal bulbs. In male development, the vaginal pouch is usually obliterated when the müllerian ducts are resorbed, although a vestigial blind vaginal pouch known as the prostatic utricule can sometimes be demonstrated; it is the site of prostate formation later in development.

The prostate gland and bulbourethral glands of Cowper in the male are outgrowths of the urogenital sinus; their differentiation is mediated by DHT and requires the presence of androgen receptors. In the female, the paraurethral glands of Skene and the vestibular glands of Bartholin have homologous origins.

Mechanism of Androgen Action

The effects of testosterone are varied and tissue specific and reflect the sum of its action and the actions of its conversion products, DHT and estradiol.

TABLE 22-5 -- Homologies Between Male and Female Sexual Structures

<table>
<thead>
<tr>
<th>Male Derivative</th>
<th>Primordial Structure</th>
<th>Female Derivative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gonad</td>
<td>Indifferent gonad derived from</td>
<td>Granulosa cells</td>
</tr>
<tr>
<td>Sertoli cells</td>
<td>Coelomic epithelium</td>
<td>Mesenchymal cell mass</td>
</tr>
<tr>
<td>Leydig cells</td>
<td>Mesonephric elements</td>
<td>Interstitial cells</td>
</tr>
<tr>
<td>Rele testes</td>
<td>Septa and tunica albuginea</td>
<td>Rele ovaris</td>
</tr>
<tr>
<td>Tunica vaginalis</td>
<td>Spermatogoniasperm</td>
<td>Primordial germ cells</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oogoniaova</td>
</tr>
<tr>
<td>Genital Ducts</td>
<td>Ductuli efferentes</td>
<td>Mesonephric tubules</td>
</tr>
<tr>
<td></td>
<td>Aberrant ductules</td>
<td>Paroophoron</td>
</tr>
<tr>
<td></td>
<td>Epididymis</td>
<td>Mesonephric (wolffian) ducts</td>
</tr>
<tr>
<td></td>
<td>Vas deferens</td>
<td>Gartner ducts</td>
</tr>
</tbody>
</table>
Müllerian ducts
Bartholin glands
Upper vagina

External Genitalia

Penis
Corpora cavernosa
Labia majora
Prostate
Bulbourethral glands (of Cowper)
Prostatic utricle (vagina masculina)

Vagina (lower)

It is thought that testosterone and DHT have different roles. The testosterone-receptor complex modulates the secretion of LH by the hypothalamic-pituitary unit and processing of the mRNA, the RNA moves to the cytoplasm, where it is translated by cytoplasmic ribosomes and results in synthesis of new proteins and hence and functional key roles for residues in the androgen receptor by the study of a range of mutations that cause androgen resistance (see later). After transcription and presence of ligand; thus the ligand is internalized and does not reside on the surface of the androgen receptor. Much information has been gained about structural ligand-binding domain that binds to receptor-interacting motifs, LXXLL, in co-regulators (L is leucine, X is any amino acid).

The mechanism of dimerization, nuclear localization signaling, and ligand binding.

In the absence of ligand, steroid hormone receptors (but not the thyroid hormone receptor) are complexed to chaperone proteins that are heat shock proteins and form large receptor complexes. Ligand binding causes a conformational change in the receptor, additional phosphorylation of serine residues located in the amino-terminal trans-activation domain, dissociation of the receptor from heat shock proteins, and "activation" of the receptor.

The monomeric activated steroid-receptor complex is smaller (<45S) than the unactivated complex (>85S). The steroid-receptor complex dimerizes and binds to palindromic steroid-responsive elements in genomic DNA that are upstream from CAAT and TATA boxes. The AF regions interact with an intermediary group of proteins termed co-regulators to form protein-protein interactions in a ligand-dependent manner to either increase (co-activator) or decrease (co-repressor) gene transcription. The multiprotein complex includes co-regulators such as CBP, SRC-1, and ARA70, the last one believed to be a protein more specifically associated with the androgen receptor. Co-regulators appear to act as a physical bridge to link nuclear receptor transcription factors to the general transcriptional machinery, such as RNA polymerases.

RNA polymerase and other transcription factors and co-activators are recruited to initiate transcription of the steroid response gene at a point 19 to 27 kb downstream of the TATA box. The mechanism of trans-activational transcription is ill understood but the ligand-dependent AF-2 region is located within one of the -helices (H12) of the ligand-binding domain that binds to receptor-interacting motifs, LXXLL, in co-regulators (L is leucine, X is any amino acid). The crystal structures of several nuclear receptor ligand-binding domains, including most recently the androgen receptor, have been characterized. They have in common 12 -helices in the form of a sandwich fold. The most carboxyl terminus is helix 12, which becomes realigned by closing like a lid to form a hydrophobic pocket in the presence of ligand; thus the ligand is internalized and does not reside on the surface of the androgen receptor. Much information has been gained about structural and functional key roles for residues in the androgen receptor by the study of a range of mutations that cause androgen resistance (see later). After transcription and processing of the mRNA, the RNA moves to the cytoplasm, where it is translated by cytoplasmic ribosomes and results in synthesis of new proteins and hence androgenic effects. Despite an abundance of biologic actions of androgens ranging from male sex differentiation to muscle growth, erythropoiesis and skeletal development, little is known about the target genes that mediate these effects. Known androgen-responsive genes are the prostate-specific proteins, probasin and prostate-specific antigen. However, the myriad of androgen-responsive genes that must be developmentally regulated during male fetal sex differentiation remain to be identified and characterized.

It is thought that testosterone and DHT have different roles. The testosterone-receptor complex modulates the secretion of LH by the hypothalamic-pituitary unit and affects the stabilization of the Wolffian ducts, whereas the DHT-receptor complex acts in the fetus to promote masculinization of the urogenital sinus and external genitalia of the fetus and acts puberty to induce maturation of secondary sex characteristics. A defect in any of the essential steps in the action of...
androgen in a male fetus impairs masculinization of the urogenital sinus and external genitalia (Fig. 22-32).

**Role of Androgens in the Differentiation of the External Genitalia and Urogenital Sinus**

The induction of male differentiation of the external genitalia and urogenital sinus is affected by DHT, the 5-reduced metabolite of testosterone. Testosterone is the prohormone that is delivered through the blood stream to these target tissues, which are rich in the enzyme 5-reductase and can readily convert testosterone to DHT, even before the fetal testis acquires the capacity to secrete testosterone. DHT binds to the androgen receptor and initiates the events that lead to androgen action. As in the case of the genital ducts, there is an inherent tendency for the external genitalia and urogenital sinus to feminize in the absence of fetal gonadal androgen secretions. Complete male differentiation of the external genitalia and urogenital sinus occurs only if the androgenic stimulus is received during the critical period of development (8 to 12 weeks) in fetal life. DHT stimulates growth of the genital tubercle and induces fusion of the urethral folds and labioscrotal swellings. It also induces differentiation of the prostate and inhibits growth of the vesicovaginal septum, thereby preventing the development of the vagina. These morphogenetic effects of androgen seem to be mediated by the mesenchyme of these tissues and not by the overlying epithelium.

After about the 12th week, when the vagina has separated from the urogenital sinus, fusion of the labioscrotal folds and urethral groove cannot occur, even with an intense androgenic stimulus. The pattern of expression of the androgen receptor explicates, at least in part, why androgen exposure that begins after about the 12th to 13th week of gestation in the female fetus no longer has the capacity to masculinize the urogenital sinus and its derivatives. Androgen receptors, which are expressed as early as the 9th week of gestation, are absent or greatly diminished by the 14th week in the urothelium of the urogenital sinus and the lower vaginal epithelium as well as in the surrounding stroma. However, androgen receptors persist in the clitoris; androgenic stimulation can cause clitoral hypertrophy at any time during fetal life or after birth.

The male fetus with steroid 5-reductase-2 deficiency and impaired conversion of testosterone to DHT has defective masculinization of the external genitalia and urogenital sinus, including absence or hypoplasia of the prostate. The failure of testosterone to masculinize the fetal external genitalia has been ascribed primarily to inability of the target tissues to form DHT. The growth of the external genitalia at puberty in patients with steroid 5-reductase-2 deficiency is attributed to the postnatal expression of the isozyme 5-reductase-1.

Whereas in some species fetal pituitary gonadotropins are required to sustain the secretion of testosterone by the fetal testes, in humans, placental hCG stimulates fetal Leydig cell development and function; human fetal pituitary LH plays a role only after differentiation of the external genitalia is already advanced. This probably explains why the external genitalia of male infants with anencephaly or fetal hypopituitarism and pituitary gonadotropin deficiency usually differentiate normally but remain small. Incomplete fusion of the labial folds and retention of the vaginal pouch in male infants may therefore be caused by deficient androgen secretion or by failure of the target tissues to respond to androgenic stimulation. Conversely, if female infants are subjected in utero to androgenic stimulation from some gonadal or extragonadal source, the external genitalia can masculinize, ranging from clitoral hypertrophy to the formation of a normal-appearing penis. Thus, similar external abnormalities can be produced in the male by androgen deficiency (or failure of the target tissues to respond) and in the female by exposure to androgen from some pathologic source in the fetus or mother.
Endocrine and Paracrine Control Mechanisms in Sex Differentiation

The regulation of sex differentiation by chemical messengers involves two types of control mechanisms. One is the classic endocrine mechanism: a cell, usually in a discrete endocrine gland, secretes a hormone into the blood stream, where it is transported to a distant target tissue to regulate or induce differentiation. Testosterone is a striking example of an endocrine secretion; testosterone secreted by the fetal Leydig cell is delivered through the circulation to the anlagen of the external genitalia and urogenital sinus. Similarly, hCG synthesized by the syncytiotrophoblast acts on the Leydig cell to stimulate testosterone secretion.

The second type of regulation in sex differentiation is paracrine control. This local and more primitive regulatory mechanism involves the dissemination of a hormone from its site of synthesis to its target cells by local diffusion through the extracellular space. Examples of this delivery system for chemical messengers are the action of AMH on the müllerian duct and the action of testosterone on the wolffian duct (in this instance testosterone is a paracrine secretion).
Hormonal Sex Differentiation

Sex differentiation is not complete until the secondary sexual characteristics have matured, fertility is attained, and the ultimate goal, reproduction, becomes possible (Fig. 22-33) (Figure Not Available). These developments occur during puberty. In the past, puberty was regarded as a de novo event because of the dramatic changes brought about by the maturation of the gonads and the increased secretion of gonadal steroids. However, the development of gonadal function is actually a continuum extending from the differentiation of the gonad and the ontogeny of the hypothalamic- pituitary-gonadal system in the fetus, through puberty, to the attainment of full sexual maturation and fertility. Puberty is not an isolated event but rather a critical stage in a sequence of complex maturational changes. The hypothalamic-pituitary unit (including the pulsatile secretion of hypothalamic LHRH and FSH and LH) functions in the fetus, is suppressed to a low level of activity for about a decade during childhood, and is reactivated at the onset of puberty. The hormonal changes and the neuroendocrinology of puberty, including adrenarche and gonadarche, are reviewed in Chapter 24.

Sex Differentiation in the Hypothalamus

Although the control of gonadal function in both sexes is mediated by both FSH and LH, the secretory patterns of the gonadotropins differ in males and females. The male pituitary characteristically secretes both FSH and LH in a pulsatile but relatively constant and sustained manner, whereas in the mature female the pulsatile secretion of FSH and LH is cyclic and is characterized by a preovulatory gonadotropin surge that leads to ovulation.

In 1936, Pfeiffer reported that the rat pituitary becomes differentiated during the early postnatal period according to the nature of the gonads, but the cyclic secretory pattern characteristic of the female pituitary is not an innate property of the pituitary itself. The pituitary of a male animal, when grafted under the hypothalamus of an adult female, is fully able to sustain the rhythm of repeated estrous cycles. When the male pituitary is grafted elsewhere in the recipient, ovulation fails to occur. Therefore, the hypothalamus or higher neural centers function differently in the two sexes.

In addressing the issue of whether genes encoded by the sex chromosomes cause sex differences in development, it is essential to consider the sex chromosomes themselves. ZFY is expressed in the hypothalamus and frontal and temporal cortex of the human adult male brain, as well as fetal brain, but its function is unknown. SRY is expressed in the hypothalamus and frontal and temporal cortex of the human adult male brain, as well as fetal brain, but its function is unknown.
PSYCHOSEXUAL DIFFERENTIATION

A counterpart to Jost's radical and seminal contribution to our understanding of the developmental effects of androgens and AMH in male sex differentiation was the discovery in the laboratory of W. C. Young by Phoenix, Goy, Gerall, and Young in 1963 more than a decade later of the "organizing action" of prenatal administration of testosterone to the pregnant guinea pig, a long gestation species (compared with the rat and mouse), on the mating behavior of the affected female offspring. Their study, similar to earlier, but less definitive, work by Dantchakoff, described male mounting behavior and "defeminization" of the female pseudohermaphrodite guinea pig and emphasized a critical period (30 to 65 days of gestation) for the exposure to the testosterone. This was the first study to demonstrate unambiguously the fact that prenatal exposure to androgen could irreversibly masculinize the fetal brain and lead to male-type behavior postnatally. It extended to the central nervous system the developmental actions of the fetal tests and its androgenic secretions.

Sexually dimorphic behavior may be classified into four broad categories (Table 22-6): (1) core gender identity, which is defined as the fundamental sense of self as being male or female; (2) gender role behavior, which refers to participation in stereotypical masculine and feminine activities; (3) gender or sexual orientation, which refers to preference for sexual partners of the same or other sex; and (4) cognitive differences.

Studies in lower species suggest that the sexual role adopted at maturity is determined by the hormonal environment in early life. As with other aspects of sex differentiation, there appears to be an innate tendency to develop female sexual postures. Development of male patterns of sexual behavior in lower species is influenced to a large extent by exposure to androgens, in particular testosterone, and their aromatization to estrogens during the prenatal and perinatal period. This organizing capacity of testosterone administered at a critical stage of development has been localized to specific areas of the brain. The sex hormones act through specific nuclear receptors in the nervous system as well as by nongenomic actions. There is a difference between the sexes in the volume of the sexually dimorphic nucleus in the preoptic area of the rat brain, and structural differences of the brains of the two sexes are now recognized in many species, including humans. However, in humans, aromatase does not contain the transport of estradiol to the central nervous system, and testosterone-binding globulin (TβG), also called sex hormone-binding globulin (SHBG), has a greater affinity for testosterone than for estradiol. Therefore, the protection hypothesis does not apply to be applicable in humans. The human, masculinization of the brain does not result from estrogen aromatized from testosterone but from testosterone and related C11 steroids. Further, XY women with the complete androgen resistance syndrome do not have a "masculinized" brain; androgen action is ineffective and fetal estrogen synthesized in the central nervous system by aromatization does not induce masculine behavior. Apparently.

fetal estrogens do not have a critical effect on the human on either feminization or masculinization of the brain; the human fetus is bathed in circulating estrogens that do not exhibit a sex dimorphism. Human males are more frequently left handed than are females, a difference that has been ascribed, at least in part, to the prenatal effect of testosterone on the brain.
Sexual identity is usually established by 18 to 30 months of age. If at puberty discordant secondary sexual characteristics are allowed to mature, some individuals may develop doubts about their true gender identity. This has been described in some patients with 5-reductase-2 deficiency, 17-hydroxysteroid dehydrogenase 3 (17-HSD 3) deficiency, or 45,XX/46,XY mosaicism. About half to two thirds of these XY individuals assigned a female sex changed their gender role to male consequent to virilization by the retained testes at puberty. Imperato-McGinley and associates reported on a geographic and cultural isolate of male pseudohermaphrodites with 5-reductase deficiency who masculinized at puberty. In these patients, a change in gender behavior was common but not invariable. Furthermore, differentiation of the sexually dimorphic nucleus of the preoptic area of the human hypothalamus is detectable between about 5 years of age and puberty, although its role in gender identity is unknown. These clinical studies have cast increasing doubt on the hypothesis that gender identity is irreversibly fixed by environmental factors by 2 to 3 years of age. They also emphasize the effect of gonadal steroids at puberty on gender identity and behavior and attest to the plasticity of gender identity in the cultural-genetic isolates studied. In another genetic isolate of patients with 5-reductase deficiency in the Eastern Highlands Province of Papua New Guinea, gender role appeared to be affected primarily by social-experiential and cultural factors rather than by hormonal mechanisms.

Stronger credence is now given to the role of early hormonal influences on sexually dimorphic behavior in humans, and in, for example, the rhesus monkey. As noted studies of women and girls with prenatal virilization caused by virilizing CAH or maternal ingestion of progestagens demonstrate no effect on gender identity in well-managed patients. However, gender-related behavior can be affected and result in general, but not always, in a variable degree of masculine sex-dimorphic behavior as compared with unaffected girls. Prenatally androgenized girls have more interest in outdoor play and competitive sports and are more "tomboystyle" than are unaffected girls. As a group, they are more career oriented and tend to lack a strong interest in doll play and mothering. The pattern is persistent and is not abnormal for female behavior in our culture. According to some studies, a higher proportion of young women with virilizing CAH rated themselves as bisexual or homosexual; another study did not support this contention. The interpretation and implications of psychosexual studies in CAH women have been critically assessed.

We have discussed our present, and far from complete, knowledge of the sex dimorphism in brain structure, the effect of androgens on sex differentiation of the brain, and sex differences in the expression in brain of certain genes on the sex chromosomes, which can be considered important components of the intrinsic or nature component of gender identity. The evidence related to hypothalamic males, patients with the complete form of androgen resistance, and prenatally virilized girls supports the thesis that exposure to androgens before birth can contribute to the programming of sexually dimorphic behavior. Even in XY individuals who had normal testicular function and androgen responsiveness in fetal life and had either a severe malformation of the penis or traumatic loss in the neonatal period, androgens have a "facilitative," not a deterministic, preordained role in gender identity. These hormonal factors are not predictably decisive in patients with classic abnormalities of sex differentiation. More important elements in the development of gender identity are the assigned sex of rearing, the reinforcement that this assignment receives during infancy and early childhood by the parents and siblings, and reinforcement by appropriate gonadal steroid secretion or replacement therapy at the normal age of puberty. If this reinforcement is weak because of ambiguous attitudes of the parents and the community setting, the outlook for attaining a normal gender identity in adult life is diminished.

In sum, the gender neutrality at birth dogma is no longer tenable. There is also no convincing support for the hypothesis that fetal androgens, through their action on the organization and function of the developing central nervous system, are the sole determining factor in the evolution of male gender identity. We do not know if there is a critical period for masculinization of the fetal brain as is the case in the guinea pig. And quantitative, even grossly qualitative, aspects of the magnitude of androgen exposure and the variation in the central nervous system response leading to masculinization of the brain are not known. Rather, the complex, still poorly understood interaction of nature (androgens and genes) and nurture (primarily socialization, parental interaction and reinforcement, and self experience) results in genetic identity. That no one factor has proven to be an invariant determinant is a manifestation of the plasticity and complexity of the process and of the variation of biologic and psychosocial factors, even the dominance of one over the other in a given individual. For example, SRY and ZFY are expressed in the male brain, but yet XY individuals with complete androgen resistance (well represented by those with a deletion or null mutation of the X-linked gene encoding the androgen receptor) exhibit female gender identity. Female pseudohermaphrodites with virilizing CAH exhibit varying degrees of fetal androgen effects on the brain, but the vast majority express female gender identity, especially those in whom glucocorticoid treatment has suppressed the postnatal androgen excess. The few studies of ablation penis discussed earlier, and the studies of Reiner and associates in XY patients with cloacal exstrophy orchidectomized in the neonatal period and assigned a female sex of rearing, support the critical role of the secretion of testosterone by the normal fetal testes in masculinizing the brain. However, even in these rare disorders some affected XY individuals had a female gender identity whereas others had a strong male gender identity. In addition, the outcomes of genital surgical techniques, once quite discouraging, have greatly advanced, including the repair of severe hypospadias often at a single stage and feminizing genioplasty of the masculinized external genitalia of female pseudohermaphrodites.

Our understanding of the complexity of determinants of gender identity and gender role behavior has advanced remarkably in recent years. But there are large gaps in our knowledge and understanding; particularly limiting is the lack of long-term follow-up into adulthood of intersex individuals treated in the past two to three decades, which leaves little room for dogma.
CLASSIFICATION OF ERRORS IN SEX DIFFERENTIATION

In the past, individuals with hermaphroditism were classified according to their gonadal morphology. In the terminology of Klebs, a true hermaphrodite is a person who possesses both ovarian and testicular tissue. A male pseudohermaphrodite is one whose gonads are exclusively testes but whose genital ducts or external genitalia, or both, exhibit the phenotype of a female or incompletely differentiated male. A female pseudohermaphrodite is a person with ovaries whose external genitalia exhibit some masculine characteristics. We have classified errors in sex differentiation by a modification and expansion of this broad framework and have attempted to blend etiologic mechanisms and clinical entities into a simplified rational classification (Table 22-7). The clinical and etiologic heterogeneity of syndromes with similar anatomic findings merits emphasis. The format of this edition thus maintains the traditional pseudohermaphroditism terminology, which is useful clinically for diagnostic algorithms. In discussing intersexuality with patients and parents it is important to refer to the masculinized female and the undermasculinized male.

TABLE 22-7 -- Classification of Anomalous Sexual Development

I. Disorders of Gonadal Differentiation
   A. Seminiferous tubule dysgenesis (Klinefelter syndrome)
   B. Syndrome of gonadal dysgenesis and its variants (Turner syndrome)
   C. Complete and incomplete forms of XX and XY gonadal dysgenesis
   D. True hermaphroditism

II. Female Pseudohermaphroditism
   A. Androgen-induced
      1. Congenital virilizing adrenal hyperplasia
      2. CYP19 (P450arom) aromatase deficiency
      3. Glucocorticoid receptor gene mutation
      4. Androgens and synthetic progestagens transferred from maternal circulation
   B. Other teratologic factors (nonandrogen-induced) associated with malformations of intestine and urinary tract

III. Male Pseudohermaphroditism
   A. Testicular unresponsiveness to hCG and LH (Leydig cell agenesis or hypoplasia due to hCG/LH receptor defect)
   B. Inborn errors of testosterone biosynthesis
      1. Enzyme deficits affecting synthesis of both corticosteroids and testosterone (variants of congenital adrenal hyperplasia)
         a. STAR deficiency (congenital lipoid adrenal hyperplasia)
         b. Side-chain (P450sc) cleavage deficiency heterozygote
         c. 3-Hydroxysteroid dehydrogenase/4,5-isomerase type 2 (3-HSD-2) deficiency
         d. CYP17 (P450c17[17,20 lyase]) deficiency
         e. Smith-Lemli-Opitz syndrome: 7-dehydrocholesterol reductase deficiency
      2. Enzyme defects primarily affecting testosterone biosynthesis by the testes
         a. CYP17 (P450c17[17,20 lyase]) deficiency
         b. 17-Hydroxysteroid dehydrogenase type 3 (17-HSD 3) deficiency
   C. Defects in androgen-dependent target tissues
      1. End-organ resistance to androgenic hormones
         a. Syndrome of complete androgen resistance and its variants (testicular feminization and its variant forms)
         b. Syndrome of incomplete androgen resistance and its variants (Reifenstein's syndrome)
         c. Androgen resistance in phenotypically normal males (infertile and fertile)
      2. Defects in testosterone metabolism by peripheral tissues; 5-reductase-2 (SRD5A2) deficiency (pseudovaginal perineoscrotal hypospadias)
   D. Dysgenetic male pseudohermaphroditism
      1. XY gonadal dysgenesis (incomplete)
      2. XO/XY mosaicism, structurally abnormal Y chromosome, SRY mutation
      3. Deryns-Drash syndrome (WT1 mutation)
      4. Fraiser syndrome (mutation of WT1 splice site junction mutation-deleting KTS)
      5. WAGR syndrome (WT1 deletion)
      6. Campomelic dysplasia (SOX9 mutation)
      7. SFI mutation
      8. DAX1 (duplication)
      9. WNT4 (duplication)
      10. 9p (DMRT1 deletion)
      11. 10q
      12. ATRX syndrome (XN2 mutation)
      13. Testicular regression syndrome
   E. Defects in synthesis, secretion, or response to antimüllerian hormone: persistent müllerian duct syndrome (female genital ducts in otherwise normal men; herniae uteri inguinale)
   F. Maternal ingestion of progestagens
   G. Environmental chemicals (endocrine disrupters)

IV. Unclassified Forms of Abnormal Sexual Development
   A. In males
Disorders of Gonadal Differentiation and Sex Chromosome Anomalies

Not all patients with anomalies of sex chromosomes have abnormal gonads; conversely, congenital defects in gonadal differentiation are not always caused by chromosomal errors. The association is so frequent, however, that these topics are inseparable. Exceptions to this association are of special importance in defining the genetic and chromosomal determinants of gonadogenesis. (see Table 22-4).

Seminiferous Tubule Dysgenesis: Klinefelter's Syndrome and Its Variants

47,XXY Seminiferous Tubule Dysgenesis (Typical Klinefelter's Syndrome)

Seminiferous tubule dysgenesis is a common cause of primary hypogonadism and male infertility (Table 22-8). This syndrome, as defined by Klinefelter and associates, usually becomes manifest first during adolescence as gynecomastia, a variable degree of androgen deficiency, small atrophic testes with hyalinization of the seminiferous tubules, aggregation of Leydig cells, aspermatogenesis, and increased plasma gonadotropins, especially plasma FSH.

In 1956, several groups found that a high proportion of patients with this syndrome are X chromatin positive despite their phenotypic male appearance. In 1959, Jacobs and Strong and Ford and co-workers first reported a 47,XXY sex chromosome constitution in patients with this disorder, explaining the positive sex chromatin pattern. Various other sex chromosome compositions, including mosaicism, were described subsequently. Virtually all these variants have in common the presence of at least two X chromosomes and a Y chromosome, except for the rare group that has a 46,XX sex chromosome complement by karyotype analysis of multiple tissues.

The differentiation of testes and lack of ovarian differentiation in patients with 47,XXY and, more strikingly, in those with 49,XXXXY complements indicate that a single Y chromosome and the expression of the testis-determining gene (SRY) are sufficient to bring about testis organogenesis and male sex differentiation in the presence of as many as four X chromosomes.

Clinical Features

In the postpubertal patient, the only constant clinical features are a male phenotype, small testes, firm testes; seminiferous tubule dysgenesis; azoospermia; Leydig cell hyperplasia; and blunted in adulthood. Impairment in verbal I.Q. is slight (10 to 20 points), and the mean I.Q. falls between 85 and 90, with a wide range of variation. Most boys with Klinefelter's syndrome require help in reading and spelling. Severe retardation is uncommon. One study of 13 47,XXY males monitored from birth through adolescence noted a mean I.Q. 21 points lower than that of a matched control group. Despite neurocognitive difficulties including reading and writing, 12 of 13 graduated from high school and 4 attended college. In general, lack of motor skills hindered participation in competitive sports; 3 of 13 were successful in achieving personal goals and in their family relationships. The families of these 3 boys were among the most stable and supportive of the study group. Most patients with Klinefelter's syndrome do not have behavioral disorders, and, in spite of verbal deficits, adult men with Klinefelter's syndrome are not significantly different as a group from other hypogonadal males or even normal controls as far as education, employment, socioeconomic status, social adjustment, and criminal behavior are concerned.

Patients with a 47,XXY karyotype tend to be taller (mean height 180 cm compared with 174 cm for their normal adult brothers) than average because of disproportionate length of the legs. This finding is present before clinical signs of puberty are evident and may not be accompanied by a proportional increase in arm span. The prepubertal onset suggests that disproportionate leg length is not related to androgen deficiency or delayed epiphyseal closure. Androgen deficiency after the age of puberty augments the prepubertal deviation in skeletal proportions. XXXY individuals have three pseudoautosomal SHOX genes (one in each X chromosome and one on the Y chromosome).

Prepubertally, the basal plasma concentration of FSH and LH and the response to LH-RH are within the normal range for age. The timing and onset of secondary sexual characteristics and puberty were reported as normal in one study and as delayed in another. With the onset of puberty, testicular histology becomes abnormal and testosterone synthesis is impaired. In postpubertal patients the plasma concentrations of testosterone tend to be low, the levels of plasma estradiol are normal or increased, and the gonadotropin levels are elevated. Testosterone responses to hCG appear to be normal in childhood and early adolescence and blunted in adulthood. Potency is usually diminished in the adult, and Leydig cell reserve is impaired, as reflected by a subnormal increase in the concentration of serum testosterone after administration of hCG and an increased concentration of LH in plasma. The testosterone production rate, levels of total and free testosterone, and rates of metabolic clearance of testosterone and estradiol tend to be low, whereas plasma estradiol levels are normal or elevated.

Gynecomastia and signs of androgen deficiency, such as diminished facial and body hair, a female escutcheon, a small phallus, poor muscular development, and a further increase in the disproportion between leg and body length, usually become evident during or after puberty. The testicular failure in Klinefelter's syndrome appears to progress with age. Gynecomastia, which occurs in about 90% of patients, is considered to be secondary to an increased ratio of serum estradiol to testosterone (see also Chapter 18).

Associated Abnormalities

Abnormalities in thyroid function include a diminished thyroid response to thyrotropin, decreased
The frequency of diabetes mellitus is increased. Nielsen reported that 19% had impaired glucose tolerance and that 8% had overt diabetes. The prevalence of diabetes mellitus was also increased in the parents. The patients with diabetes mellitus were usually younger than 50 years of age, and the type II diabetes was usually mild. Insulin resistance with secondary hyperinsulinemia may be the cause of glucose intolerance.

47,XXY patients with gynecomastia have an increased predisposition to cancer of the breast. In a survey of 187 males with breast cancer, 8 patients with chromatin-positive seminiferous tubule dysgenesis were detected, about 18 times the expected prevalence. Whether this increased incidence is solely the consequence of the gynecomastia is unclear (see Chapter 17). Further, whereas male breast cancer is rare, a Swedish study of Klinefelter patients concluded that there is a 50-fold increased risk of developing breast cancer relative to normal males. Infiltrating ductal carcinoma is the most common histologic type. Even though a study of 696 men with Klinefelter’s syndrome did not report an overall increase in cancer incidence, the prevalence of germ cell tumors (particularly in the mediastinum) that secrete hCG and cause LHRH-independent sexual precocity is increased. Twenty to 50 percent of boys 8 years of age or older with primary mediastinal germ cell tumors have Klinefelter’s syndrome. The latter diagnosis is suggested by the association of prepubertal-size testes with sexual precocity. Routine screening is not indicated, given an estimated incidence of germ cell tumors of 1.5 per 1000 persons with Klinefelter’s syndrome but the diagnosis of Kleinfelter’s syndrome should be considered in all boys with a gynecomastia at any site (see Chap. 24).

A British study of mortality and cancer incidence in 646 patients with Klinefelter’s syndrome showed that mortality was increased from diabetes, cardiovascular and respiratory disease, as well as lung and breast cancer.

About 25% of adults with Klinefelter’s syndrome have osteoporosis, but it is uncommon in patients on testosterone replacement. Chronic pulmonary disease and varicose veins with stasis ulcers may also be more prevalent in adults. Patients with both androgen resistance and a 47,XXY karyotype have been reported. These patients had female or ambiguous male genitalia and some clinical features of the 47,XXY karyotype. This combined defect is probably caused by fertilization of an oocyte containing two X chromosomes, each bearing a defect in the X-linked gene coding for the androgen receptor (e.g., uniparental disomy for the X chromosome), by a normal sperm containing a Y chromosome. Similarly, a null mutation, which causes intrauterine death in males usually with X-linked dominant incontinentia pigmenti, is associated with survival in affected XXY patients.

Several males with 46,XY plus a marker chromosome have been reported. In these patients, the marker found was a small ring X chromosome that did not express XIST and hence was not inactivated. The abnormal dosage of X-active chromosome genes caused developmental delay and dysmorphic features.

Surveys of the prevalence of 47,XXY fetuses by karyotype analysis of unselected newborns indicate an incidence of about 1 per 800 to 1000 males, the most common human chromosomal abnormality. No racial or geographic predilection has been observed.

Testicular Lesions.

Klinefelter’s syndrome accounts for 10% to 20% of men attending a male infertility clinic. Changes in the histologic structure of the testis become more marked with age in 47,XXY individuals. A limited number of studies of fetal testes have been reported, and the findings are variable. Grumbach and associates reported normal histology for the testes of a 1700-g chromatin-positive premature infant, as was similarly reported in a 49,XXXXY 21-week fetus, but examination of several other affected fetuses suggested that the germinal epithelium was deficient and that germ cells were heterotopic; these observations indicate that the histology of the fetal testes varies. In three infants aged 3 to 12 months with a 47,XXY karyotype, spermatogonia were decreased. In later childhood, testicular biopsies have revealed small tubules with progressive reduction in spermatogonia. In considering the testicular lesion, it seems that a normal or near-normal complement of germ cells is present early in fetal life. During late gestation and early infancy, a drastic loss of spermatogonia ensues, possibly because of an exaggeration of the normal apoptosis of spermatogonia in the neonatal period. In addition, excessive germ cell loss could result from defective maturation or failure of the germ cells to migrate to the periphery of the tubule and align in opposition to the basement membrane. An experimental XXY mouse model has been established. The histology of the testes in adult mice is consistent with the human syndrome. Progressive loss of germ cells in XXXY mice was observed by 10 days after birth.

With the approach of adolescence, even before pubertal signs are well advanced, the actions of pituitary gonadotropins on the intrinsically defective testis induce progressive hyalinization of the seminiferous tubules and pseudoadenomatous clumping of Leydig cells. Despite this clumping, the mean volume of Leydig cells usually is normal. After pubescence, the testes are characterized by small dysgenetic tubules with arrested development, fibrosis, and hyalinization. The result is testes that are small in size and firm in consistency. Peritubular elastic tissue is usually absent or diminished. That gonadotropin secretion plays a direct or indirect role in the progressive degeneration of the testes was suggested in a 7-year-old 48,XXXY boy with true precocious puberty and elevated urinary gonadotropin levels. Unlike the relatively normal testicular architecture in most boys of this age with Klinefelter’s syndrome, the testes of this boy exhibited extensive hyalinization and fibrosis of the tubules and clumping of Leydig cells.

Hyalinization of the tubules is usually extensive but varies in degree from patient to patient and even between the testes of the same patient. The fibrosis tends to progress with age, and in some older patients few tubules can be identified. Conversely, in some patients the tubules are lined by Sertoli cells, tubular fibrosis is relatively slight, and the histologic appearance resembles that of germinal cell aplasia. Rarely, spermatogenesis is found in isolated tubules. This finding could represent hidden mosaicism in the gonad or possibly mitotic nondisjunction or anaphase lag in germ cells giving rise to 46,XY cells that would then go on to spermatogenesis. There have been sporadic reports of paternity; most fertile individuals proved to have sex chromosome mosaicism, and in others acceptable
distinguish them from patients with typical Klinefelter's syndrome. Analyses of sperm chromosomes from a 46,XY/47,XXY male showed that 1% of sperm had a 24,XY haplotype, which suggests that some 47,XXY cells can undergo meiosis and form XY-bearing spermatooza. The technique of intracytoplasmic sperm injection (ICSI) has been used with some success for achieving fertility, including the birth of twins. However, there is a risk of trisomy 21 in offspring of patients with Klinefelter's syndrome.

47,XXY males may develop through nondisjunction of the sex chromosomes during either the first or second meiotic division in either parent or, less commonly, through mitotic nondisjunction in the zygote at the time of or after fertilization. Fertilization of a 46,XX ovum by a Y-bearing sperm or of an X ovum by a 46,XY-bearing sperm would yield a 47,XXY zygote. Mitotic nondisjunction of the sex chromosomes in a 46,XY zygote could yield a 47,XXY and a 45,Y daughter cell. Because the 45,Y cell line is nonviable, only the 47,XXY cell line would survive.

These abnormalities of meiosis almost always occur in patients with normal sex chromosome constitution. However, Rosenkrantz described two 47,XXY patients whose mothers were, respectively, 47,XXX and 46,XX/47,XXX mosaic. Whether a 47,XXY karyotype is derived more frequently than previously suspected from a polysomic X constitution in the mother remains to be determined.

In a study of 47,XXY males, Jacobs and colleagues found that the XXY constitution resulted from paternal nondisjunction in the first meiotic division in 53% of the patients, from maternal nondisjunction during the first meiotic division in 34%, and from maternal nondisjunction during the second meiotic division in 9%. Three percent of cases appeared to be related to postzygotic mitotic nondisjunction. Ferguson-Smith and colleagues and others reported a positive association with advanced maternal age in 47,XXY patients, although this association is less marked than in trisomy 21. The association with advanced maternal age correlates with first meiotic division errors. Some studies in X Y Y individuals indicate that paternal nondisjunction does not depend on age. A finding reminiscent of that in patients with autosomal trisomy. Nevertheless, the prevalence of XY, YY, and XX disomy in human sperm appears to be increased in older men.

Furthermore, in a study of the sperm of healthy men who were fathers of boys with Klinefelter's syndrome, the frequency of YY sperm (7.5 per 10,000) was 31% higher among fathers in their 40s and 100% higher for those in their 50s.

Rarely, Klinefelter's syndrome is associated with a supernumerary X chromosome that is structurally abnormal, for example, an X-autosome translocation or an isochromosome for the long arm of the X.

Among a small, rare group of phenotypic females with a 46,XXY karyotype, six (one set of monozygotic twins) had the complete androgen resistance syndrome. Both X chromosomes harbored a deleted or mutant gene encoding the androgen receptor. One 47,XXY fertile woman had a structurally abnormal Y chromosome devoid of the SRY gene attributable to an illegitimate X-Y interchange.

Etiologic Factors.

An important factor in the origin of the 47,XXY chromosome constitution is advanced maternal age and maternal nondisjunction. As discussed previously, an important effect in chromosome abnormalities may be a consequence of the long diplonete stage of human ova. Oocytes remain suspended in prophase of the first meiotic division from birth to ovulation, which may not occur for decades. The defective segregation of the two X chromosomes could be caused, at least in part, by reduction of the length of the chiasma as the length of the diplonete stage increases (see section on aneuploidy). As in the syndrome of gonadal dysgenesis, the prevalence of twinning in sibships of 47,XXY individuals may be increased.

 Genetic factors that predispose to nondisjunction have been demonstrated in lower species. Although chromosome abnormalities are usually sporadic, families have been reported in which leukemia and various chromosome abnormalities have occurred in siblings and relatives. In addition, patients with more than one form of trisomy seem to occur more frequently than expected by chance alone. A role for radiation, viruses, environmental toxins, folate deficiency, or autoimmune as a predisposing factor has not been established.

Diagnosis and Treatment.

The diagnosis of Klinefelter's syndrome in the postpubertal male is suggested by the typical phenotype and hormonal changes and confirmed by the finding of a 47,XXY karyotype or a variant sex chromosome complement in blood, skin, or gonads. Treatment should be directed toward androgen replacement therapy when there is evidence of androgen deficiency. In general, parenteral androgens are more effective in inducing virilization and are safer than oral preparations.

Hepatic tumors and abnormalities in liver function have been associated with chronic administration of oral androgens that have substitutions at the 17-position (e.g., a methyl group), but such abnormalities are not a problem with testosterone ester preparations such as propionate or enanthate or the testosterone dermal patch. Testosterone enanthate in oil, 200 mg intramuscularly every 2 to 3 weeks, is recommended for full replacement therapy, but it is wise to begin therapy at a lower dose (e.g., 50 mg intramuscularly every 4 weeks) to avoid rapid virilization and bone maturation, especially in adolescent males. The testosterone dermal patch is a useful, but more expensive, therapeutic approach. In general, conspicuous and long-standing gynecomastia does not diminish as a result of androgen replacement. However, in some patients, especially those with less striking gynecomastia, regression can occur with androgen replacement therapy. Severe or psychologically disturbing gynecomastia is corrected by reduction mammoplasty or by liposuction.

The diagnosis of Klinefelter's syndrome should be suspected in prepubertal patients with one or more of the following: (1) long legs, (2) smaller than normal penis, (3) learning disorders, and (4) developmental delay in speech and language. Similarly, based on the outcome of cytogenetic surveys of newborns, it is recommended that all boys with undescended testes or a micropenis or who have gynecomastia should have a karyotype performed. Many of these features are amenable to therapy, so early detection and intervention may be beneficial. Nielsen and Pels et al. suggested that prepubertally diagnosed patients with Klinefelter's syndrome should be offered therapy with testosterone at 11 to 12 years of age to initiate puberty and to prevent the physical and psychological complications of hypogonadism. We have employed a regimen that begins replacement therapy with 50 mg of testosterone enanthate in oil intramuscularly each month at a bone age of 12 years. When the bone age has advanced to 14 years, the dose may be increased to 100 mg intramuscularly monthly. After several years of treatment at these doses, a height is usually attained that is appropriate for genetic height potential of the individual and pubertal progression is usually adequate. When full masculinization is desired, an adult replacement dose of 200 mg every 2 weeks or 300 mg every 3 weeks may be given.
Approximately 20% of 46,XX males are SRY negative. These males have a much higher prevalence of genital ambiguity, such as micropenis, hypospadias, and peritubular and interstitial fibrosis occurs. In some patients the morphology of the testes is similar to that in 47,XXY patients. The diagnosis may be suspected from the clinical features. Male phenotype and psychosocial orientation and are similar clinical and endocrinologically to individuals with classic Klinefelter's syndrome except for minor differences. Postpubertally, as in Klinefelter's syndrome, they have decreased libido and potency may not appear until the fourth or fifth decade. The Leydig cells appear hyperplastic. In some patients the morphology of the testes is similar to that in 47,XXY males; they have small testes with azoospermia. A 46,XX karyotype is present in about 1 of every 20,000 phenotypic males. In general, the greater amount of Y material present, the more virilized the phenotype. Several theories have been advanced to explain this type of sex reversal: (1) loss of a Y chromosome early in embryogenesis, (2) cryptic sex determination. Several theories have been advanced to explain this type of sex reversal: (1) loss of a Y chromosome early in embryogenesis, (2) cryptic sex differentiation, (3) translocation between a Y chromosome and an X chromosome or autososome resulting in the presence of the testis-determining gene (or genes) on an X chromosome or autosome, (4) a mutation involving either an autosomal or X-linked gene in the pathway to testis differentiation. Such 46,XX patients have had three copies of the pseudoautosomal region in their constitution, most are mentally retarded, although this may in part relate to ascertainment bias, because a significant proportion of cases are detected in screening mental and psychiatric hospitals. The 48,XXXX karyotype is usually associated with tall stature (the mean height of 26 patients was 181 cm), disproportionately low extremities, gynecomastia, delinquent behavior, and unusual dermatoglyphic patterns. Peripheral vascular disease, especially varicose veins and stasis dermatitis, has been observed. Secondary sexual characteristics are poorly developed, and testicular histology is similar to that of 47,XXY patients. The sex chromatin pattern is indistinguishable from that of the 47,XXY groups; however, two fluorescent Y bodies are present in a high proportion of somatic nuclei. Further resolution of meiosis and nondisjunction must occur in paternal meiosis. In two informative matings the Xb blood groups indicated that the factor contributed an X as well as two Y chromosomes, suggesting that an X ovum was fertilized by an XY sperm (arising from successful nondisjunction in the first and second meiotic divisions). The 48,XXXX karyotype in a patient whose mother was 47,XXX could have arisen by the fertilization of an XX ovum by a YY sperm. The diagnosis of 46,XY/47,XXY mosaicism may result from nondisjunction or anaphase lag in 47,XXX zygote.

46,XX males.

A 46,XX karyotype is present in about 15% of all 20,000 phenotypic males. These patients have a male phenotype and psychosocial orientation and are similar clinically and endocrinologically to individuals with classic Klinefelter's syndrome except for minor differences. Postpubertally, as in Klinefelter's syndrome, they have decreased libido and potency. Increased androgen deficiency. With the increase in the number of X chromosomes, the severity and frequency of somatic anomalies such as short neck, epicanthal folds, radioulnar synostosis, and clindactyly increase. Mental retardation, somatic anomalies, and small testes also occur in 46,XXXXY patients.

46,XXX.

The 46,XXX karyotype has been reported in more than 100 patients since the first report by Fraccaro and colleagues in 1960. The diagnosis may be suspected from the clinical features. Male phenotype and psychosocial orientation are similar clinically and endocrinologically to individuals with classic Klinefelter's syndrome except for minor differences. Postpubertally, as in Klinefelter's syndrome, they have decreased libido and potency. Compared with males with a 47,XXY karyotype, 46,XXY males have a lower frequency of intellectual and psychosocial problems; they are shorter (mean height, 168 cm) than 47,XXY or normal males; they have varying degrees of testosterone deficiency, gynecomastia, and small testes with azoospermia. Testosterone production is usually decreased, as is the response to hCG. Both basal and LH-RH-induced FSH and LH levels are increased. There is a 10% incidence of hypospadias that is attributed to a deficiency of testosterone formation by the fetal Leydig cells. 46,XX males with genital abnormalities usually lack evidence of SRY DNA in their genome and manifest a greater prevalence and degree of gynecomastia than 46,XX men. In contrast to the morphology in Klinefelter's syndrome. In contrast to the morphology in Klinefelter's syndrome, the diagnosis may be suspected. The morphology of the testes is similar to that in 47,XXY males; they are older (mean age, 45 years) at the time of diagnosis than 47,XXY patients. Decreased libido and potency may not appear until the fourth or fifth decade. In general, the greater amount of Y material present, the more virilized the phenotype.

46,XY/47,XXY Mosaicism

46,XY/47,XXY mosaicism is the second most common karyotype in chromatin-positive men. The presence of a normal XY cell line in these patients can modify the clinical expression of the 47,XXY cell line. In general, these patients manifest a lesser degree of gynecomastia, androgen deficiency, and testicular pathology. As a group they are older (mean age, 45 years) at the time of diagnosis than 47,XXY patients. Decreased libido and potency may not appear until the fourth or fifth decade. These findings suggest that both anomalous Y-to-X exchange and maternal or paternal nondisjunction occurred in these patients. Approximately 20% of 46,XX males are SRY negative. These males have a much higher prevalence of genital ambiguity, such as micropenis, gynecomastia, and azoospermia. SRY sequence and the PAR.

Approximately 20% of 46,XX males are SRY negative. These males have a much higher prevalence of genital ambiguity, such as micropenis, gynecomastia, and azoospermia. SRY sequence and the PAR.
associated with true hermaphroditism. These observations are consistent with the origin of a small proportion of 46,XX males as a consequence of a mutation in a downstream autosome or X-linked gene (or genes) involved in the testis-determining cascade. They also suggest that 46,XX males and 46,XX true hermaphrodites may arise by similar pathogenic mechanisms. In sum, the 46,XX male syndrome arises mainly as a result of a Y-to-X translocation; other possible mechanisms include mutation in an X chromosomal or autosomal gene and cryptic mosaicism that involves a Y-bearing cell line in all at least the Sertoli cells.

### Syndrome of Gonadal Dysgenesis: Turner's Syndrome and Its Variants

In 1938 Turner described seven phenotypic women with short stature, sexual infantilism, webbing of the neck, and cubitus valgus. Studies of this syndrome and its variants have made a major contribution to the evolution of current concepts of sex differentiation. (For reviews, see references [18] and [19].)

In the early 1940s Albright and colleagues and Vamey and associates found that the excretion of urinary gonadotropin was increased in affected adolescents and adults. Wilkins and Fleischmann soon thereafter described the gonads as bilateral, pale *streaks* of connective tissue situated in the mesosalpinges and devoid of any germ cells. Wilkins proposed, in light of Jost's fetal castration experiments in the rabbit, that some of these functionally agonal patients might be genetic males, because fetal castration of either sex invariably leads to a female phenotype. The discovery in 1954 that many of these patients, contrary to their phenotype, were X chromatin negative seemed initially to confirm that hypothesis, but after techniques became available for analysis of the chromosome constitution Ford and co-workers reported that the sex chromosome constitution in a 14-year-old phenotypic female with this syndrome was 45,X rather than 46,XY. Work in many laboratories thereafter defined more precisely the chromosomal basis of this and related disorders.

The absence of a second sex chromosome (X chromosome monosomy with haploinsufficiency) is associated with five cardinal features: female phenotype, short stature, sexual infantilism owing to rudimentary gonads, a variety of associated somatic abnormalities, and embryonic lethality. These features may be modified by the presence of lesser degrees of sexual chromosome deficiency. It is therefore useful to consider the syndrome of gonadal dysgenesis and its variants as a continuum of features ranging from those of the typical 45,X phenotype to that of a normal female or male. The functional importance of chromosomal additions to the basic 45,X pattern can be deduced from the extent to which they modify normal, in at least some cases, the short stature, sexual infantilism, and somatic anomalies that typify the 45,X patient.

Partial sex chromosome monosomy (haploinsufficiency) may be attributed to a structurally abnormal second sex chromosome (X or Y), sex chromosome mosaicism involving a 45,X cell line, or both. Even though the modified clinical forms are almost invariably associated with partial sex chromosome monosomy, the contrary is not necessarily true; partial sex chromosome monosomies can cause the typical clinical picture found in 45,X patients.

**Typical 45,X Gonadal Dysgenesis (Turner's Syndrome)**

In patients with the cardinal features of sex chromosome monosomy, the X chromosome pattern is negative in about 60%; most of these patients have a 45,X sex chromosome constitution (Table 22-9). Significant variability occurs in expression of the somatic anomalies associated with sex chromosome monosomy ([Fig. 22-38](#)).

### Clinical Aspects

The typical patient (see [Fig. 22-38](#)) is often recognizable by the distinctive facies, in which micrognathia; epicantal folds; prominent, low-set, rotated and/or deformed ears; a fish-like mouth with a narrow, high-arched palate; piosis; and strabismus are present with varying degrees of frequency. The chest is usually square and shield-like with microthelia and inverted nipples. The areolae appear to be widely spaced. The neck is short and broad, and the hairline in back is low. Webbing of the neck is present in 25% to 40%, and

![Diagram of the short arms of the X and Y chromosomes during meiotic pairing. A: A crossover (dashed lines) usually occurs between the pseudoautosomal regions of the X and Y chromosomes. Anomalous but equal crossovers (solid lines) can occur that result in an X chromosome with the sex-determining region (SRY) and a Y chromosome deficient in the SRY. It is estimated to occur at the PRKX and PRKY locus in 40% of SRY XX males. Zygotides with these sex chromosomes will become XX males of XY females as indicated. B, Anomalous unequal crossovers (solid lines) during male meiosis can result in an X chromosome with an SRY gene as well as the pseudoautosomal regions of both the X and Y chromosomes. SRY, sex-determining region Y; ZFY, zinc finger Y.](#)

coarctation of the aorta occurs in 10% to 20%. Those with coarctation usually also have webbing of the neck. Additional anomalies include congenital lymphedema of the feet and hands (30%) ([Fig. 22-39](#)) or puffyiness of the dorsum of the fingers; short fourth metacarpals (50%); renal abnormalities (40%); high-arched palate; various skeletal anomalies; including cubitus valgus, Madelung's deformity of the wrist, genu valgum, and scoliosis; increased number of pigmented nevi; tendency to keloid formation; abnormal nails; recurrent otitis media, which may result in conductive hearing loss (as well as progressive sensorineural loss of hearing); unexplained hypertension; and, rarely, gastrointestinal bleeding secondary to intestinal telangiectasia, hemangiomatosis, or dilated veins. Money and others have reported that impairments of directional sense and space-form recognition, visual-motor coordination, and motor learning are common; this perceptual disability results in a lower mean performance I.Q. than in the general population and is evidence of diffuse or multifocal cerebral dysfunction, whereas verbal ability (including comprehension and vocabulary) is normal. In general, patients with gonadal dysgenesis do not tend to differ from siblings in overall intelligence; however, mental retardation (I.Q. < 70) may be more common than the 1% to 3% incidence reported in the general population. Gender identity and sexual attitudes are feminine. Girls with gonadal dysgenesis are generally more immature and distractable and have less self-esteem, poorer peer relations, and more difficulty at school than peers. Skuse and associates described poorer adjustment and "social cognition" in 45,X M(40), maternally derived X chromosome) individuals with Turner's syndrome. This difference was attributed to genomic imprinting, the silencing of some gene on the maternal X chromosome located in the pericentric region of the short arm or on the long arm of the X chromosome. This putative gene, which apparently escapes inactivation, is the first imprinted gene proposed for the human X chromosome. Further studies of parental origin of the X chromosome in Turner's syndrome are helping to elucidate behavioral phenotypes for a number of neurodevelopmental aspects.

As adults, women with gonadal dysgenesis exhibit more conservative sexual attitudes and a more negative body image. Lack of ovarian function (including infertility), rather than height, was the major concern of adult women with the syndrome of gonadal dysgenesis in one study. Severe psychopathological manifestations are uncommon. In general, most women with gonadal dysgenesis are independent, self-sufficient, and socially active. One study of 63 women with Turner's syndrome reported normal psychological well-being but more difficulties in establishing social and partner relationships compared with the general population.

### TABLE 22-9 — Clinical Features of 45,X Gonadal Dysgenesis (Turner's Syndrome)

<table>
<thead>
<tr>
<th>Karyotype</th>
<th>45,X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inheritance</td>
<td>Sporadic; meiotic or miotic disjunction</td>
</tr>
<tr>
<td>Genitalia</td>
<td>Female</td>
</tr>
<tr>
<td>Wolffian duct derivatives</td>
<td>Absent</td>
</tr>
<tr>
<td>Müllerian duct derivatives</td>
<td>Normal female</td>
</tr>
<tr>
<td>Gonads</td>
<td>Streak</td>
</tr>
<tr>
<td>Habitus</td>
<td>Short stature; sexual infantilism at puberty; somatic stigmata</td>
</tr>
</tbody>
</table>
Increased plasma LH and FSH concentrations; decreased plasma estradiol levels

The risk of anorexia nervosa and inflammatory bowel disease may be increased. The epiphysis Bonnie-Ullrich syndrome has been applied to phenotypic female infants with lymphedema of the distal extremities and loose folds of skin over the back of the neck in addition to the typical features of gonad dysgenesis (Fig. 22-40). In the neonate, pleural effusions and ascites that clear spontaneously are not uncommon and pericardial effusion has been reported. The serous effusions and the lymphedema are attributable to hypoplasia and other defects of the lymphatic system. 45, X abortuses commonly exhibit generalized edema and a large hygroma of the neck. Postnatally, the latter abnormality results in webbing of the neck. These findings can be detected by prenatal ultrasonography. Shephard and Fante suggested that the severe edema that occurs secondary to hypoplastic lymphatic and lymphatic duct hypoplasia in 45, X fetuses is responsible for many of the malformations involving the ears, hairline, neck, nipples, nails, and kidneys. The increased incidence of congenital heart disease associated with webbed neck, especially coarctation of the aorta (40%), has led to the suggestion that lymphatic obstruction is involved in the pathogenesis of both types of deformation.

The prevalence of cardiovascular abnormalities has varied from 20% to 50%. In a Danish study of 179 patients, 26% had cardiovascular malformations. Coarctation of the aorta was the most common, occurring in 10% of patients, primarily those with a 45, X karyotype. The prevalence of bicuspid aortic valves in Turner syndrome determined by echocardiography has ranged from 5% to 34%. Bicuspid aortic valves carry an increased risk for subacute bacterial endocarditis and tend to evolve with age into stenotic and/or insufficient aortic valves. An increased incidence of mitral valve prolapse has been reported in patients with gonadal dysgenesis. Other studies have reported an increased incidence of partial anomalous venous drainage and hypoplastic left heart syndrome. On echocardiography, 8% to 29% of patients with the syndrome of gonadal dysgenesis have aortic root dilation, and rupture from aortic dilation has been reported in more than 20 patients. Therefore, all patients with Turner's syndrome should have a thorough baseline cardiac evaluation, including an echocardiogram or magnetic resonance imaging study or both in infancy and again at adolescence. Patients with increased risk factors for dissection and rupture (e.g., those with coarctation, hypertension, and aortic root dilatation) require yearly follow-up and therapeutic measures to decrease the risk of dissection. Pregnancy (after in vitro fertilization) carries an increased risk of fatal aortic dissection. Patients with a bicuspid aortic valve should be given prophylactic antibiotic therapy before surgery or dental procedures, to prevent subacute bacterial endocarditis.

The most common renal abnormalities are rotation of the kidney, horseshoe kidney, duplication of the renal pelvis and ureter, and hydronephrosis secondary to ureteropelvic obstruction. Complete absence of the kidney and gross renal ectopia in 7 of 141 patients was reported by Lippe and associates. Malformation of the kidneys and upper collecting system including an abnormal vascular supply are so common that intravenous urography, a renal sonogram or renography should be obtained routinely at the time of diagnosis.

Skeletal maturation is normal or slightly delayed in childhood and lags further in adolescence as a result of gonadal steroid deficiency. In most cases, the skeleton exhibits localized areas of rarefaction (fish-net appearance), especially of the hands, feet, elbows, and upper femurs. Bicorporeal girls with gonadal dysgenesis have normal bone density for height and age but decreased density at the wrist for bone age and body mass index. The risk of wrist fractures is increased. Bechtold and colleagues, using quantitative computed tomography of the forearm and human growth hormone (hGH) in estrogen-treated adolescent and young adult patients, detected low total volumetric bone mineral density owing to decreased cortical thickness and suggested bone strength may be inadequate for the relatively high body weight. Patients not treated with estrogen often develop a severe form of the post-menopausal type of osteoporosis and may develop fractures and vertebral collapse. Lanes and associates reported in young women with Turner's syndrome the failure to obtain normal peak bone mass despite estrogen replacement therapy begun in adolescence. Osteochondrosis-like changes of the spine, vertebral hypoplasia, and scoliosis are common.

In addition to the metacarpal sign (shortening of the fourth metacarpal), Kosowicz described a carpal sign characterized by a more acute angular configuration of the proximal row of carpal bones. Madelung's or "bayonet" deformity of the wrist, a feature of the Leni-Weill syndrome caused by haploinsufficiency of the SHOX gene, is present in about 10% of patients. Cubitus valgus (an increased carrying angle) occurs in half, is a consequence of a developmental abnormality of the trochlear head, and develops after birth. The knee may show deformities of the medial tibial and femoral condyles with obliterated tibial epiphyses and medial projections of the tibial metaphyses that can result in genu valgum. The pelvis tends to have a male-type infer. Midface hypoplasia is common. The growth of the temporal bone, condylar cartilage, and sphenoid-occipital synchondrosis is abnormal. Scoliosis is present in approximately 10% of patients, secondary either to...

Hormone profile: Increased plasma LH and FSH concentrations; decreased plasma estradiol levels

| LH, luteinizing hormone; FSH, follicle-stimulating hormone |

**Figure 22-38** Variation in physical appearance in five patients with the typical form of the syndrome of gonadal dysgenesis. All of these patients had a 45,X karyotype, and all had differences between height age and chronologic age of 5 years or more. (Modified from Grumbach MM. Some considerations of the pathogenesis and classification of anomalies of sex in man. In Astwood EB [ed]. Clinical Endocrinology. New York, Grune & Stratton, 1960, pp 407-436.)

**Figure 22-39** A patient aged 14 years, 10 months with the typical form of the syndrome of gonadal dysgenesis. The X chromatin pattern was negative and the karyotype was 45,X. She was short (height, 134.5 cm; height age, 9 years, 5 months), was sexually infantile except for the appearance of sparse pubic hair, and exhibited characteristic stigmata of the syndrome: a short webbed neck. These findings can be detected by prenatal ultrasonography. Shephard and Fante suggested that the severe edema that occurs secondary to hypoplastic lymphatic and lymphatic duct hypoplasia in 45,X fetuses is responsible for many of the malformations involving the ears, hairline, neck, nipples, nails, and kidneys. The increased incidence of congenital heart disease associated with webbed neck, especially coarctation of the aorta (40%), has led to the suggestion that lymphatic obstruction is involved in the pathogenesis of both types of deformation.

The most common renal abnormalities are rotation of the kidney, horseshoe kidney, duplication of the renal pelvis and ureter, and hydronephrosis secondary to ureteropelvic obstruction. Complete absence of the kidney and gross renal ectopia in 7 of 141 patients was reported by Lippe and associates. Malformation of the kidneys and upper collecting system including an abnormal vascular supply are so common that intravenous urography, a renal sonogram or renography should be obtained routinely at the time of diagnosis.

Skeletal maturation is normal or slightly delayed in childhood and lags further in adolescence as a result of gonadal steroid deficiency. In most cases, the skeleton exhibits localized areas of rarefaction (fish-net appearance), especially of the hands, feet, elbows, and upper femurs. Bicorporeal girls with gonadal dysgenesis have normal bone density for height and age but decreased density at the wrist for bone age and body mass index. The risk of wrist fractures is increased. Bechtold and colleagues, using quantitative computed tomography of the forearm and human growth hormone (hGH) in estrogen-treated adolescent and young adult patients, detected low total volumetric bone mineral density owing to decreased cortical thickness and suggested bone strength may be inadequate for the relatively high body weight. Patients not treated with estrogen often develop a severe form of the post-menopausal type of osteoporosis and may develop fractures and vertebral collapse. Lanes and associates reported in young women with Turner's syndrome the failure to obtain normal peak bone mass despite estrogen replacement therapy begun in adolescence. Osteochondrosis-like changes of the spine, vertebral hypoplasia, and scoliosis are common.

In addition to the metacarpal sign (shortening of the fourth metacarpal), Kosowicz described a carpal sign characterized by a more acute angular configuration of the proximal row of carpal bones. Madelung's or "bayonet" deformity of the wrist, a feature of the Leni-Weill syndrome caused by haploinsufficiency of the SHOX gene, is present in about 10% of patients. Cubitus valgus (an increased carrying angle) occurs in half, is a consequence of a developmental abnormality of the trochlear head, and develops after birth. The knee may show deformities of the medial tibial and femoral condyles with obliterated tibial epiphyses and medial projections of the tibial metaphyses that can result in genu valgum. The pelvis tends to have a male-type infer. Midface hypoplasia is common. The growth of the temporal bone, condylar cartilage, and sphenoid-occipital synchondrosis is abnormal. Scoliosis is present in approximately 10% of patients, secondary either to...
Short stature is an invariable feature in 45,X individuals and may be evident in utero. Intrauterine growth retardation is common, and the average birth weight (2.83 ± 0.57 kg) and length (48.2 ± 3.2 cm) are 1 standard deviation (SD) below the mean value for normal infants of comparable gestational age. The growth retardation is evident by the middle of the second trimester of gestation and affects all long bones. The mean final height of patients in different series ranged from 142 to 147 cm. In a survey of 12 European countries, a total of 661 patients were found, 51% of whom were 45,X and had a mean final height of 144.3 ± 6.7 cm; none had received estrogen therapy before the age of 14 years. The ratio of sitting to standing height is frequently increased by late childhood and reflects the greater retardation in growth of the legs. Growth curves specific for Turner’s syndrome are available, including growth curves for some nationalities (Fig. 22-41).

Postnatal growth failure is evident from early infancy. By age 3 years, the mean deficit in height is 3.0 ± 1.5 SD. By a bone age of 9 years the difference in mean height between patients and normal individuals (16 cm) is close to the difference at maturity (20 cm). Hence, there is little additional loss of height relative to normal individuals after a bone age of 9 years despite the lack of a pubertal growth spurt (see Fig. 22-41). This may be a result of prolongation of growth as a consequence of the effect of estrogen deficiency on bone maturation. Although final height in untreated patients may not be achieved until late in the second decade of life, there is no major gain in height after the age of 16 years. Final height correlates with birth weight and target height.

The short stature is not attributable to a deficiency of growth hormone, insulin-like growth factor I (IGF-I, also called somatomedin-C), IGF-II, or adrenal or gonadal steroids (see Chapter 23). Decreased amplitude and frequency of growth hormone pulses have been reported after 8 years of age. Likewise, IGF-I levels that are normal up to 10 years of age are low thereafter; however, administration of either estrogen or growth hormone induces a rise in the concentration of plasma IGF-I. The changes in growth hormone secretory dynamics and IGF-I concentrations after 8 to 10 years of age are probably secondary to the lack of the estrogen-induced rise in plasma growth hormone concentration and IGF-I levels at puberty. The cause of the progressive growth failure is attributable, at least in part, to the missing PAR1.

Sexual Infanticilism.

The genital tract and external genitalia in this syndrome are female in character but immature. Located in the mesosalpinges parallel to the fallopian tubes are long, attenuated, pale, fibrous streaks of connective tissue. Typically, these streak-like or spindle-shaped structures consist of fibrous stroma arranged in whorls similar to those in ovarian stroma, but they lack primordial follicles. Vestigial mediulary elements and rudimentary mesonephric tubules like those found in the primitive genital ridge are common at the hilus. After puberty, aggregates of epithelioid cells resembling Leydig or hilus cells are present in variable quantity. Ultrasonography of the pelvis or MRI commonly detects the small uterus and streak gonads.

Singh and Carr studied the gonadal ridges of eight spontaneously aborted embryos and fetuses ranging in gestational age from 5 weeks to 4 months. Primordial germ cells were observed in all eight specimens. Until the third month of gestation, no appreciable differences were noted between these gonads and those from 46,XX fetuses; after that, connective tissue stroma increased and formation of follicles was impaired, suggesting that primordial germ cells seed the 45,X gonad, that many degenerate during oocyte formation and folliculogenesis, and that surviving oocytes undergo accelerated apoptosis. Jirasek noted that oocytes in patients with 45,X karyotype degenerate after formation of the primary follicle, possibly because the surrounding follicular cell layer is incomplete. Two active X chromosomes appear to be required for the normal development of human oogonia and oocytes. Follicles are common in the gonadal streaks of 45,X infants at birth but are uncommon by late childhood and adolescence. Nevertheless, spontaneous puberty, menarche, and fertility can occur in putative 45,X patients.

Longitudinal studies of both basal and LH-releasing hormone secretions demonstrate a lack of feedback inhibition of the hypothalamic-pituitary axis by the dysgenetic gonads in infants and young children with gonadal dysgenesis. In 58 patients aged 2 to 20 years, plasma FSH levels were elevated in those aged 2 days to 4 years and decreased to normal values between 5 and 10 years of age (Fig. 22-42). After 10 years, the plasma FSH level rose again into the castrate range. Therefore, the pattern of plasma FSH concentration followed a diphase curve similar to but higher than that in normal infants and children. The pattern of change in LH levels was comparable, but the concentrations were one third to one tenth those of FSH. LH-releasing hormone-induced LH and FSH responses exhibited a diphase pattern with age, similar to those of basal levels. In patients younger than 5 years of age, both the mean basal levels and the rise in gonadotropin levels induced by administration of LH-releasing hormone were increased. Between ages 5 and 10 years, basal levels of FSH and LH and LH-releasing-evoked responses were less than those of younger patients with gonadal dysgenesis (see Fig. 22-42). In some patients between ages 6 and 10 years, both FSH and LH concentrations and the LH-releasing-induced gonadotropin responses were comparable to those in normal children. After age 11 years a striking rise in basal and readily releasable LH and FSH levels was observed. Therefore, between the ages of 5 and 10 years, basal and LH-releasing-evoked gonadotropin responses may not reflect the functional status of the gonads in all patients with gonadal dysgenesis.

Astragalian strephos is the rule in 45,X gonadal dysgenesis, exceptions have been documented. Primary follicles have been described in the gonadal ridges of some 45,X individuals at adolescence, and this correlates with the rare occurrence of menarche and a variable but attenuated period of regular menses. In a large study of Turner’s syndrome patients older than 12 years, including, in addition to 45,X patients, those with X chromosome mosaicism and structural abnormalities, spontaneous pubertal development occurred in 16%, including menarche at a mean age of 13 years. Regular menses occurred 9 years post menarche in just over one third of this group. The presence of a second X chromosome was much more evident in those with spontaneous puberty. The figure for spontaneous pregnancy appears to be 3% to 5%. In addition to variability in the rate of follicular atresia, another possible explanation for the presence of
the section on clinical aspects. The deviation in parental origin of the retained X chromosome in 45,X individuals is a consequence of mitotic nondisjunction, which occurs in 15% to 30% of patients. The prevalence of thyroid antibodies and hypothyroidism (or hyperthyroidism) increases during childhood and adolescence and is the presence of a cryptic Y cell line. Males with a 45,X karyotype have a Y to X chromosome or a Y-autosome translocation involving variable segments of the euchromatic (sex-determining) region of the Y chromosome. Translocations have been reported involving the short arm of chromosomes 5, 14, 15, and 18 and the X chromosome. Most patients have had either minor or major anomalies not usually associated with the syndrome of gonadal dysgenesis, such as the cri-du-chat syndrome. These additional anomalies are no doubt related to the autosomal involvement in the translocation and to the degree of deletion involved.

Incidence in Abortuses, Newborns, and Twins.

The incidence of gonadal dysgenesis is approximately 1 per 2000 live female births, and approximately 50% of the patients have a 45,X karyotype. There is, in addition, a considerable loss of 45,X embryos and fetuses. About 7% of spontaneous abortuses have a 45,X constitution. It is estimated that the frequency of 45,X zygotes is 2%, probably the most common chromosome anomaly in humans but fewer than 1% of 45,X conceptuses survive to term. Hook and Warburton have analyzed chromosome karyotypes in embryonic and fetal deaths and demonstrated a significant disparity between the 45,X karyotype and those with mosaicism and/or an isochromosome for the long arm of the X chromosome (Xq;). They postulated a “fetoprotective” effect of more than one dose of some locus or loci on the long arm of the X chromosome.

Turner's syndrome carries a threefold increase in mortality and a reduction in life expectancy of 6 to 13 years; it is less in 45,X patients than in non-45,X Turner syndrome.

The incidence of autoimmune disorders is increased; the most prevalent is autoimmune thyroiditis and Graves' disease, which occurs in 15% to 30%. The prevalence of Crohn's disease and ulcerative colitis is increased and the inflammatory bowel disease is often severe.

As discussed earlier, congenital renal anomalies are common (about ninefold greater than the general population) and the risk of the obstructive uropathy and pyelonephritis is greatly increased. In addition, vascular malformations involving the kidney are more prevalent.

During childhood, problematic otitis media is common and may result in conductive hearing loss. Abnormalities in the growth of the temporal bone, condylar cartilage, and sphenoid synchondrosis result in an abnormality in the positioning of the external auditory meatus and the relation of the middle ear to the eustachian tube in patients with the syndrome of gonadal dysgenesis. These changes, along with abnormalities in the shape of the palate, are thought to be responsible for the increased incidence of otitis media. Sensorineural deafness is present in about two thirds of adult patients. The frequent episodes of otitis media and the sensorineural hearing loss are independent variables in gonadal dysgenesis; the sensorineural hearing loss may be related to loss of genes on the X chromosome responsible for the gonadal dysgenesis phenotype.

In a Danish study, Turner's syndrome women had a fivefold increased risk of developing cancer of the colon and rectum. As noted previously, the prevalence of anorexia nervosa is increased; its onset usually coincides with the initiation of estrogen treatment.

Origin of 45,X Constitution and Phenotype.

A 45,X chromosome constitution (see Fig. 22-1A and Fig. 22-4) may be a consequence of nondisjunction or chromosome loss during gametogenesis in either parent that results in a sex or ovum lacking a sex chromosome. Although errors of mitosis in a normal zygote often lead to mosaicism, a pure 45,X constitution may arise at the first cleavage division from anaphase lag with loss of a sex chromosome or, less likely, from mitotic nondisjunction with failure of the complementary 47,XXX or 47,XYY cell line to survive. However, the percentage of aborted fetuses with a 45,X karyotype is the same as in liveborn infants. Therefore, imprinting does not appear to affect fetal survival in 45,X individuals. Although some studies showed no effect of imprinting on phenotype, 45,X patients appear to have an increased prevalence of cardiovascular anomalies and webbed neck; furthermore, the height in 45,X correlates more strongly with maternal height than with midparental height. A putative imprinted gene on the maternal X chromosome, which affects social cognition, is discussed earlier in the section on clinical aspects.

The very high embryonic and fetal mortality of 45,X conceptuses, as opposed to those with 45,X/46,XX or 45,X/46,XY mosaicism or a structurally abnormal X
mosaic for all or part of the Y chromosome. The similar results obtained by Binder and colleagues indicate a low level of Y chromosome mosaicism in 45,X and chromatin-positive individuals with gonadal dysgenesis. The significance of low-percentage mosaicism for Y chromosome material and its relation to gonadal differentiation and the risk of malignancy are still to be determined. Cryptic Y chromosome identification by sensitive PCR techniques may, however, not show concordance between clinical signs of hyperandrogenism and molecular studies on gonadal tissue. In such Y-negative gonads there is hirsus cell hyperplasia. The routine screening of all Turner's syndrome patients for the presence of SRY is not recommended.

The classic gonadal dysgenesis phenotype is usually associated with absence of all or a proximal portion of the short arm of the X or Y chromosome. The haploinsufficiency of genes in these loci that are not inactivated is postulated to cause the phenotype. Page and co-workers suggested that the genes that encode ribosome protein S4X (RPS4X), and its homologue on the Y chromosome (RPS4Y), are candidate genes. However, the location of this gene on the long arm of the X chromosome and the fact that it is expressed in patients with 46,XXp and 46,XY karyotypes make it an unlikely candidate. Zimm and associates using both cytogenetic and molecular analyses found evidence of a critical region, Xp11.2p22.1, that included loci for short stature (in addition to the more distal SHOX gene), gonadal failure, high arched palate, and thyroid autoimmunity.

Ogata and Matsuo postulated that the gonadal dysgenesis phenotype is multifactorial in origin. According to their construct, the phenotype is related to (1) quantitative loss or alteration of euchromatic or noninactivated genes (haploinsufficiency), leading to global nonspecific developmental defects; (2) haploinsufficiency of pseudoautosomal and/or Y-specific growth genes and lymphogenic genes, resulting in short stature and the "deformative" stigmata; and (3) oocyte loss and gonadal dysgenesis from impaired or failed chromosome pairing during meiotic prophase.

The underlying cause of this sex chromosome abnormality is not known. An increased frequency of thyroid autoimmunity in patients with the syndrome of gonadal dysgenesis and in their parents suggests that the genetic predisposition to develop autoantibodies in one or both parents is associated with an increased prevalence of the 45,X constitution and other chromosome abnormalities in the offspring. Infants with the syndrome of gonadal dysgenesis have been born after artificial insemination and in vitro fertilization. Familial occurrence of 45,X gonadal dysgenesis is rare.

**Diagnosis and Treatment**

Phenotypic females with the following features should have a karyotype analysis: (1) short stature (>2.5 SD below the mean height value for age), (2) somatic stigmata associated with gonadal dysgenesis, and (3) delayed adolescence with increased plasma or urinary gonadotropin levels. Although determination of the X chromatin pattern (Barr body) is a rapid method of screening, karyotype analysis is the definitive procedure. The concentration of plasma FSH and inhibins is useful in assessing the functional status of the gonads.

Even though many severely affected 45,X individuals with prominent dysmorphic features (lymphedema, loose folds of skin over the back of the neck) or coarctation of the aorta are recognized at birth or early infancy, the diagnosis in the less severely affected may be delayed until childhood when short stature is evident or until the age of puberty when secondary sex characteristics fail to appear. It is important to obtain a karyotype on all girls with unexplained short stature especially those with even subtle dysmorphic features of the syndrome of gonadal dysgenesis. Saverständti and Davenport found that (excluding...}

### TABLE 22-10 -- Suggested Follow-up of Adults with Syndrome of Gonadal Dysgenesis (Turner’s Syndrome)

<table>
<thead>
<tr>
<th>Baseline</th>
<th>Annual</th>
<th>35 Yearly</th>
</tr>
</thead>
<tbody>
<tr>
<td>Karyotype</td>
<td>Physical examination (e.g., BMI, blood pressure, CVS)</td>
<td>Echocardiography</td>
</tr>
<tr>
<td></td>
<td>Renal and pelvic ultrasound</td>
<td>Bone densitometry</td>
</tr>
<tr>
<td></td>
<td>Thyroid function</td>
<td>Audiogram</td>
</tr>
<tr>
<td></td>
<td>Echocardiography</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fasting lipids</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Thyroid autoantibodies</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fasting blood glucose</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gonadotropins</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Liver function</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Renal function</td>
<td></td>
</tr>
<tr>
<td>BMI, Body mass index; CVS, cardiovascular system.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>TABLE 22-10 -- Suggested Follow-up of Adults with Syndrome of Gonadal Dysgenesis (Turner’s Syndrome)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>Karyotype</td>
</tr>
<tr>
<td>Renal and pelvic ultrasound</td>
</tr>
<tr>
<td>Echocardiography</td>
</tr>
<tr>
<td>Thyroid autoantibodies</td>
</tr>
<tr>
<td>Gonadotropins</td>
</tr>
<tr>
<td>Renal function</td>
</tr>
<tr>
<td>BMI, Body mass index; CVS, cardiovascular system.</td>
</tr>
</tbody>
</table>


those in whom the diagnosis was made in infancy) overall delay in diagnosis was 7.7 years. Early diagnosis is important for more optimal long-term management, treatment, and counseling, including, if appropriate and with informed parental consent, the use of recombinant hGH treatment before the child falls below -2.0 SD in height, to achieve better growth in childhood and potentially normal adult stature.

The following studies should be done in women when the diagnosis of the syndrome of gonadal dysgenesis is made: an intravenous pyelogram or ultrasonographic examination to exclude a renal anomaly; an echocardiogram or MRI study to assess cardiovascular function; periodic hearing examination and evaluation of thyroid function and thyroid antibodies; regular measurements of plasma glucose levels after adolescence; and monitoring for scoliosis and bone density in late adolescence and adulthood for evidence of osteopenia. Guidelines for the diagnosis and management for Turner's syndrome in childhood and adulthood have been updated following an international consensus workshop and reviewed by Elsheikh and co-workers. There is particular emphasis on long-term monitoring for cardiovascular disease (including hypertension, aortic dilatation and the risk of aortic dissection), regular screening for thyroid dysfunction, recognition of hearing impairment which worsens in adult life and early planning for any request for assisted conception. Table 22-10 presents the suggested follow-up of adults with Turner’s syndrome.

Therapy is directed toward augmenting stature, correcting somatic anomalies, inducing secondary sexual characteristics and menses, and counseling. As noted, the short stature in gonadal dysgenesis is not related to a deficiency of growth hormone, insulin-like growth factors, thyroid hormone, or adrenal or gonadal steroids. However, administration of pharmacologic doses of biosynthetic hGH increases growth rate and augments final height by a mean of 5 to 10 cm. The heterogeneity in response appears to be related to the chronologic age at the start of therapy, the duration of therapy, the dose and frequency of growth hormone administration, the use of oxandrolone and/or estradiol, the growth standards used, the height of the parents, and the growth hormone peak elicited by pharmacologic stimulation. Rosenfeld and co-workers have the longest and most extensive study. Starting in 1983, 70 patients between the ages of 4.7 and 12.4 years with normal growth hormone responses to provocative stimuli were studied. They were randomly assigned to (1) a control group for 1 year; (2) the anabolic steroid oxandrolone at a dose of 0.125 mg/kg by mouth daily; (3) growth hormone 0.125 mg/kg subcutaneously three times per week; or (4) a combination of oxandrolone and growth hormone. After 12 to 24 months, all groups except the growth hormone alone group were placed on combination therapy; however, the oxandrolone dose was lowered to 0.0625 mg/kg because of signs of virilization. After 3 years, most patients received daily growth hormone rather than thrice weekly; however, the dose of 0.375 mg/kg per week was unchanged. The mean height in patients who completed therapy was 151.7 cm, for a net mean gain of 9 cm over projected height and historical controls. Long-term studies now indicate that when growth hormone treatment alone is started early enough, most girls will benefit and some will achieve normal final height.
Attaining a final height of 150 cm is now a realistic target. 14 It is important to individualize the dose of growth hormone depending on the patient's growth response to the usual dose (0.375 mg/kg per week in six or seven divided doses). Growth hormone treatment should be considered when the patient's height falls below -2.0 SD on the growth curve for normal girls.

Growth hormone therapy is approved by the U.S. Food and Drug Administration for the treatment of patients with Turner's syndrome, and it was approved earlier in Japan and many Western European countries. It is prudent to discuss growth hormone therapy with the parents and patients, including its efficacy and side effects, in all patients with gonadal dysgenesis whose height is more than 2.0 SD below the mean value for age, especially those whose growth rate is less than 5 cm/year. Studies are in progress to evaluate the effects of initiating growth hormone therapy at 5 to 6 years of age before the growth deficit is severe. Growth hormone therapy is usually continued until the growth rate falls to less than 2 cm/year or the bone age exceeds 15 years. Supraphysiologic doses of growth hormone induced insulin resistance but not hypoglycemia; the increased insulin values returned to the normal range when growth hormone treatment was discontinued. 14

Estrogen therapy has commonly been deferred until age 15 or later on the assumption that treatment at an earlier age leads to rapid skeletal maturation and diminished height. This premise was based largely on the fact that pharmacologic doses of estrogens can accelerate bone maturation and lead to premature epiphyseal fusion without a proportionate increase in height. Studies in patients with aromatase deficiency and in a patient with a mutation in the estrogen receptor indicate that estrogen rather than androgen is the principal gonadal hormone involved in bone maturation and fusion and bone mineral accretion. 15 16 17 We examined the effect of early low dose, conjugated estrogen therapy on linear growth, bone age, and the development of secondary sexual characteristics in a group of patients with gonadal dysgenesis. 18 Low-dose conjugated estrogens (9 µg/kg body weight per day) or ethinyl estradiol (141 ng/kg body weight per day) was given to 21 patients with the syndrome of gonadal dysgenesis who had a mean age of 13 years and mean bone age of 10.7 years. Growth rate was transiently accelerated but declined to below the pretherapy rate after 12 months of therapy. The final height of the patients treated with low-dose estrogen was not different from that of control nontreated patients or that of a group of six girls with Turner's syndrome in whom normal ovarian function was present and spontaneous puberty ensued. Hence, no increase or decrease in mean final height was noted in our study. However, girls who received estrogen with a bone age of less than 11 years were shorter than the girls who began low-dose estrogen therapy after a bone age of 11 years. 19 Similar results have been obtained subsequently in other studies. 20 21 Initially it is important to use the lowest dose of an estrogen preparation (including an estradiol patch) 22 that gradually will induce pubertal development.

Serious psychological effects are frequently associated with a prolonged delay in the treatment of sexual infantilism. 23 The institution of low-dose, conjugated estrogen, synthetic estrogen therapy or transdermal estradiol patch 24 alone at approximately 13 years of age (bone age >11 years) elicits a brief growth spurt without inordinate advancement of skeletal maturation or reduction in final height and induces the development of secondary sexual characteristics at an age comparable to that of normal peers, thereby obviating the undesirable psychological consequences and deficient bone mineralization of a prolonged delay in sexual maturation. Studies in which growth hormone treatment was combined with early estrogen therapy (i.e., in patients younger than 12 years of age) indicate a shorter final height than those patients treated with the growth hormone alone. 25 26 The number of years of growth hormone therapy before estrogen therapy is a critical factor in predicting height gained, and hence the time of initiation of estrogen therapy in the growth hormonetreated patient has an important influence on final height. 27 Earlier introduction of growth hormone therapy or the use of higher doses of growth hormone or both to induce normal or near-normal height for age 28 29 permits the initiation of estrogen therapy by about 13 years of age, an important psychologic consideration. Therefore, in the treatment of girls with Turner's syndrome, the goal of increased adult height must be balanced against the desire for sexual maturation in each individual patient.

Estrogen replacement therapy may improve certain neuropsychologic deficits (nonverbal processing speed, motor function, and memory). 30 31 However, the neurocognitive deficits in adult women with Turner's syndrome were not altered significantly by estrogen replacement. 32 33 A number of instances of endometrial carcinoma have been reported in patients with gonadal dysgenesis. 34 The evidence suggests that estrogens, especially when unopposed by progesterone, can produce a progression of histologic changes from endometrial hyperplasia to invasive carcinoma (see also Chapter 16). To clarify the relation between estrogen therapy and endometrial pathology in gonadal dysgenesis, Rosenwaks and colleagues 35 36 studied 41 patients receiving estrogen replacement therapy. Increased risk of abnormal endometrial histology correlated with (1) a lifetime dosage of conjugated estrogens of more than 2500 mg, (2) more than 7 years of estrogen therapy, and (3) a daily dosage of conjugated estrogens greater than 1.25 mg. Progestagens can modify the effect of estrogens on endometrial histology. It is therefore prudent to treat patients with gonadal dysgenesis with low-dose cyclic estrogen replacement therapy, with progestagen added at the end of each cycle. Further studies are necessary to assess the optimal dose of estrogen that reduces the risk of endometrial carcinoma while concurrently preventing osteoporosis.

Rarely, patients with a 45,X karyotype and no cytogenetic evidence of Y chromosome material develop gonadoblastomas. 37 However, a study employing multiple Y-specific DNA probes indicates that 3.4% of apparent 45,X patients have Y chromosomal material present. 38 These 45,X patients may be at risk for gonadoblastoma formation. Most patients with a 45,X karyotype have little or no risk of neoplastic transformation of the streak gonads.

Replacement Therapy.

We routinely initiate therapy (depending on the height) at about 13 years of age with 0.3 mg of conjugated estrogen or 5 µg of ethinyl estradiol by mouth or very low dose transdermal 17-estradiol. 39 The oral dose is gradually increased over the next 2 to 3 years to 0.6 to 1.25 mg of conjugated estrogens or 10 µg of ethinyl estradiol daily for the first 21 days of the month. The patient is maintained on the minimal effective dose of estrogen needed to maintain secondary sexual characteristhaps, permit withdrawal bleeding, and prevent osteopenia. Medroxyprogesterone acetate, 5 to 10 mg/day, is given from the 10th through the 21st day of the month to ensure more physiologic menses and to reduce the risks of endometrial and breast cancer. There is only limited clinical experience on the use of transdermal estradiol patch in adolescent girls, but with this approach (while more expensive) the natural estradiol reaches the systemic circulation directly without first undergoing metabolism by the intestine and liver. 40 The common late adolescent and adult oral dose of estrogen replacement therapy often fails to increase to normal the size of the uterus (especially an adult fundal-cervical ratio as assessed by pelvic ultrasonography). The attainment of a mature heart-shaped uterine configuration is important only if the patient elects to become pregnant by oocyte donation and in vitro fertilization. 41

An important part of the management is the education of the patient and family. 42 A frank discussion with the parents of the pathophysiology of the condition is appropriate when the diagnosis is made and reinforced at later sessions. Thereafter, the child should be given as much information about her condition as she can comprehend to allay any false fears or anxieties. An honest assessment of reproductive function based on clinical findings as well as hormone levels should be given to the patient when appropriate. Advances in in vitro fertilization and embryo transplantation make pregnancy possible for these patients, but the miscarriage rate is high 43 and the risk of aortic rupture is increased during pregnancy. The importance of medical and psychosocial management throughout life in patients with the syndrome of gonadal dysgenesis must be emphasized. Social and psychosocial support from the parents and the physician usually results in a well-adjusted woman.

Partial Sex Chromosome Monosomy and Clinical Variants of the Syndrome of Gonadal Dysgenesis

Partial sex chromosome monosomy may or may not modify the expression of the classic 45,X phenotype. 44 45 Forty to 50 percent of patients with the typical syndrome of gonadal dysgenesis are X chromatin positive. This group usually has a structurally abnormal X chromosome or sex chromosome mosaicism involving a 45,X cell line. Chromatin-positive and chromatin-negative variants of gonadal dysgenesis are discussed here in relation to the more usual types of sex chromosome abnormalities to which they may be associated. The diagram in Figure 22-44 shows the variable effect of partial sex chromosome monosomy (haploinsufficiency) on the typical features of the syndrome.

In patients with sex chromosome mosaicism, the ratio in each gonad, 45,X primordial germ cells and blastemal components to those with a normal 46,XX or 46,XY constitution is probably the major determinant of whether the ultimate gonadal structure is a streak, a dysgenetic or hypoplastic ovary, or testis. 46 47 48 We believe the evidence supports the idea that, after X-ray, the primitive gonad, primordial germ cells that bear a 45,X constitution degenerate more rapidly, quite likely by apoptosis, than do 46,XX cells, resulting in a streak, hypoplastic, or normal ovary. 49 Similarly, if the gonadal blastemal components, in particular the Sertoli cells, do not contain an appropriate number of 46,XY cells, testicular development does not take place. 50 51

The quantitative relation between 45,X cells and those with a 46,XX or 46,XY pattern in peripheral tissues may also be responsible for the variable effect of mosaicism on stature and associated somatic stigmata. 52
In patients with a single cell line (euploid) containing a structurally abnormal sex chromosome, the somatic and gonadal consequences appear to be related to the nature and degree of the short or long arm deficiency of the second X or Y chromosome. Table 22-11 summarizes the correlation between structural abnormalities of the X and Y chromosomes and the clinical manifestations. The use of deletion mapping of the human sex chromosomes to clarify the relation of phenotype to karyotype has limitations. Structural abnormalities are often associated with mosaicism because of loss of the structurally abnormal sex chromosome from the stem cell line. Furthermore, structural rearrangements of chromosomes are complex. However, the advent of chromosome banding and molecular genetic techniques has facilitated the analysis of structurally abnormal sex chromosomes. At present, the data suggest that (1) ovarian determinants are located on both the long and short arms of the X chromosome and that patients with short arm deletions proximal to band Xp21 or long arm deletions proximal to band Xq25 usually have streak gonads and sexual infantilism; and (2) the short arm of the X chromosome (and to a lesser extent the long arm) contains loci that, if deleted, result in short stature and the somatic stigmata of the syndrome of gonadal dysgenesis (see section on biologic functions of the X chromosome and Y chromosome).

45,X/46,XX, 45,X/47,XXX, and 45,X/46,XX/47,XXX Mosaicism.

45,X/46,XX mosaicism is the most common finding in patients with chromatin-positive gonadal dysgenesis and is second in frequency only to 45,X; the mosaic karyotype arises through loss of one X chromosome in a 46,XX conceptus. Patients with this form of mosaicism usually exhibit fewer of the associated somatic anomalies, are not invariably short, may menstruate, and may even be fertile. One gonad may be of the streak type, and the contralateral gonad may be either a hypoplastic or a normal ovary; alternatively, both ovaries may be either normal or hypoplastic. During a family survey for a leucocyte anomaly, a normal grandmother with 45,X/46,XX/47,XXX mosaicism was discovered fortuitously. Some appreciation of the variable clinical features may be gleaned from nine patients with these forms of mosaicism studied by Morishima and Grumbach. All had normal female external genitalia. Of seven who attained pubertal age, four developed some female secondary sexual characteristics, and two were menstruating regularly. One of the two had had three pregnancies. In some, no important somatic abnormalities were detected, and two were of normal stature. One of the 45,XY/46,XX patients had a webbed neck, coarctation of the aorta, and other gonadal dysgenesis stigmata but was of normal stature and menstruated regularly. A 12-year-old 45,X/46,XX/47,XXX patient had primary hypothyroidism and autoimmune thyroiditis.

Patients with the Xqi structural abnormality (isochromosome for the long arm of the X) have an X chromosome that consists primarily of two long arms (Xq) and lacks a short arm (Xp); it arises mainly as a consequence of a break in sequence in the proximal short arm and not by centromere misdivision (Fig. 22-45) and occurs in about 15% of Turner's syndrome individuals (about 1 in 13,000 female live births). The Xqi chromosomes may be either monocentric or dicentric X. In a review of 89 cases, 29 were monocentric. Of these, only 5 of 17 were associated with mosaicism for a 45,X cell line. In contrast, 49 of 60 patients with a dicentric isochromosome had a 45,X cell line. Dicentric X isochromosomes are more unstable than monocentric forms and probably result more frequently in sex chromosome mosaicism through loss of the heteromorphic dicentric X chromosome. In 14 patients studied with molecular biologic techniques, Xp markers were found in three dicentric Xq chromosomes and in three monocentric Xq chromosomes. In five instances the Xqi was paternally derived. Isochromosome for the long arm of the X is the most common form of structural rearrangement of the X chromosome.

Patients with a long arm X isochromosome are invariably short and have streak gonads, although some menstruate spontaneously. In general, the somatic stigmata of gonadal dysgenesis are less severe than in 45,X patients. Coarctation of the aorta and severe lymphedema of the hands and feet are less common in 46,XXX patients. Webing of the neck, if present, is usually slight. The findings indicate that absence of the short arm on the second X, even in the presence of an X chromosome composed of two long arms, leads to short stature, failure of ovarian development, and some somatic stigmata of gonadal dysgenesis. The prevalence of autoimmune thyroiditis, decreased glucose tolerance, and inflammatory bowel disease may be higher in patients with structural abnormalities of the X chromosome, especially 46,X,qX, than in 45,X individuals.

Structurally abnormal X chromosomes are usually late replicating (except in balanced X-autosome translocations), and they give rise to the X chromatin body.

46,Xqpi.

There is controversy about the existence of an isochromosome for the short arm of the X chromosome. Of the 11 reported cases, 3 have been revised to long arm deletions, 4 were reported as presumptive, and 2 have been questioned on cytogenetic grounds. The controversy revolves around the difficulty in distinguishing Xp1 from deletions of the long arm.
of the X chromosome, because the banding pattern of Xp is quite similar to that of Xq from the centromere to Xq24. There have been no reports of either high-resolution chromosome banding or molecular genetic analysis in a patient with 46,XXpi.

46,XX or 45,XXI,XXii.

A ring X chromosome usually occurs as part of 45,XX/46,XX mosaicism or a more complex karyotype (see Fig. 22-45). Short stature is present in most patients, and most have minor stigmata of gonadal dysgenesis; some have a webbed neck or coarctation of the aorta. Approximately one third have spontaneous menses and develop secondary sexual characteristics. A mother and daughter with 45,XX/46,XX have been described. Although most patients with 45,XXI,XXI have the gonadal dysgenesis phenotype, patients with severe mental retardation, syndactyly, and abnormal facies have been reported. XIST, a gene that is transcribed only by the inactive X chromosome, is usually not expressed in these severely affected patients. All genes analyzed on the proximal short and long arms of the ring X were expressed, suggesting that the abnormal phenotype in these patients is caused by disomy for these genes on the X chromosome resulting from lack of dosage compensation. A ring X chromosome is associated with significant learning difficulties and some behavior problems. The degree of mental impairment is usually more severe in patients with a ring X chromosome than in those with active X chromosomes. The molecular phenotype associated with active or inactive ring X chromosomes confounds predicting prognosis in an affected fetus detected by prenatal diagnosis.

The proportion of X chromatin-positive cells is decreased in patients with a ring X chromosome, and the X chromatin bodies tend to be small. The ring X chromosome, with rare exceptions, exhibits late DNA replication.

The ring X chromosome arises by loss of both ends (telomeres) of the chromosome and union of the proximal breaks; as a consequence, a variable amount of chromosomal material is lost from each arm (see Fig. 22-45). Ring chromosomes are unstable, and the size of the ring varies in different cells of the same subject. In relation to gonadal dysgenesis, studies of patients with a ring X chromosome have established that loss of both telomeres of an X chromosome need not lead to the development of streak gonads.

It is sometimes difficult to be sure of the cytogenetic origin of the ring chromosome, a distinction that is critical in view of the increased risk of gonadal tumors associated with dysgenetic gonads and Y cell lines. Molecular cytogenetic analysis with specific X and Y chromosome probes has made identification easier.

Deletions of the short arm of the X chromosome (Xp) are rare and are frequently associated with 45,X mosaicism. Phenotypic-karyotypic analysis of 40 nonmosaic patients indicated considerable variation in somatic stigmata and gonadal function. Patients with a terminal deletion of the short arm of the X (distal to Xp22) can have normal ovarian function and no somatic stigmata of gonadal dysgenesis with the possible exception of a modest degree of short stature. Patients with deletions proximal to Xp22 usually have short stature, variable stigmata of gonadal dysgenesis, and gonadal dysfunction (Fig. 22-46). A lymphedema critical region is proposed at Xp11.4.

The abnormal X chromosome is usually the late DNA-replicating X and is responsible for the small X chromatin body in interphase nuclei in these patients. Of interest is the report of a familial group of seven patients with the syndrome of gonadal dysgenesis secondary to a deletion of the short arm of an X chromosome in which the disorder was transmitted by carriers of a balanced translocation between the X and chromosome 1.

Table 22-11 -- Relation of Structural Abnormalities of X and Y to Clinical Manifestations of Gonadal Dysgenesis

<table>
<thead>
<tr>
<th>Karyotype</th>
<th>Pheno-</th>
<th>Sexual</th>
<th>Short</th>
<th>Somatic Anomalies of Syndrome of Gonadal Dysgenesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loss of an X or Y</td>
<td>45,X</td>
<td>Female</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Isochromosome for long arm of an X</td>
<td>46,X(Xq)</td>
<td>Female</td>
<td>(occ. 1)</td>
<td>+</td>
</tr>
<tr>
<td>Deletion of short arm of an X</td>
<td>46,X,del(X)(p21)</td>
<td>Female</td>
<td>+, ±, or -</td>
<td>(-)</td>
</tr>
</tbody>
</table>
| Deletion of long arm of an X | 46,X,del(X)(q21) | Female | + | - (?)
| Deletion of both arms of an X (ring X) | 46,Xr(X)(q22;q25) | Female | or ± | + |
| Loss of short arm of Y | 46,del(Y)(p11) | Ambiguous | + | + |

X chromosome in which the disorder was transmitted by carriers of a balanced translocation between the X and chromosome 1. Familial Xp deletions have been detected in seven additional families with a variable phenotype and short stature as the only major phenotypic abnormality.

46,Xq and 45,XII,XX.

Patients have been reported with a deletion of the long arm of the X chromosome (Xq). In general, patients with only a 46,XXq cell line are normal in stature or have moderately short stature and exhibit few manifestations of gonadal dysgenesis but have primary amenorrhea, sexual infantilism, and streak gonads. We have studied two patients with 46,XXq karyotypes, and the findings in one are summarized in Figure 22-47. Exceptions to the rule that 46,XXq patients lack stigmata of gonadal dysgenesis and are of normal height were reported before chromosome banding techniques became available. Such cases may represent either hidden mosaicism or complex structural rearrangements of the X chromosome, including inversions and interstitial rather than terminal deletions. Studies with FISH and Southern blotting suggest that gonadal dysgenesis stigmata are absent when the break point is distal to Xq24 and that deletions distal to Xq25-q28 do not reduce stature. Most Xq patients sooner or later have ovarian failure or premature menopause irrespective of the size of the deletion.

Isodicentric X.

Isodicentric X chromosomes are large X chromosomes with two C bands. These chromosomes replicate

Figure 22-45 Structural abnormalities of the X chromosome. The normal X at the left is G banded. A dark band on the short arm and two major dark bands on the long arm are visible. The first Xq and the ring X chromosome (Xr) are not banded. They show late replication with tritiated thymidine. Note symmetry of the arms of the second Xq. Even with G banding, it is difficult to distinguish this chromosome from a possible short arm isochromosome. The long arm isochromosome (Xq) appears to be dicentric. The two chromosomes to the far right are isodicentric X chromosomes. Both have two C bands but only one functional centromere. There is a mirror-like band pattern on both sides of a point between the two C bands. The first isodicentric X chromosome presumably represents a break in the long arm of X at q22 with fusion of chromatids and duplication of the entire chromosome. The second isodicentric X chromosome appears to represent a terminal break in the short arm so that reduplication of the chromatids has produced what appears to be almost two X chromosomes.

late, form a large bipartite sex chromatin body, and apparently have one functionally suppressed centromere (see Fig. 22-45). The banding pattern of isodicentric X chromosomes reveals a mirror image about a point between the two centromeres (C bands). These chromosomes are usually associated with mosaicism for a 45,X cell line and presumably arise by chromatid break and fusion of sister chromatids. This event would produce an aneural fragment that would be lost during cell division and thereby result in a 45,X cell line. Phenotypic-karyotypic correlations in these patients are similar to those in Xp and Xq patients.
X-Autosome Translocations.

X-autosome translocations have been reviewed. In general, women with a break in the X chromosome between Xp13 and Xp26 manifest infertility, confirming the belief that this region contains genes critical to gonadal differentiation and function. Male carriers of a balanced X-autosome translocation with an X chromosome break in the "critical region," Xp13 to Xp26, are usually infertile.

X Chromatin-Negative Variants of Gonadal Dysgenesis

The pattern of sex chromosome mosaicism and structural abnormalities of the Y chromosome is similar to that for the X chromosome. Usually, as a consequence of its effect on gonadal differentiation, a Y-bearing cell line modifies the typical female phenotype of the syndrome by causing a variable degree of masculine differentiation of the genital tract.

A highly diverse phenotype is encountered in these forms of mosaicism, ranging from phenotypic females to individuals with ambiguous external genitalia to phenotypic males. As in 45,X/46,XX mosaicism, short stature and the associated somatic abnormalities, although frequently present, are inconsistent features and may vary independently of each other and of gonadal differentiation. In the review by Zah and co-workers of 60 patients with 45,X/46,XY mosaicism, two thirds were reared as females.

Of nine patients with 45,X/46,XY or 45,X/46,XY/47,XXY mosaicism studied by Morishima and Grumbach, one was a phenotypic female, one was a phenotypic male, and seven had ambiguous external genitalia (Table 22-13). The gonads varied from bilateral streaks in the phenotypic female to bilateral dysgenetic testes. In others the development was asymmetric: one patient had a streak in one mesosalpinx and a rudimentary testis on the contralateral side (so-called mixed gonadal dysgenesis); another had a normal testis in the scrotum and a streak in the hernia. The proportion of X chromatin bodies that were slightly small. Plasma gonadotropin levels were elevated: LH was 5.6 µg/L (LER-960), and follicle-stimulating hormone level was 36.5 µg/L (LER-869). The buccal smear showed a normal proportion of X chromatin bodies in interphase nuclei, which were conspicuously small. Karyotype analysis and autoradiography revealed a 46,X0p karyotype. The X chromosome that had been deleted close to the centromere, but a small segment of the short arm was visible distal to the centromere. C. A 20-year-old phenotypic female with a chief complaint of dysfunctional uterine bleeding. She had short stature, slight puffiness of hands and feet, and short fourth metacarpals. Female secondary sexual characteristics appeared at age 11, and menarche at age 13 was followed by regular menses, which later became irregular. The buccal smear contained nuclei with a normal proportion of small sex chromatin bodies. Bilateral ovaries were identified grossly and histologically during an appendectomy. Karyotype was 46,X0p. The extent of deletion of the short arms of the abnormal X chromosome in this patient is less than that seen in patients in A and B. A segment of the short arm is readily discernible above the centromere. It appears that, in these three patients with X0p karyotypes, somatic and gonadal manifestations of the syndrome of gonadal dysgenesis correlated with the magnitude of deletion of the short arm of the X chromosome.

45,X0,X, 45,XY, 45,X0XY, 45,X0XY/47,XXY, and Related Abnormalities (Table 22-12)

A brief review of the chromosome analyses of 58 patients in this category is presented. Presence of an 80% proportion of X chromatin in the buccal smear indicated a mosaicism with one cell line of greater than 20%. Habitus:

| Karyotype: | 45,X/46,XY |
| Genitalia: | Female ambiguous male with normal gonadal function |
| Wolffian duct derivatives: | Duct differentiation, contingent on functional integrity of homolateral fetal gonad (i.e., streak) |
| Müllerian duct derivatives: | Uterus, fallopian tubes, dysgenetic testes variable structures; testes wolffian duct derivatives |
| Gonad: | Streak gonads dysgenetic testes normal testes; Streak gonad + dysgenetic testis "mixed gonadal dysgenesis"; risk gonadal neoplasm (gonadoblastoma) |
| Habitus: | Short, stigmata of gonadal dysgenesis; genitalia: streak gonads female with sexual infantilism at puberty; dysgenetic testes ambiguous genitalia; if gonadoblastoma present gynecomasia secondary to estradiol production; testes normal male differentiation |
| Hormone profile: | Increased plasma FSH and LH; decreased testosterone concentrations |

FSH, follicle-stimulating hormone; LH, luteinizing hormone.

During screening of a family for bone marrow transplantation donors, we discovered an adult with 45,X/46,XY mosaicism. Except for short stature, no stigmata of the syndrome of gonadal dysgenesis were present, and he had normal adult male genitalia (Fig. 22-49). Plasma gonadotropin and testosterone levels, both basally and in response to LH-RH, were within the normal range for adult men. On pelvic ultrasonography, müllerian duct derivatives were absent, and the testes appeared normal and homogeneous. The sperm count was 17 million/mL, and the patient demonstrated fertility, as evidenced by the fathering of a "normal" 46,XY fetus. The finding of a short, otherwise normal fertile male with 45,X/46,XY mosaicism extends the phenotypic spectrum of this disorder.

Ascertainment bias may be responsible for the lack of well-differentiated males in reports of this disorder. A review of the chromosome analyses of 58 patients in the literature who were ascertained because of ambiguous genitalia shows a preponderance of 45,X cells, suggesting that only individuals whose abnormality...
Case | External Genitalia | Urogenital Sinus | Phallic Enlargement | Gonads | Female | Genital Ducts | Male \\
--- | --- | --- | --- | --- | --- | --- | --- \\
1 | Female | - | - | Rt. streak? | Rt. fallopian tube? | Uterus | Rt. \\
2 | Ambiguous | + | + | Rt. testis | Rt. fallopian tube | Uterus | Rt. vas deferens \\
3 | Ambiguous | + | + | Rt. not found | Rt. fallopian tube | Uterus | Rt. \\
4 | Ambiguous | + | + | Rt. dysgenetic testis | Rt. fallopian tube | Vestigial uterus | Rt. vas deferens \\
5 | Ambiguous | + | + | Rt. dysgenetic testis | Rt. fallopian tube | Uterus | Rt. vas deferens \\
6 | Ambiguous | + | + | Rt. dysgenetic testis | Rt. fallopian tube | Uterus | Rt. \\
7 | Ambiguous | + | + | Rt. dysgenetic testis | Rt. fallopian tube | Uterus | Rt. vas deferens \\
8 | Ambiguous | + | + | Rt. dysgenetic testis | Rt. fallopian tube | Uterus | Rt. vas deferens \\
9 | Male | - | - | Normal penis | Rt. | Rt. fallopian tube | Vestigial uterus | Rt. \\

Because 45,X/46,XY mosaics not only may harbor gonadoblastomas, roughly 30% of which are associated with germ cell tumors, but also have an increased prevalence of CIS, gonadal biopsy is indicated in all individuals with a male phenotype and 45,X/46,XY mosaicism. If the testis is histologically normal and is in the scrotum or can be placed in the scrotum, it can be retained. However, careful close follow-up is mandatory. Müller, Skakkebaek, and associates recommend ultrasonography of the gonads and biopsy of the retained testis at the start of puberty. If biopsy and ultrasonography show lack of evidence of CIS, they suggest an annual ultrasonographic examination and a second biopsy at 20 years of age. The absence of CIS at age 20 suggests that the risk of a gonadal germ细胞 tumor is minimal. Page and his group proposed the location of a gonadoblastoma susceptibility locus on the Y chromosome detected Y chromosomal material in 12.2%; the occurrence of gonadoblastoma was estimated at 7% to 10% of those with Y chromosome DNA.

The restricted local or paracrine action of the testis on the development of genital ducts is well demonstrated in patients with asymmetric gonadal development. In such patients, development of male ducts and involution of the müllerian structures are also asymmetric and parallel the degree of testicular development on each side. As discussed previously, local action of the testis on müllerian duct regression is mediated through AMH, whereas unilateral differentiation of male ducts is mediated by high local levels of testosterone in the wolffian ducts and their derivatives. The presence of Sertoli cells in the ipsilateral gonad correlates with the absence of müllerian structures on the same side in patients with 45,X/46,XY mosaicism. This observation is consistent with the local secretion of AMH by embryonic and fetal Sertoli cells. Male differentiation of the external genitalia is, however, brought about by the systemic effects of testosterone secreted by a fetal testis and
Figure 22-49. The external genitalia of a normally differentiated male with 45,X/46,XY mosaicism. Karyotype analyses revealed 10% and 68% mosaicism for a 45,X cell line in blood and skin, respectively. Gonadotropin levels, both basal and LHRH-stimulated, and plasma testosterone levels were normal. Fertility was documented in vitro and by the conception of a normal male fetus.

Figure 22-49. 45,X/46,XY mosaicism with a feminizing gonadoblastoma. A. A 20-year-old female with many stigmata of the syndrome of gonadal dysgenesis, including short stature, multiple new, cubitus valgus, and hyperconvex, small nails. The buccal smear was X chromatin negative; on fluorescence microscopy, 30% of interphase nuclei had a single Y body. Karyotype was 45,X/46,XY. The patient had spontaneous development of pubic and axillary hair at age 12. At age 18, breast development was noted. Her height was 139 cm (-6.1 SD) and weight 39 kg (-2.5 SD). Bone age was 17 years; an intravenous pyelogram was normal. The concentration of plasma gonadotropins at 20 years of age was elevated; plasma luteinizing hormone was 8 μU/mL (LER-860) and follicle-stimulating hormone was 50 μU/L (LER-889). The concentration of plasma estradiol was 95 pmol/L (26 pg/mL), and that of estrone was 117 pmol/L (32 pg/mL); the plasma testosterone level was less than 0.7 nmol/L (0.2 ng/mL). On exploratory laparotomy, normal-appearing fallopian tubes and a uterus were found. The right gonad was a typical "streak," with whorls of fibrous connective tissue. B. The left gonad was replaced by a 1.3 × 1.1 × 1.0 cm tumor mass, which, on histologic section, revealed well-defined nests and islands of Sertoli-Leydig-like cells and germ cells, as well as calcification consistent with diagnosis of gonadoblastoma. C. Higher magnification illustrates aggregates of germ cells and small epithelial cells resembling immature Sertoli cells, as well as cells indistinguishable from Leydig cells. After gonadectomy the concentration of plasma estradiol was prepubertal (<18 pmol/L [<5 pg/mL]).

Diagnosis and Treatment.

The presence of functional testicular elements can be detected before puberty by the rise in the concentration of serum testosterone above prepubertal values after a course of hCG and the determination of serum AMH and inhibin B.

In some patients with 45,X/46,XY or 45,X/47,XXY mosaicism, the brightly fluorescent portion of the Y chromosome is absent. Caspersson and colleagues noted the absence of bright fluorescence of the Y chromosome in four of seven patients. In one patient, a Y-to-chromosome-2 translocation during gametogenesis was suspected. In other such patients there was no evidence of translocation or deletion of the Y chromosome. Fluorescence, C banding, and replication of the Y chromosome were altered compared with normal.

Magenis and Donlon studied 12 structurally altered Y chromosomes with a panel of banding techniques. They concluded that the nonfluorescent Y chromosome is an isodicentric chromosome that most likely arises from a chromatic break at the heterochromatic-euchromatic junction on the long arm of the Y chromosome with sister chromatid fusion and duplication of the Y. Subsequent studies with Y-specific DNA probes verified their interpretation but revealed variability in the break point on the Y. Isodicentric chromosomes are more prone to mitotic errors that result in a 45,X cell line. In patients bearing a Y cell line who have "dysgenetic gonads," gonadal extirpation is prudent. Some patients have mosaicism with a cell line containing a minute nonfluorescent chromosome fragment whose exact nature (i.e., either X or Y) is not apparent by standard chromosome banding techniques. Molecular DNA analyses or FISH should be performed in these cases to ascertain the origin of the fragment.

Mixed, asymmetric, or atypical gonadal dysgenesis is a term that has sometimes been used to describe patients with a streak gonad on one side and a testis on the other. Although this association is common in 45,X/46,XY mosaicism, these gonadal findings are not specific for this mosaicism and also occur with a 46,XY karyotype (e.g., in familial 46,XY gonadal dysgenesis).

45,X/46,XY mosaicism probably arises mainly through anaphase lag and is frequently associated with structurally abnormal Y chromosomes. Interchromosomal rearrangements with loss of the structurally abnormal Y may be a common mechanism for the production of 45,X/46,XY mosaicism. 45X/46,XY mosaicism in phenotype female patients with typical features of 45,X Turner's syndrome has been associated with a mutation in the 5' non-HMG box region of the SRY gene. In these patients the SRY mutation apart from the 45,X cell line would have prevented the differentiation of testicular tissue.

Diagnosis and Treatment.

The diagnosis is established by demonstration of 45,X/46,XY mosaicism in blood, skin, or gonadal tissue. A Y chromosome, even one lacking the distal fluorescent portion of its long arm, can be recognized by its size and morphologic appearance (parallel long arms and short, fuzzy short arms) and the presence of a segment of Giemsa 1 positive heterochromatin; it can also be identified by use of Y-specific DNA probes for the centromeric region as well as for the long and short arms. The decision as to the sex of rearing should be based on the potential for normal function of the external genitalia. In patients assigned a female gender role, the gonads should be removed, and the external genitalia should be repaired by clitoral recession, vaginoplasty, and labioscrotal reduction. Estrogen therapy should be initiated at the age of normal puberty to induce female secondary sexual characteristics. In affected infants for whom a male gender assignment is selected, all gonadal tissue except that which appears functionally and histologically normal and is in the scrotum should be removed, and prostatic tests should be placed in the reconstructed scrotal sac if appropriate. In these patients, removal of the müllerian duct remnants is indicated, as is repair of any hypospadias.

As discussed previously, most 45,X/46,XY mosaic males detected by amniocentesis are born with normal male genitalia. In these infants, it is prudent to perform MRI of the pelvis and testes to detect any müllerian structures and any inhomogeneity of the testes suggestive of dysgenesis or neoplasm. Hypothalamic-pituitary-gonadal integrity can be assessed by serial determinations of plasma gonadotropin and testosterone levels. Plasma AMH and inhibin B levels are markers of the functional integrity of the Sertoli cells. If there is hormonal or imaging evidence of testicular dysgenesis, testicular biopsy or gonadectomy (or both) is indicated in infancy. Even in the absence of evidence for testicular dysgenesis, close follow-up is indicated, including a testicular biopsy at puberty and at age 20 to ascertain malignant potential (e.g., carcinoma in situ). The need for androgen replacement therapy at adolescence depends on the capacity of the testis to secrete testosterone.

Structural Abnormalities of the Y Chromosome

Structural abnormalities of the Y chromosome that are of clinical significance are rarer than those of the X chromosome. This may be because the abnormal Y chromosome, being smaller than most structurally abnormal X chromosomes, is more readily lost from the cell during mitosis. The 45,X composition may therefore occur as a consequence of a structural abnormality of the Y chromosome that is lost at an early cleavage division. Deletions of both the long and short arms of the Y chromosome, as well as rings, isochromosomes of both arms, and dicentric chromosomes, have been described, and the PAR regions of the Y chromosome have
been defined. Proximal to the PAR region on the short arm of the Y chromosome are the sex-determining region and the sex-determining gene SRY. In general, males with deletions of the short arm of the Y chromosome do not differentiate as males and may manifest gonadal dysgenesis stigmata, especially lymphedema. They are not short. The findings in these patients support earlier evidence that genes responsible for the stigmata of gonadal dysgenesis are primarily on the short arms of the X and Y chromosomes, whereas a gene for stature (PHOG/SHOX on PAR1) and certain skeletal abnormalities resides on the proximal long arm of the Y chromosome. Deletions of the long arm of the Y chromosome and, to a lesser extent, the long arm of the X chromosome do not result in gonadal dysgenesis somatic stigmata.

The genes mapped to the long arm of the Y chromosome include a locus for H-Y antigen (SMCY) as assessed by the cytotoxic T cell assay, the pseudogene for STS, the YRBMI family of genes, and a putative locus that affects spermatogenesis.

A gene (GBY) that has a role in gonadoblastoma formation in patients with dysgenetic gonads is co-localized to a 1- to 2-Mb region (regions 3 and 4) near the centromere of the Y chromosome. A small number of patients with putative isochromosomes for the short and long arms of the Y chromosome are not mosaics. Three fourths of the patients reported with short arm isochromosomes, Ypi, were phenotypic males; the others had ambiguous genitalia. Seven patients with long arm isochromosomes, Yqi, were phenotypic females with gonadal dysgenesis. Gonadal dysgenesis stigmata and short stature were present in half the patients. In other words, the phenotype in 46,XYq11 individuals was that expected with the loss of Yp and thus SRY and the genes that prevent gonad dysgenesis stigmata, whereas patients with Ypi invariably had male differentiation. Patients with Yp Y chromosomes usually have an associated 45,XX cell line and hence are mosaics. The phenotypes vary from that of a normal male (depending on the presence of the SRY gene) through individuals with ambiguous genitalia and male pseudohermaphroditism to women with infanticile female external genitalia and bilateral streak gonads. The variation in phenotype is best explained by the effect of the 45,X cell line as well as the loss of genes on the Y chromosome.

Instances of apparently balanced Y-autosome translocations are known; usually the distal heterochromatic region of the long arm of the Y chromosome is translocated to either a D or G autosome. Other reciprocal balanced and unbalanced Y-autosome translocations have been reported in which male sex differentiation is normal.

As noted previously, Y-to-X translocations involving the sex-determining portion of the Y chromosome and a variable portion of the short arm of the X chromosome (usually the PAR) are found in more than 85% of XX males. There have also been reports of males with a 46,Y(Xp22:Yq11) karyotype who inherited the X(Y) from their mothers. These males have an intact Y chromosome and a variable deletion of the short arm of the X chromosome. The phenotypic features are variable and correlate with the extent of the deletion of the short arm of the X chromosome. Hence, short stature, mild mental retardation, chondrodysplasia punctata, STS deficiency with ichthyosis, anemia, and hypogonadotropic hypogonadism (Kallmann's syndrome) are variable features of this contiguous gene deletion syndrome. Phenotypically normal females with a 46,X(X;Y) karyotype have been reported. All females with Xp22:Yq11 translocations have been normal except for short stature and increased fetal wastage. In patients with Y-to-X translocation, careful cytogenetic and molecular DNA studies should be performed to define the exact breakpoints on the X and Y chromosomes. One can assess the functional integrity of the gonads by measuring the concentration of basal and LH/HRH-evoked plasma gonadotropins and plasma steroids. Pelvic MRI is helpful in defining the pelvic contents and gonads.

Y-autosome translocations have been reported in 45,X males. In general there appear to be two preferential locations, on chromosomes 5p and 18p, resulting in monosomy for 5p and 18p and multiple congenital anomalies. Three unrelated phenotypic males with an XQY karyotype had severe mental retardation, generalized hypotonia, and microcephaly with evidence of partial X chromosome disomy. Aberrant XY exchange involving the distal long arm of the X and the distal euchromatic region of the long arm of the Y chromosome results in functional disomy for X-linked genes, and the phenotype has been designated the XYq syndrome.

Pure Gonadal Dysgenesis

This term has been applied to phenotypic females with a 46,XX or 46,XY karyotype who have rudimentary streak gonads and remain sexually infantile but are of normal or tall stature and lack the somatic stigmata of gonadal dysgenesis. At puberty, they exhibit the usual effects of prepubertal castration, and plasma and urinary gonadotropin values are increased. The X chromatin pattern may be either positive or negative. Some X chromatin-negative patients have clitoral enlargement, which may be present at birth or first become manifest at puberty; clitoral enlargement is rarely present in X chromatin-positive patients. The designation pure gonadal dysgenesis was introduced by Hamrden and Stewart in 1959 in their report of a 19-year-old phenotypic female with the described phenotype and 46,XY karyotype. It is now known that a variety of etiologic factors may lead to the development of this clinical picture. We have chosen to restrict the term pure gonadal dysgenesis to patients with XX or XY gonadal dysgenesis (see later discussion). The latter is also referred to as complete XY sex reversal.

Familial and Sporadic Complete 46,XX Gonadal Dysgenesis and Its Incomplete Forms

46,XX gonadal dysgenesis is characterized by normal stature, sexual infantilism, bilateral streak gonads (similar in structure to those of 45, gonadal dysgenesis), normal female internal and external genitalia, primary or secondary amenorrhea, elevated gonadotropin levels, absence of the somatic stigmata of the syndrome of gonadal dysgenesis, and a 46,XX karyotype (Table 22-14).

The habitus is often eunuchoid. Rare patients have had a few somatic abnormalities, including cubitus valgus, but none has had the typical gonadal dysgenesis phenotype. McDonough and associates reviewed the phenotypic and cytogenetic findings in 82 phenotypic female patients with primary gonadal failure. Sex chromosome anomalies were found in association with ovarian failure in 52 of 82 patients, all of whom were less than 160 cm tall. Conversely, all patients taller than 160 cm with ovarian failure had either a 46,XX or a 46,XY karyotype.

Occasionally, women with clitoral enlargement, hirsutism, and other signs of virilization have serum testosterone levels above the range for normal women. The streak gonads secrete testosterone, presumably from nests of hilus cells. The high concentration of gonadotropins apparently leads to hilus cell hyperplasia and a modest increase in circulating androgen levels, which, in the presence of meager estrogen production, have potent biologic action.

Families may have multiple siblings affected, and within families the expression of the disease may vary in affected siblings. The gonads may range from bilateral streak gonads to hypoplastic ovaries with varying degrees of ovarian function resulting in secondary rather than primary amenorrhea. In the

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Complete Syndrome</th>
<th>Incomplete Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Karyotype:</td>
<td>46,XX</td>
<td>Same</td>
</tr>
<tr>
<td>Inheritance:</td>
<td>Autosomal recessive in familial cases (sensorineural deafness in about 10%); FSH receptor mutation in some cases</td>
<td>Same</td>
</tr>
<tr>
<td>Genitalia:</td>
<td>Normal female</td>
<td>Same</td>
</tr>
<tr>
<td>Wolffian duct derivatives:</td>
<td>Absent</td>
<td>Same</td>
</tr>
<tr>
<td>Müllerian duct derivatives:</td>
<td>Normal female</td>
<td>Same</td>
</tr>
<tr>
<td>Gonads:</td>
<td>Bilateral streak gonads</td>
<td>Hypoplastic ovary and streak or bilateral hypogonadotropic ovaries</td>
</tr>
<tr>
<td>Habitus:</td>
<td>Normal stature, no somatic stigmata of gonadal dysgenesis</td>
<td>Sexual infantilism</td>
</tr>
<tr>
<td>Hormone profile</td>
<td>Increased plasma FSH and LH concentration</td>
<td>Plasma estradiol variable: decreased or normal</td>
</tr>
</tbody>
</table>

FSH, follicle-stimulating hormone; LH, luteinizing hormone.
Familial cases transmission is consistent with an autosomal recessive trait. A locus for XX gonadal dysgenesis was mapped to the short arm of chromosome 2 in a large group of Finnish women. A homozygous missense mutation in the gene encoding the FSH receptor was detected; it resulted in a substitution of valine at residue 189 of the FSH receptor protein, and receptor activity was decreased but not absent in transfected cells. The affected women with XX gonadal dysgenesis had either primary or secondary amenorrhea, variable development of secondary sexual characteristics, and elevated gonadotropins. However, women with the FSH receptor mutation had primary follicles present in their ovaries, whereas women with primary ovarian failure of other causes had no or few ovarian follicles. Males homozygous for the same mutation of the FSH receptor had variable degrees of spermatogenetic failure but not azoosperma. Abnormal folliculogenesis in patients without the FSH receptor mutation may result from the effect of a mutant gene on germ cell migration, the gonadal blastema, or the rate of germ cell attrition, a defect in the putative ovary-organizing factor or its receptor, or accelerated atresia.

Familial 46,XX gonadal dysgenesis has been associated with sensorineural deafness. Genetic heterogeneity is suggested by concordance of the gonadal defect with deaf mutism in these families and by other families in which short stature, 46,XX gonadal dysgenesis, microcephaly, and arachnodyactyly occurred in affected siblings. Hamet and colleagues described three sisters with renal failure, adrenal hyperplasia, hypertension, sensorineural deafness, and primary hypogonadism. A kindred with cerebellar ataxia and hypergonadotropin hypogonadism and a family with mental retardation, streak gonads, myopathy, and various neurologic abnormalities have been described. The association of XX gonadal dysgenesis and eyelid dysplasia, ptosis, and epicanthus can occur in families. Haploinsufficiency of the FOXL2 gene on chromosome 3q23 (a member of the winged helix/forkhead family of transcription factor) causes autosomal dominant blepharophimosis/ptosis/epicanthus inversus syndrome (BPES type 1), which mainly involves a null mutation. In the mouse, Foxl2 is expressed in ovarian follicle cells and the developing eyelid.

Sporadic cases of 46,XX gonadal dysgenesis are heterogeneous. For example, ovarian hypoplasia has been associated with aneuplody, especially trisomy 13 and trisomy 18. Patients with 46,XX gonadal dysgenesis should be distinguished from those with ovarian failure caused by infection (e.g., mumps in childhood, autoimmune oophoritis) and from patients with antibodies to gonadotropin receptors, biologically inactive FSH, galactosemia, or biosynthetic errors that affect estrogen formation (e.g., deficiency of steroid 17-hydroxylase and/or 17,20-lyase [CYP17]).

Rare 46,XX patients with absent gonads, hypoplastic müllerian duct derivatives, and other congenital anomalies have been described in sibships with similarly affected 46,XY phenotypic females, suggesting an autosomal recessive mode of inheritance. The term 46,XX gonadal agenesis or agenadism has been applied to these patients because of the complete absence of gonads, even gonadal streaks, and the occurrence of müllerian duct abnormalities. In contrast to 46,XY gonadal dysgenesis, gonadal neoplasms are rare in 46,XX gonadal dysgenesis. The diagnosis of 46,XX gonadal dysgenesis is based on finding a normal karyotype in a sexually infantile phenotypic female with hypergonadotropin hypogonadism. In sporadic cases, it is important to confirm the presence of streak or hypogynadistic gonads by ultrasonography, pelvic MRI, or laparoscopy. Replacement therapy with estrogen is similar to that for patients with 45,X gonadal dysgenesis.

46,XY gonadal dysgenesis was first described by Swyer. This syndrome in its complete form is characterized by a female phenotype, normal to tall stature, bilateral diencephalic gonads, sexual infantilism with primary amenorrhea, eunuchoid habitus, and a 46,XY karyotype. Common features of gonadal dysgenesis are usually absent. The internal structures are female with bilateral fallopian tubes, a uterus, and a vagina. Clitoral enlargement is not uncommon, and gonadal neoplasms, especially gonadoblastoma and germinoma (seminoma, dysgerminoma), occur in 10% to 30%. In individuals with the incomplete or variant form, both the internal and external genitalia may be ambiguous. Breast development after the normal age of puberty suggests the presence of an estrogen-secreting gonadal tumor, especially a gonadoblastoma. Plasma and urinary gonadotropin levels are increased. In some patients the concentration of serum testosterone is higher than in adult women, presumably because of the secretion of androgens from the hilus cells of the streak gonads. A male proportion of single fluorescent Y bodies is present in interphase nuclei. Excluded from this syndrome are patients with variants of the syndrome of gonadal dysgenesis (e.g., 45,X/46,XY mosaicism) and those with microscopically visible structural abnormalities of the Y chromosome.

XY gonadal dysgenesis is a heterogeneous condition that can result from deletions of the short arm of the Y chromosome, mutations in the SRY gene, mutations in a variety of sex-determining genes, or duplications of the DAX1 gene on the X chromosome. As noted previously, patients with extensive contiguous deletions of the short arm of the Y chromosome can manifest gonadal dysgenesis stigmata as well as XY gonadal dysgenesis. Point mutations in the SRY gene have been found in 10% to 20% of patients with complete XY gonadal dysgenesis. Presumably,

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Complete Syndrome</th>
<th>Incomplete Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Karyotype:</td>
<td>46,XY</td>
<td>Same</td>
</tr>
<tr>
<td>Inheritance:</td>
<td>Familial cases consistent with X linked (or male-limited autosomal dominant); SRY mutation or deletion in 15%</td>
<td>Same: No SRY mutation detected</td>
</tr>
<tr>
<td>Genitalia:</td>
<td>Female</td>
<td>Ambiguous</td>
</tr>
<tr>
<td>Wolffian duct derivatives:</td>
<td>Absent</td>
<td>Rudimentary hypoplastic</td>
</tr>
<tr>
<td>Müllerian duct derivatives:</td>
<td>Normal</td>
<td>Variable, rudimentary hypoplastic</td>
</tr>
<tr>
<td>Gonads:</td>
<td>Bilateral streak gonads</td>
<td>Bilateral dysgenetic testes or streak gonad + dysgenetic testes (mixed gonadal dysgenesis)</td>
</tr>
<tr>
<td>Risk:</td>
<td>Increased risk of gonadal tumor, gonadoblastoma</td>
<td>Same</td>
</tr>
<tr>
<td>Habitus:</td>
<td>Sexual infantilism at puberty</td>
<td>Variable degree of virilization at puberty</td>
</tr>
<tr>
<td>Hormone profile:</td>
<td>Increased plasma FSH and LH and decreased testosterone concentrations postpubertally</td>
<td>Same</td>
</tr>
</tbody>
</table>

FSH, follicle-stimulating hormone; LH, luteinizing hormone.

most patients with XY gonadal dysgenesis who do not have a mutation in the SRY gene have mutations either in upstream or downstream genes in the testes determination and differentiation cascade. Almost all the mutations in the SRY gene in patients with the complete form of XY gonadal dysgenesis, except for example, one in the 3' region and a deletion in the 5' region of the gene, are in the DNA-binding (HMG) box domain of the SRY protein. These mutations are usually nonsense or missense single-amino-acid substitutions (15 of 23 mutations) and affect DNA binding or binding. Familial aggregates of 46,XY gonadal dysgenesis have been reported. The presence of complete 46,XY dysgenesis in "daughters" of fathers who have the same mutation in the SRY HMG box is perplexing. Possible explanations for this phenomenon include (1) another mutation in the father that allows for normal testicular determination and differentiation; (2) segregation of another polymorphic locus that interacts with SRY; (3) variations in the timing and the critical level of SRY expression achieved; and (4) mosaicism. Two novel SRY missense mutations causing reduced DNA binding have been identified in XY females and their fathers. Analyses of DNA from the fathers indicated that they were mosaic for both the wild-type and mutant SRY genes and quite likely are mosaic for the SRY mutation in their testes.

Familial 46,XY gonadal dysgenesis without an apparent mutation in the SRY may be transmitted as an X-linked, sex-limited
autosomal dominant or possibly as an autosomal recessive trait. Within a family, affected individuals may vary in the appearance of the external genitalia and the development of secondary sexual characteristics. Usually, the external genitalia and internal genital ducts are female, and the patient is sexually infantile (complete form). However, affected siblings may have ambiguous external genitalia, ambiguous genital ducts, and a urogenital sinus (incomplete or variant form). The spectrum of findings suggests heterogeneity in transmission or variable expression of the mutant gene in the same cohort. In a family reported by Chemke and colleagues, two siblings had 46,XY gonadal dysgenesis with bilateral streak gonads and another had the incomplete form with genital ambiguity, bilateral dysgenetic testes, and müllerian derivatives. We studied an infant born to the "normal" 46,XX sister in this family who had a 46,XY karyotype, ambiguous external genitalia, bilateral dysgenetic testes, and müllerian duct derivatives; inheritance in this family is consistent with an X-linked recessive or male-limited autosomal dominant trait.

46,XY gonadal dysgenesis can be caused by a duplication of the DAX1 locus on the short arm of the X chromosome (Xp21). This gene, which encodes congenital adrenal hypoplasia, is postulated to act as a suppressor of SRY. Duplication of DAX1 in 46,XY individual leads to sex reversal; deletion of the DAX1 gene has no effect on testicular development. A further example of gene duplication causing XY sex reversal has been reported for the WNT4 gene. The mechanism postulated for the sex reversal is WNT4-induced up-regulation of DAX1.

46,XY gonadal dysgenesis occurs in association with campomelic dysplasia, 844 which is caused by a mutation in the SOX9 gene on 17q21. Mutations in a single allele for this gene in males can result in both XY gonadal dysgenesis and campomelic dysplasia (see Fig. 22-17). 845,846 46,XX affected females have normal ovarian development. Homozygous mutations are probably lethal, owing to the critical function of SOX9 in chondrogenesis. No SOX9 mutations have been reported in patients with XY gonadal dysgenesis alone.

9p- and 10q- deletions also can cause 46,XY gonadal dysgenesis. 847,848 The DMRT1 at 9p24.3 is a candidate sex-determining gene that would be deleted in the 9p- patients; the gene(s) on 10q- affecting testsis determination is not known. 849,850 Mutations of WNT1 result in the Denys-Drash, Frasier, and WAGR syndromes (see later discussions). Various names have been applied to this syndrome. 840,841,842 46,XY gonadal dysgenesis has also been associated with multiple congenital anomalies (genito-palato-cardiac syndrome) as in the Smith-Lemli-Opitz syndrome (see later).

Like some 45,X individuals, some patients with 46,XY gonadal dysgenesis have ovarian follicles in the dysgenetic gonads. Cussen and MacMahon described germ cells and follicles in the underdeveloped gonads of a 3-month-old 46,XY female. Subsequent examination at age 3 years 10 months revealed bilateral streak gonads and a gonadoblastoma in one gonad. Some ovarian follicles persisted and functioned at puberty in 46,XY phenotypic females who underwent spontaneous puberty and experienced menarche. Examination of the gonads after development of secondary amenorrhea revealed only gonadal streaks with a few hilus cells. We recommend that these patients not be classified as true hermaphrodites.

There is a high prevalence of gonadal tumors, especially gonadoblastoma and germinoma, in 46,XY gonadal dysgenesis, and they can occur bilaterally in childhood. Bilateral prophylactic gonadectomy is indicated in all patients with 46,XY gonadal dysgenesis, even those who have deletions or mutations involving the SRY gene. The sex rearing of patients with the incomplete form of 46,XY gonadal dysgenesis is determined by the extent of genital ambiguity and the age at diagnosis. Patients raised as females should be placed on estrogen replacement therapy at age 12 to 13 years and should eventually be cycled monthly with estrogen and progestagen. In patients raised as males, testosterone replacement therapy should begin at the age of puberty.

"Male Turner's Syndrome"

Many phenotypic males have been reported with short stature, webbed neck, and other somatic abnormalities associated with the syndrome of gonadal dysgenesis in whom the testes were hypoplastic and frequently undescended. The resemblance of these males to females with 45,X gonadal dysgenesis suggested a pathogenetic parallelism or gonadal dysgenesis in the male. However, with rare exceptions, this relation is no longer tenable. A few patients with the phenotypic features of gonadal dysgenesis syndrome have a sex chromosome abnormality (e.g., 45,X/46,XY mosaicism) and represent a variant form of gonadal dysgenesis. The other patients have "Male Turner's Syndrome"

Among the group of phenotypic males previously classified as having male gonadal dysgenesis syndrome, a distinctive clinical entity was identified that led to the identification of its counterpart in the female and its designation as separate from the syndrome of gonadal dysgenesis. Various names have been applied to this new well-established syndrome (Noonan's syndrome, 46,XX and 46,XY Turner's phenotype, pseudo-Turner's syndrome, Ultich's syndrome), but we prefer to exclude Turner" from the designation. Table 22-16 lists the clinical features of 2 phenotypic males and 12 phenotypic females with this entity studied by us in 1962. These patients have a characteristic facies that includes hypertelorism, down-slanting palpebral tissues, epicanthal folds, ptosis, low-set anteriorly rotated ears, and, frequently, a webbed neck and short stature (Fig. 22-51); in 12 of the 14 cases, congenital heart disease was present. The most common cardiac malformation is pulmonary valvular stenosis (50% to 60%) often with dysplasia. Hypertrophic cardiomyopathy occurs in 20% to 30% of patients and can manifest in the neonatal period; the electrocardiogram is frequently abnormal. Atrial septal defects are found in 10% of patients. The echocardiogram is abnormal in 50% to 80% of individuals with Noonan's syndrome. Coarctation of the aorta and aortic stenosis, the most common cardiovascular anomalies in the syndrome of gonadal dysgenesis, are infrequent (10%). Pectus carinatum superiorly, pectus excavatum inferiorly, and cubitus valgus are often present. Feeding problems in infancy are common, including gastric dysmotility. Mental development is impaired in about 15% of patients. There are few reports of cognition and intelligence in individuals with Noonan's syndrome. Several reports have described the Noonan phenotype predicted a specific pattern of defects in cognitive functioning. Refractive errors and strabismus are common. Lymphedema occurs in 15% of patients. Sixty-five percent of patients have easy bruising or excessive postoperative bleeding as a result of a variety of coagulation defects; the most common is partial
<table>
<thead>
<tr>
<th>Trait</th>
<th>Affected</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Broad apex nasi</td>
<td>2/2</td>
<td>11/12</td>
</tr>
<tr>
<td>Low-set and/or malformed ears</td>
<td>2/2</td>
<td>8/12</td>
</tr>
<tr>
<td>High-arched palate</td>
<td>2/2</td>
<td>8/12</td>
</tr>
<tr>
<td>Neck</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short</td>
<td>2/2</td>
<td>10/12</td>
</tr>
<tr>
<td>Webbing</td>
<td>2/2</td>
<td>10/12</td>
</tr>
<tr>
<td>Low hairline</td>
<td>2/2</td>
<td>10/12</td>
</tr>
<tr>
<td>Chest</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shield-like</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wide-spaced nipples</td>
<td>2/2</td>
<td>11/11</td>
</tr>
<tr>
<td>Pectus excavatum</td>
<td>2/2</td>
<td>5/12</td>
</tr>
<tr>
<td>Cardiac abnormalities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pulmonic stenosis (PS)</td>
<td>2/2</td>
<td>11/12</td>
</tr>
<tr>
<td>PS and ventricular septal defect</td>
<td>0/2</td>
<td>1/10</td>
</tr>
<tr>
<td>Atrial septal defect (ASD)</td>
<td>2/2</td>
<td>6/10</td>
</tr>
<tr>
<td>ASD with anomalous pulmonary venous return</td>
<td>0/2</td>
<td>1/10</td>
</tr>
<tr>
<td>Endocardial cushion defect (ECD)</td>
<td>0/2</td>
<td>2/10</td>
</tr>
<tr>
<td>ECD + patent ductus arteriosus and mitral insufficiency</td>
<td>0/2</td>
<td>1/10</td>
</tr>
<tr>
<td>Both PS and ASD</td>
<td>2/2</td>
<td>3/10</td>
</tr>
<tr>
<td>Patent ductus arteriosus (PDA)</td>
<td>0/2</td>
<td>2/10</td>
</tr>
<tr>
<td>Undiagnosed heart disease</td>
<td>0/2</td>
<td>2/10</td>
</tr>
<tr>
<td>Incompletely evaluated</td>
<td>0/2</td>
<td>2/12</td>
</tr>
<tr>
<td>Extremities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cubitus valgus</td>
<td>2/2</td>
<td>9/12</td>
</tr>
<tr>
<td>Gracile fingers</td>
<td>1/2</td>
<td>8/12</td>
</tr>
<tr>
<td>Short stubby fingers</td>
<td>1/2</td>
<td>2/12</td>
</tr>
<tr>
<td>Lymphedema</td>
<td>0/2</td>
<td>3/12</td>
</tr>
<tr>
<td>Dystrophic nails</td>
<td>2/2</td>
<td>2/12</td>
</tr>
<tr>
<td>Shortened fourth metacarpal(s)</td>
<td>0/2</td>
<td>3/12</td>
</tr>
<tr>
<td>Clinodactyly of fifth finger(s)</td>
<td>1/2</td>
<td>2/12</td>
</tr>
<tr>
<td>Palmar simian crease</td>
<td>1/2</td>
<td>1/12</td>
</tr>
<tr>
<td>Undescended testes</td>
<td>2/2</td>
<td></td>
</tr>
<tr>
<td>Delayed puberty</td>
<td>1/1</td>
<td>3/3</td>
</tr>
<tr>
<td>Skeletal retardation</td>
<td>2/2</td>
<td>8/10</td>
</tr>
<tr>
<td>Mental development</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retarded</td>
<td>2/2</td>
<td>4/12</td>
</tr>
<tr>
<td>Borderline</td>
<td>0/2</td>
<td>5/12</td>
</tr>
<tr>
<td>Normal</td>
<td>0/2</td>
<td>3/12</td>
</tr>
<tr>
<td>Intrauterine growth retardation</td>
<td>1/2</td>
<td>4/12</td>
</tr>
<tr>
<td>Renal collecting system</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>2/2</td>
<td>7/8</td>
</tr>
<tr>
<td>Abnormal</td>
<td>0/2</td>
<td>1/8</td>
</tr>
<tr>
<td>Normal karyotype</td>
<td>2/2</td>
<td>12/12</td>
</tr>
</tbody>
</table>

factor XI deficiency. The chromosome constitution is normal, and gonadal differentiation is appropriate for the phenotypic and chromosomal sex. Cryptorchidism is common in males, and the testes may be hypoplastic and exhibit germinal aplasia. Puberty is delayed, and androgen deficiency is not uncommon. However, 50% of affected males have normal testicular function, including fertility in the absence of cryptorchidism. Affected females have functioning ovaries and, although the onset of puberty may be delayed, female secondary sexual characteristics eventually emerge.

Noonan’s syndrome has an incidence of 1 in 1000 to 1 in 2500 live births and about 50% of cases are sporadic. Familial clusters are usually consistent with autosomal dominant inheritance in this genetically heterogeneous disorder. Linkage analysis in a large Dutch kindred suggested location of a gene for Noonan’s syndrome on the long arm of chromosome 12 (12q24). Missense mutations in PTPN11, the gene encoding the nonreceptor protein tyrosine-2 phosphatase (SHP-2) that contains two Src homology-2 (SH2) domains, is a cause of Noonan’s syndrome, accounting for about 50% of affected individuals. A gain of function arising from excess SHP-2 activity is postulated as the disease mechanism. The abnormality of gonadal function and the higher incidence of congenital heart disease in males may play a part in the apparently higher maternal transmission of the mutant gene.

The diagnosis is based on the constellation of stigmata, the most prominent of which are short stature, webbed neck, pectus excavatum, ptosis, and right-sided congenital heart disease in a patient with a normal sex chromosome constitution. The differential diagnosis of this syndrome is extensive and includes structural abnormalities of the Y chromosome (especially those involving the short arm), 45,X/46,XY mosaicism, and dysmorphic syndromes secondary to hydantoin, primidone, or alcohol exposure during gestation. At puberty, affected males may require testosterone replacement therapy. Mean final height tends to follow the lower limits of the growth curve; mean adult height was 162.5 cm in men and 152.7 cm in women according to Ranke and co-workers.

True Hermaphroditism

The diagnosis of true hermaphroditism requires the presence of both ovarian (containing follicles) and testicular tissue in either the same or opposite gonads. Failure to adhere to this definition has led to considerable confusion. Gonadal stroma arranged in whorls, similar to those found in the ovary but lacking oocytes, should not be considered sufficient evidence to designate the rudimentary gonad as an ovary. Similarly, if testicular tissue is present in the contralateral gonad, we do not consider the presence of a few oocytes in a streak gonad to be adequate evidence for the diagnosis of true hermaphroditism. Because rare female-type germ cells may be found in patients with 45,X gonadal dysgenesis, it is of little value from the clinical, cytogenetic, embryologic, or nosologic standpoint to classify as true...
hermaphrodites the 45,X/46,XY mosaics or individuals with 46,XY complete gonadal dysgenesis in whom a dysgenetic gonad contains rare oocytes. Similarly, the status of the internal and external genitalia, which invariably exhibit some degree of ambisexual development, should not be used as a criterion for classification of an individual as a true hermaphrodite. This uncommon type of intersex is relatively more prevalent in South African blacks.

Classification

True hermaphroditism is uncommon but has been reported in more than 400 individuals. 

Patients with this syndrome may be subclassified according to the type and location of the gonads.

Lateral.

A testis is present on one side, and an ovary is present on the other in about 20% of patients. The ovary is frequently found on the left side. 

Bilateral.

Both testicular and ovarian tissue are present bilaterally, usually as ovotestes, in about 30% of patients. 

Unilateral.

Testicular and ovarian tissue is present on one side and a testis or ovary is found on the other side in slightly less than one half of cases. 

A testis or ovotestis may be

<table>
<thead>
<tr>
<th>TABLE 22-17 – Clinical Features of True Hermaphroditism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Karyotype: 46,XX (most common), 46,XX/46,XY, or 46,XY (rare)</td>
</tr>
<tr>
<td>Inheritance: Familial cases (autosomal recessive, autosomal dominant transmission) rare</td>
</tr>
<tr>
<td>Genitalia: Ambiguous; cryptorchidism frequent; ovotestes possibly located in labioscrotal fold</td>
</tr>
<tr>
<td>Wolffian duct derivatives: Duct differentiation after that of the homolateral gonad</td>
</tr>
<tr>
<td>Müllerian duct derivatives:</td>
</tr>
<tr>
<td>Gonad: Testis, ovary, or ovotestis</td>
</tr>
<tr>
<td>Habitus: Breast development and virilization common at puberty</td>
</tr>
<tr>
<td>Molecular studies: Approximately 10% of 46,XX true hermaphrodites are SRY positive</td>
</tr>
</tbody>
</table>

and incomplete fusion of the labioscrotal folds. The labioscrotal folds are asymmetric in half of the patients, with the right side more predominant. In rare cases a penile urethra is present. Cryptorchidism is common, but at least one gonad is usually palpable in the labioscrotal fold or in the inguinal region. An inguinal hernia, which may contain a gonad or uterus, is present in about one half of the cases, and a vagina and a uterus are present in most patients; the uterus may be underdeveloped (hemivagina), rudimentary, or absent. The differentiation of the genital ducts usually follows that of the gonads. The ovotestis is the most common gonad in true hermaphrodites, followed by the ovary and, least commonly, the testis. In patients with a testis on one side and an ovary on the other, the development of the homolateral duct is usually consistent with that of the gonad, despite the varied appearance of the external genitalia. Most patients with an ovotestis have predominantly female development of the genital ducts. The relation between gonadal structure and differentiation of the genital tract in true hermaphroditism provides additional evidence for the local action of AMH secreted by the Sertoli cells of the embryonic testes.

Breast development is common during puberty in true hermaphrodites, and menses occur in more than half of the patients. Periodic hematuria associated with menstruation is a late clue to the diagnosis. Spermagogenesis is rare; seminiferous tubules in an ovotestis or testes are abnormal in most cases, and interstitial fibrosis of the testes is common. Ovulation is not uncommon, and pregnancy and childbirth can occur in patients with a 46,XX karyotype, whereas only one 46,XY true hermaphrodite has been reported to have fathered a child.

Few studies of hypothalamic-pituitary-gonadal function have been carried out in true hermaphrodites. Whereas an ovary or
True hermaphroditism could result from sex chromosome mosaicism (apparent or cryptic), chimerism, Y-to-autosome or Y-to-X chromosome translocation, or mutation of either X-linked or autosomal genes involved in sex determination. Most 46,XX true hermaphrodites (85%) have SRY-negative in leukocyte DNA. However, by the use of molecular genetic and histochemical techniques for the SRY gene and the SRY protein, SRY gene expression and protein has been detected in the ovotestes of XX true hermaphrodites in whom leukocytes and ribosomal DNA were negative for SRY. In six XX hermaphrodites in whom the peripheral karyotype was SRY negative, all ovotestes exhibited low levels of expression of the SRY gene and SRY protein was detected mainly in Sertoli and germ cells. These observations suggest cryptic mosaicism or chimerism of a Y-bearing cell may be more common in XX 'true' hermaphrodites than previously thought.

The (genes or) responsible for XX (SRY) true hermaphroditism has not yet been identified; however, it is probably a downstream gene in the sex determination pathway that, when mutated, allows for testicular differentiation. That Y sequence generates 46,XX males and 46,XX true hermaphrodites can occur in the same pedigree is well documented. For example, studied a pedigree with two 46,XX true hermaphrodites and one 46,XX male who were first cousins. The three patients were negative for Y sequences, including PABY (pseudoautosomal boundary Y), SRY, and ZFY. The frequency of occurrence of 46,XX true hermaphrodites and 46,XX maleness was compatible with an autosomal dominant or an X-linked dominant mode of transmission with variable expression in the case of a putative X-linked gene mutation.

It has been postulated that random inactivation would result in true hermaphroditism and nonrandom inactivation would lead to an XX male. However, no evidence for a functional inactivation in the X chromosome has been reported. Reinforcement of the external genitalia has led to a general acceptance of the X chromosome as a testis determiner. In 1987, Spurle and colleagues carried out a detailed DNA analysis of the X chromosome in 66,XX SRY-negative true hermaphrodites and found no evidence of uniparental disomy of the X chromosome. Sarafoglou and Oster reviewed the familial cases of true hermaphroditism and pedigrees in which both true hermaphroditism and XX maleness have occurred. An SRY-X translocation was detected in one pedigree, but in the other affected families an X-linked gene, a sex-limited autosomal dominant gene, or an autosomal recessive gene that could have made the sex determination.

A small number of "Y-positive" 46,XX true hermaphrodites have been reported in kindreds with 46,XX males. It has been postulated that inactivation of the X chromosome would lead to a true hermaphroditism. The SRY gene by the spread of inactivation into the translocated segment. However, in one SRY-positive 46,XX true hermaphrodite, the translocated X was randomly inactivated; on the other hand, nonrandom inactivation of the X chromosome bearing the Y translocation was found in another SRY-positive 46,XX male. Accordingly, in some instances X inactivation may play a role in the gonadal phenotype of 46,XX SRY-positive males.

Sex chromosome mosaicism arises from mitotic or meiotic errors. In contrast, 46,XX/46,XY chimerism is usually a consequence of double fertilization or, possibly, fusion of two normally fertilized ova. Chimeric individuals have two distinct populations of cells, each of which has a different genetic origin (in contrast to mosaicism). Study of 46,XX/46,XY chimeras provides evidence for the fertilization of a binucleate ovum by two sperm, one bearing an X and the other a Y. Not all patients with whole-body chimerism have true hermaphroditism. One 46,XX/46,XY patient was a phenotypic male without true hermaphroditism; a likely mechanism for the chimerism in this case, based on the blood group studies and other findings, is fusion of two zygotes or fertilization of an ovum and its polar body. The experiments of Tarkowski with XX and XY mouse blastocysts demonstrated that random fusion of two blastocyst s seldom produces XXXY true hermaphrodites; fused mouse blastocysts usually develop testes rather than ovaries or ovotestes. A 46,XY true hermaphrodite, the least pattern of inactivation of the XX chromosome, has been reported. The normal expression of chimerism). Study of 46,XX/46,XY chimeras provides evidence for the fertilization of a binucleate ovum by two sperm, one bearing an X and the other a Y. Not all patients with whole-body chimerism have true hermaphroditism. One 46,XX/46,XY patient was a phenotypic male without true hermaphroditism; a likely mechanism for the chimerism in this case, based on the blood group studies and other findings, is fusion of two zygotes or fertilization of an ovum and its polar body. The experiments of Tarkowski with XX and XY mouse blastocysts demonstrated that random fusion of two blastocyst s seldom produces XXXY true hermaphrodites; fused mouse blastocysts usually develop testes rather than ovaries or ovotestes. A 46,XY true hermaphrodite, the least pattern of inactivation of the XX chromosome, has been reported. The normal expression of the X chromosome is the result of blastocyst s seldom produces XXXY true hermaphrodites; fused mouse blastocysts usually develop testes rather than ovaries or ovotestes. A 46,XY true hermaphrodite, the least pattern of inactivation of the XX chromosome, has been reported. The normal expression of chimerism). Study of 46,XX/46,XY chimeras provides evidence for the fertilization of a binucleate ovum by two sperm, one bearing an X and the other a Y. Not all patients with whole-body chimerism have true hermaphroditism. One 46,XX/46,XY patient was a phenotypic male without true hermaphroditism; a likely mechanism for the chimerism in this case, based on the blood group studies and other findings, is fusion of two zygotes or fertilization of an ovum and its polar body. The experiments of Tarkowski with XX and XY mouse blastocysts demonstrated that random fusion of two blastocyst s seldom produces XXXY true hermaphrodites; fused mouse blastocysts usually develop testes rather than ovaries or ovotestes. A 46,XY true hermaphrodite, the least pattern of inactivation of the XX chromosome, has been reported. The normal expression of chimerism). Study of 46,XX/46,XY chimeras provides evidence for the fertilization of a binucleate ovum by two sperm, one bearing an X and the other a Y. Not all patients with whole-body chimerism have true hermaphroditism. One 46,XX/46,XY patient was a phenotypic male without true hermaphroditism; a likely mechanism for the chimerism in this case, based on the blood group studies and other findings, is fusion of two zygotes or fertilization of an ovum and its polar body. The experiments of Tarkowski with XX and XY mouse blastocysts demonstrated that random fusion of two blastocyst s seldom produces XXXY true hermaphrodites; fused mouse blastocysts usually develop testes rather than ovaries or ovotestes. A 46,XY true hermaphrodite, the least pattern of inactivation of the XX chromosome, has been reported. The normal expression of chimerism). Study of 46,XX/46,XY chimeras provides evidence for the fertilization of a binucleate ovum by two sperm, one bearing an X and the other a Y. Not all patients with whole-body chimerism have true hermaphroditism. One 46,XX/46,XY patient was a phenotypic male without true hermaphroditism; a likely mechanism for the chimerism in this case, based on the blood group studies and other findings, is fusion of two zygotes or fertilization of an ovum and its polar body. The experiments of Tarkowski with XX and XY mouse blastocysts demonstrated that random fusion of two blastocyst s seldom produces XXXY true hermaphrodites; fused mouse blastocysts usually develop testes rather than ovaries or ovotestes. A 46,XY true hermaphrodite, the least pattern of inactivation of the XX chromosome, has been reported. The normal expression of...
The prevalence of gonadal neoplasms is increased in patients with certain types of dysgenetic gonads, in particular all those with a Y-bearing cell line. Germinoma (dysgerminoma, seminoma), teratoma, and gonadoblastoma have been found. Cryptorchid testes, even when not associated with intersexuality, are also associated with an increased risk of malignancy. The probability that cryptorchid testes will undergo malignant degeneration is difficult to assess, but it is at least 10 times greater than the probability for normally descended testes. Approximately 7% of males with testicular neoplasms have been or are cryptorchid at the time of diagnosis. In addition, in one third of patients with cryptorchidism who develop carcinoma of the testis, the neoplasm occurs after orchiectomy; in patients with unilateral cryptorchidism, 25% of tumors were located in the contralateral descended testes. C I S is a premalignant lesion of the testes. It is characterized by germ cells that are larger than normal spermatogonia, have clumped chromatin, are highly aneuploid, and are positive for placent al alkaline phosphatase. C I S, also called intratubular germ cell neoplasm (IGCN), is commonly seen adjacent to germ cell tumors in adults. However, it is not found adjacent to tumors before puberty, and the cells differ morphologically from those in postpubertal males. The natural history of CIS in prepubertal patients remains to be determined. About 50% of postpubertal patients with CIS develop germ cell tumors within 5 years of diagnosis; it has been suggested that, with time, the incidence may approach 100%. C I S is thought to originate from primordial germ cells. It occurs in 2% to 4% of males with cryptorchidism, in 25% of XY individuals with ambiguous genitalia, in individuals with dysgenetic testes and an XY cell line (XY gonadal dysgenesis, XO/XY mosaicism), and in the syndrome of androgen resistance. Localized low-dose radiotherapy to the testis will eradicate CIS and germ cells but preserves Leydig cells. The progression from CIS to invasive germ cell tumors appears to correlate with expression of cyclins and cyclin-dependent kinase inhibitors.

Gonadal neoplasms are uncommon in patients with 47,XXY seminiferous tubule dysgenesis, but a small number of patients have gonadal or extragonadal germ cell tumors. Similar gonadoblastoma is rare in the streak gonads of 45,X patients and in 45,X mosaics with a normal or structurally abnormal X chromosome in the second cell line. Gonadoblastoma and dysgerminoma, mucinous cystadenoma and a hilus cell tumor with signs of utilization have been reported in gonadal dysgenesis. However, these patients have not been studied for a low percentage of Y chromosome mosaicism by molecular analyses.

Gonadoblastomas are usually composed of three elements; large germ cells, sex cord derivatives (Sertoli-granulosa cells), and stromal cells (theca cells, Leydig cells). They are found almost exclusively in patients who have a 46,XY cell line. They may be microscopic or large, especially if overgrown by other germ cell elements, and often calcified. A comprehensive review of gonadoblastoma was published by Scully (see Fig. 22-50). In 27 of 74 patients, a tumor was present in both gonads. Thirty patients were younger than 15 years of age when the tumors were diagnosed, and 10 were age 15 years or older. A few of these tumors were detected incidentally on histologic examination of dysgenetic gonads removed for other indications. In patients in whom chromosomal studies were carried out, the predominant karyotypes were 45,X/46,XY and 46,XY. Although 80% of 46,XY patients are reared as females, most display some degree of clitoromegaly or hirsutism; rarely, the tumors secrete enough estrogen to induce breast development (see Fig. 22-50). Pure gonadoblastomas can be regarded as germ cell tumors in situ and as such do not metastasize. In half the cases, however, the germ cells infiltrate the stroma of the tumor to form a seminoma. Gonadoblastomas are also associated with more highly malignant germ cell tumors such as endodermal sinus tumors, embryonal carcinoma, and choriocarcinoma (10%). There is an increased risk of gonadal tumors (gonadoblastoma and/or seminoma) in patients with 46,XY gonadal dysgenesis and particularly in familial cases. The strikingly disparate propensity for neoplastic transformation in the streak or dysgenetic gonads of patients with 46,XY gonadal dysgenesis in contrast to 46,XX gonadal dysgenesis must be emphasized. The gonadoblastoma locus (GBY) has been mapped to a critical interval on the short arm and adjacent centromeric region of the Y chromosome. Five functional genes have been identified in this region as possible candidates for the cancer-predisposing locus. TSPY, one of the encoded proteins, is preferentially expressed in the germ cells of gonadoblastoma specimens.

In view of the well-documented malignant potential of dysgenetic gonads, the question of prophylactic gonadectomy merits serious attention. The neoplasms are infrequent in childhood, but the risk rises appreciably in young adults. High gonadotropin levels may play a role in their growth, and substitution therapy with gonadotropins may afford some protection. A prudent course is to advise laparotomy or laparoscopy and removal of the dysgenetic gonads of all patients with 46,XY gonadal dysgenesis (complete and incomplete forms) and of all patients with the syndrome of gonadal dysgenesis who have a cell line with a normal or a structurally abnormal Y chromosome or who have Y chromosome material as determined by molecular genetic studies. Exceptions to this rule occur in patients who are 45,X/46,XY mosaics with normal male genitalia, histologically normal testes, and normal gonadotropin levels and in patients with 45,X/46,XY mosaicism with ambiguus genitalia who have been assigned a male gender role and in whom a histologically normal gonad is located in the scrotum. However, the fact that a gonad is located in the scrotum or labial folds and is palpable does not guarantee against a disastrous result, because seminomas can metastasize at an early stage, before a local mass is obvious. Hence, it is prudent to subject these testes postpubertally to ascertain the presence of CIS (a premalignant lesion). Patients with 45,X gonadal dysgenesis who have no suggestion of clitoromegaly are not at risk. The incidence of gonadal tumors in patients with other X chromosome abnormalities, such as 45,X/46,XX, 45,XX, 46,X(, and 46,XXp, is low; however, these patients should be examined at regular intervals and, if indicated, monitored by ultrasonography of the pelvis for signs of gonad or uterine neoplasia.

Sex Chromosome Abnormalities Unassociated with Gonadal Defects

The addition of one more sex chromosomes to the genome has a deleterious effect on cognitive function. The following five sex chromosome abnormalities are not accompanied by a typical gonadal defect but are frequently associated with mental retardation.

47,XXX

This common chromosome abnormality has the frequency of about 1 per 1000 female newborns. The prevalence of 47,XXX individuals in institutions for the mentally retarded is 4.3 per 1000, suggesting an increased risk for severe mental retardation. 47,XXX females have reduced general intelligence and low scores on tests of attention, concept formation, spatial thinking, verbal fluency, and academic skills. XXX individuals demonstrated less overall psychosocial adaptation and a greater degree of psychological disturbance, compared with patients with Klenefelter's syndrome or gonadal dysgenesis. Although some have delayed menarche or premature ovarian failure, most 47,XXXX females have normal ovarian function. 47,XXX females can rarely give birth to 47,XXX sons and the prevalence of congenital malformations is increased in the progeny of 47,XXX women. Subtle clinical features in infants ascertainment by karyotype analysis include the following: a tendency to low birth weight, decreased head circumference, advanced mean parental age, an increased incidence of clinodactyly, normal postnatal growth patterns, an increased risk of speech and language problems, and a lower mean I.Q. than the siblings or a control group. The extra X chromosome is of maternal origin in most instances, arising mainly from nondisjunction during the first or second meiotic divisions. The diagnosis of 47,XXX can be confirmed by the finding of two sex chromatin bodies in interphase cells and by the demonstration of a 47,XXX karyotype through the use of appropriate banding techniques. Because of the increased risk in the offspring of a sex chromosome abnormality (47,XXX and 47,XXY) and congenital malformations, prenatal counseling and amniocentesis should be considered in 47,XXX females who become pregnant.

48,XXX

This rare anomaly is associated with considerable phenotype heterogeneity, making identification by clinical means difficult. The most constant feature is a variable degree of mental retardation affecting speech. Ovarian function is usually normal. Average adult height is 169 cm; the prevalence of skeletal anomalies, such as clinodactyly and radioulnar synostosis, is increased. The diagnosis is suggested by finding three sex chromatin bodies in 6% to 9% of somatic nuclei and is confirmed by karyotype analysis.
The rare penta-X syndrome is invariably associated with severe prenatal and postnatal growth delay and mental retardation. Other stigmata include microcephaly, hypertelorism, epicanthal folds, up-slanted palpebral fissures, depressed nasal bridge, abnormal dentition, a short neck, congenital heart disease, clinodactyly, overlapping toes, and joint laxity. The external genitalia and gonadal function are usually normal; however, fertility is still to be documented. A proportion of interphase nuclei contain four X chromatin bodies and four late replicating X chromosomes (Fig. 22-13).

47,XXY

The first subject reported by Sandberg and associates was an essentially normal fertile man of average intelligence who was detected only because he had a daughter with Down's syndrome. However, surveys in penal institutions suggested an increase in prevalence of this anomaly, especially in tall prisoners, and gave rise to an undeserved stereotype that has been modified by later studies. Among 43 47,XXY boys 1 to 12 years of age, ascertained by routine karyotype analysis in the newborn period, no clear-cut 47,XXY syndrome emerged in childhood. No major deviations could be attributed to an extra Y chromosome, with the possible exception of a skew to the left in I.Q. scores, although full-scale I.Q. scores ranged from 80 to 140. In general, 47,XXY patients tend to be easily distractible and hyperactive and to have low tolerance of frustration. The prevalence of antisocial and other behavioral difficulties is increased. 47,XXY is a common sex chromosome abnormality, occurring in 1 per 1000 male births; it is the only aneuploidy not selected against before birth. Among the features are tall stature, nodulocystic acne, and skeletal anomalies such as radioulnar synostosis. Sexual development is usually normal, and the rare reports of hypospadias in 47,XXY patients may be coincidental. The diagnosis should be suspected in hyperactive tall men with nodulocystic acne and can be confirmed by demonstration of two fluorescent Y bodies in somatic interphase nuclei or by karyotype analysis. The additional Y chromosome most commonly arises by nondisjunction at meiosis II.

48,XXYY

The variable phenotypes associated with this rare karyotype include multiple somatic abnormalities and mental retardation. The diagnosis is confirmed by finding three fluorescent Y bodies in interphase nuclei or by karyotype analysis.

Sex Differentiation Genes

The genes known to be involved in sex differentiation are listed in Table 22-18.
### Classification of Errors in Sex Differentiation (Continued)

#### Female Pseudohermaphroditism

Female pseudohermaphroditism ([Table 22-19](#)) is the easiest of the sexual anomalies to comprehend, because the ovaries and Müllerian derivatives are normal and anatomic abnormality is limited to the external urogenital sinus and genitalia. Because in the absence of testes there is an inherent tendency for the external genitalia to feminize, a female fetus is masculinized only if exposed to androgens. The degree of fetal masculinization is determined by the stage of differentiation at the time of exposure. Once the vagina has separated from the urogenital sinus (at about the 12th fetal week), androgens cause only clitoral hypertrophy.

Even with severe masculinization of the external genitalia, the uterus and fallopian tubes are normal, because regression of the primordia for these structures, the Müllerian ducts, requires secretion of AMH by fetal testes, and this action cannot be mimicked by androgens. Although the presence of virilized genitalia usually provides prima facie evidence of an androgenic influence during gestation, ambiguous genitalia, superficially resembling those produced by androgen, are an occasional feature of other, more generalized teratologic malformations.

#### Congenital Adrenal Hyperplasia

CAH accounts for most of the cases of female pseudohermaphroditism and approximately half of all patients with ambiguous external genitalia.

#### Biochemical Variants of Congenital Adrenal Hyperplasia

Mutations in five genes, four that encode biosynthetic enzymes for steroid hormone synthesis (CYP21, CYP17, CYP11B1, CYP11A1, and HSD3B2) and one that encodes the intracellular cholesterol transport protein (StAR, steroid acute regulatory protein), can cause CAH, each resulting in distinctive biochemical consequences and clinical features ([Fig. 22-54](#) ; see Chapter 14). All (but the rare CYP11A1 heterozygous mutation) are transmitted as autosomal recessive traits. The common denominator in all six biochemical defects is impaired cortisol secretion, which results in hypersecretion of corticotropin-releasing hormone (CRH) and corticotropin and consequent hyperplasia of the adrenal cortex. Only deficiencies of 21-hydroxylase (CYP21) and 11-hydroxylase (CYP11B1), however, are predominantly virilizing disorders. In patients with "classic" forms of these two enzymatic defects, the most striking abnormality of the sexual phenotype is masculinization of the female fetus because of overproduction of adrenal androgens.

---

### TABLE 22-18 -- Genes Involved in Sex Differentiation

<table>
<thead>
<tr>
<th>Gene</th>
<th>Encodes</th>
<th>Human Locus</th>
<th>Human Phenotype (Mutation or Deletion)</th>
<th>Mouse Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMH</td>
<td>Antimüllerian hormone; member of TGF- family; causes regression of müllerian ducts</td>
<td>19p13.3</td>
<td>46XY: Persistent müllerian duct syndrome, cryptorchidism, type I</td>
<td>KO: Persistent müllerian ducts, Leydig cell hyperplasia</td>
</tr>
<tr>
<td>AMH type 2 receptor (interacts with type 1 receptor)</td>
<td>Serine/threonine kinase: type II receptor for TGF- related proteins</td>
<td>12q13</td>
<td>46,XY: Persistent müllerian duct syndrome, type II</td>
<td>KO: Persistent müllerian duct syndrome</td>
</tr>
<tr>
<td>HCG</td>
<td>Beta-subunit of hCG</td>
<td>19p13.32</td>
<td>Not reported; presumably lethal</td>
<td>NA</td>
</tr>
<tr>
<td>HCG/LH receptor</td>
<td>Receptor</td>
<td>2p21</td>
<td>46XY: male pseudohermaphrodite</td>
<td>KO: Males normal at birth; postnatal arrest in sexual development</td>
</tr>
<tr>
<td>SIAR</td>
<td>Steroidogenic acute regulatory protein</td>
<td>8p11.2</td>
<td>46,XY: male pseudohermaphrodite with adrenal and gonadal steroid deficiency</td>
<td>KO: Mouse phenotype same as human</td>
</tr>
<tr>
<td>CYP11A1</td>
<td>Enzyme that catalyzes conversion of cholesterol to pregnenolone</td>
<td>15q23.24</td>
<td>46,XY: heterozygote, male pseudohermaphrodite, lateonset adrenal insufficiency</td>
<td>KO: (phenotype in the rabbit same as human, lethal)</td>
</tr>
<tr>
<td>HSD3B2</td>
<td>Enzyme that catalyzes conversion of C_{19}^- and C_{19}^-4'-steroids to 4'-steroids by the gonads and adrenals</td>
<td>1p13</td>
<td>46,XY: Severe deficiencymale pseudohormaphrodite with adrenal insufficiency, 46,XX: 17-corticosterone with adrenal insufficiency</td>
<td>Milder mutation: as above without clinical aldosterone or glucocorticoid deficiency</td>
</tr>
<tr>
<td>AR</td>
<td>Androgen receptor</td>
<td>Xq1112</td>
<td>46,XY: female ambiguous &quot;normal&quot; male genitalia</td>
<td>Same as human</td>
</tr>
<tr>
<td>SDR5A2</td>
<td>(5-reductase-2)</td>
<td>2p23</td>
<td>46,XY: male pseudohermaphrodite; pseudovaginal perineoscrotal hypospadia; normal Wolffian ducts; virilization at puberty</td>
<td>KO: 5-reductase 1 and 2 XY: no clinical abnormalities</td>
</tr>
</tbody>
</table>

---
<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Function</th>
<th>Chromosome Location</th>
<th>Female Pseudohermaphroditism (KO)</th>
<th>Male Pseudohermaphroditism (XY)</th>
<th>Male Normal</th>
<th>Female Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP19 (P450arom)</td>
<td>Enzyme that converts C19-steroids to C18-estrogens in the placenta (protecting the female fetus from masculinization), gonads, extraglandular tissues</td>
<td>15q21</td>
<td>46,XX: female pseudohermaphrodite virilization at puberty, tall stature, no female secondary sexual characteristics, ovarian cysts, osteoporosis</td>
<td>46,XY: normal male differentiation, tall stature, osteoporosis, elevated gonadotropins, lack of epiphyseal fusion, macro-orchidism</td>
<td>KO: normal sex differentiation</td>
<td></td>
</tr>
<tr>
<td>HSD3B1</td>
<td>Enzyme that catalyzes conversion of C21 and C23-steroids to 4-steroids in extra-adrenal and extraglandular tissues</td>
<td>1p13</td>
<td>Not reported, presumably lethal, 3HSD1 is expressed in placenta</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>CYP17 (P450c17)</td>
<td>Enzyme that catalyzes the 17-hydroxylation of 17,20-steroids to 4-steroids in the placenta (protecting the female fetus from masculinization)</td>
<td>10q2425</td>
<td>46,XY: male pseudohermaphrodite with hypertension (17-hydroxylase deficiency)</td>
<td>46,XX: normal differentiation with hypertension and primary gonadal failure (17-hydroxylase deficiency)</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>CYP21 (P450c11)</td>
<td>Enzyme that catalyzes the 21-hydroxylation of 17-hydroxyprogesterone to 11-deoxycortisol and 21-hydroxylation of progesterone to dehydroepiandrosterone by the adrenal and gonad</td>
<td>8q2122</td>
<td>46,XX: female pseudohermaphrodite with adrenal insufficiency (cortisol and aldosterone deficiency)</td>
<td>46,XY: sexual precocity (macrogenitosomia praecox), adrenal insufficiency</td>
<td>KO: normal genitalia, adrenal insufficiency</td>
<td></td>
</tr>
<tr>
<td>HSD17B3</td>
<td>Enzyme that converts androstenedione to testosterone in the testis</td>
<td>9q22</td>
<td>46,XY: male pseudohermaphrodite with virilization and ± gynecomastia at puberty</td>
<td>46,XX: no clinical manifestations</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>DHCR7</td>
<td>Enzyme (3-hydroxysteroid 7-reductase) that catalyzes the conversion of 7-dehydrocholesterol</td>
<td>11q1213</td>
<td>46,XY: ambiguous genitalia, hypospadias type II</td>
<td>KO: resembles severe form of Smith-Lemli-Opitz syndrome, craniofacial anomalies, intrauterine growth retardation, death in first day of life secondary to poor sucking reflex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WNT7A</td>
<td></td>
<td>Not reported in the human</td>
<td>XY: persistence of the müllerian ducts due to lack of AMH receptor</td>
<td>XX: lead to abnormal development of the müllerian ducts and uterus resulting in infertility</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WNT5A</td>
<td></td>
<td>Not reported in the human</td>
<td>XY: absence of the phallus</td>
<td>XX: lack of external genitalia</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

KO, Knockout; AMH, antimüllerian hormone; hCG, human chorionic gonadotropin; TGF-, transforming growth factor-beta.

---

**TABLE 22-19 – Classification of Female Pseudohermaphroditism**

1. **Androgen-Induced**
   A. Fetal source
      1. Congenital adrenal hyperplasia
         a. Virilism only, defective adrenal 21-hydroxylation (CYP21)
         b. Virilism with salt-losing syndrome, defective adrenal 21-hydroxylation (CYP21)
         c. Virilism with hypertension, defective adrenal 11-hydroxylation (CYP11B1)
         d. Virilism with adrenal insufficiency, deficient 3-HSD 2 (HSD3B 2)
   2. P450 aromatase (CYP19) deficiency
   3. Glucocorticoid receptor gene mutation

B. Maternal source
   1. Iatrogenic
      a. Testosterone and related steroids
      b. Certain synthetic oral progestagens and rarely diethylstilbestrol
   2. Virilizing ovarian or adrenal tumor
   3. Virilizing luteoma of pregnancy
   4. Congenital virilizing adrenal hyperplasia in mother

C. Undetermined source
   1. Virilizing luteoma of pregnancy
and androgen precursors. Affected males have no abnormalities of the genitalia. These autosomal recessive disorders of steroid biosynthesis, of which more than 90% are caused by 21-hydroxylase deficiency, are discussed in this section as causes of female pseudohermaphroditism.

Patients with 3-HSD, CYP17, CYP1A1, or STAR deficiencies have defects that not only block cortisol synthesis but also impair the production of gonadal steroids by the gonads and by the adrenal glands. Affected males have varying degrees of male pseudohermaphroditism because of deficient testosterone production by the fetal Leydig cells, whereas affected females may or may not exhibit virilization. If present, virilization in females is much less severe than in CYP21 and CYP11B1 deficiencies. These forms of CAH in the male are discussed in the section on male pseudohermaphroditism. Administration to the pregnant rat of selective synthetic inhibitors of the enzymes involved in adrenal and testicular steroid biosynthesis produced abnormalities of sex differentiation in the offspring that are the counterparts of CAH in humans and

![Figure 22-53 Female pseudohermaphroditism induced by prenatal exposure to androgens. Exposure after 12th fetal week leads only to clitoral hypertrophy (diagram on left). Exposure at progressively earlier stages of differentiation (depicted from left to right in drawings) leads to retention of the urogenital sinus and labioscrotal fusion. If exposure occurs sufficiently early, the labia fuse to form a penile urethra. (From Grumbach MM, Ouchterlm JR. The effects of androgens on fetal sexual development: androgen-induced female pseudohermaphroditism. Front Pediatr 1965; 11:157-185. Reproduced with permission of the publisher. © The American Pediatric Society.)](image)

served to clarify the role of steroidogenic enzymes in the control of fetal sex differentiation. CYP21 (21-Hydroxylase) Deficiency

Simple Virilizing Form of CYP21 Deficiency.

Deficiency of CYP21 (CYP21A2; cytochrome P450C21), the most common cause of ambiguous genitalia in infants, is inherited (as are the other forms) as an autosomal recessive trait. The simple virilizing form of CYP21 deficiency has an incidence of about 1 per 50,000 persons and accounts for approximately 25% of subjects with CYP21 deficiency. Deficiency in the 21-hydroxylation of 11-deoxycorticosterone results in elevated circulating testosterone levels. Before the 12th week of gestation, high fetal androgen levels lead to a varying degree of labioscrotal fusion and clitoral enlargement in the affected female fetus; exposure to androgen after week 12 causes isolated clitoromegaly. (See Rule of Androgenic Dominance in the Differentiation of the External Genitalia and Urogenital Sinus.) Exposure to excess androgens in the male during gestation can result in subtle penile enlargement (e.g., testotoxicosis, virilizing adrenal hyperplasia).

The genitalia of females with the virilizing forms of CAH (CYP21 and CYP11B1) may exhibit a spectrum of masculinization from simple enlargement of the clitoris to complete labioscrotal fusion with a penile urethra (see Fig. 22-53). In severe cases the urogenital sinus is usually preserved and serves as a common outlet for both the urethra and vagina. The hypersecretion of androgens and androgen precursors begins weeks before the 12th week of gestation, especially in patients

![Figure 22-54 Diagram of the steroid biosynthetic pathways and the biosynthetic defects that result in congenital adrenal hyperplasia. The defect in patients with "lipoid adrenal hyperplasia" is not (except for one reported case) in the CYP11A1 (cholesterol side-chain cleavage) enzyme but in STAR, the sterolgenic acute regulatory protein. This protein is involved in the transport of cholesterol from the outer mitochondrial membrane to the inner membrane where the CYP11A1 enzyme is located. CYP11B1 (11-hydroxylase) catalyzes 11-hydroxylation of deoxycorticosterone and 11-deoxycorticisol primarily. CYP17 (17-hydroxylase/17,20-lyase) catalyzes both 17-hydroxylation and splitting of the 17,20 bond, but for the latter it has preferential \( ^{17} \)-17-hydroxylation activity (see text). CYP19 (aromatase) catalyzes the conversion of androstenedione to estrone and testosterone to estradiol. CYP11B2 (aldosterone synthetase) catalyzes the conversion of corticosterone to aldosterone. 3-HSD I and 3-HSD II, 3-hydroxysteroid dehydrogenase/isomerase types I and II; CYP21 (P450C21), 21-hydroxylase; 17-HSD 3, 17-hydroxysteroid dehydrogenase type 3. In the human, deletion or a homoygous null mutation of CYP11A1 (P450scs) is probably lethal in utero but a heterogeneous mutation caused congenital lipoid adrenal hyperplasia (see text).](image)

who manifest more than simple clitoromegaly. In addition to the traditional pathway, Auchs has raised the possibility that 5-reduced \( ^{5} \)-steroids such as 17-hydroxyloiprogrenin may be an important substrate for conversion to dihydrotestosterone by peripheral tissues. The uterus and fallopian tubes (mullerian structures) and the ovaries are normally formed, except in rare cases. Wolffian duct development is consistently absent regardless of the degree of virilization of the external genitalia in affected females. Thus, internal genital masculinization corresponds to gonadal sex in both affected females and males.

Postnatally, secretion of testosterone by the adrenal gland and conversion of androstenedione to testosterone in peripheral tissues result in continued virilization of the untreated patient. In the simple virilizing form of CYP21 deficiency, the 21-hydroxylation of \( ^{17} \)-deoxycorticosteroids and \( ^{17} \)-deoxycorticisol is primarily impaired. However, even in patients with "mild," late-onset CYP21 deficiency the 21-hydroxylation of mineralocorticoids is defective, as evidenced by elevated plasma 21-deoxycorticosterone levels after corticosteroid stimulation. Untreated patients with simple virilizing CYP21 deficiency usually, but not always, have normal plasma renin levels and normal aldosterone secretion rates. Untreated patients, increased androgen production leads to the early appearance of pubic hair, acne, clitoromegaly (or penile enlargement in the male), increased muscular development, other signs of virilization, rapid growth during childhood, and disproportionate increase in the rate of skeletal maturation, which results in premature closure of the epiphyses and short stature in adolescence and adulthood.

In patients with severe CYP21 deficiency, both virilization and salt loss can occur. This variant, which occurs in about 75% of patients with classic CYP21 deficiency (1 in 15,000 live births), is caused by a severe or complete defect in adrenal 21-hydroxylation that leads to impaired cortisol (adrenal fasciculata) and aldosterone (adrenal glomerulosa) secretion and increased plasma renin. Electrolyte and fluid losses due to aldosterone deficiency cause hyponatremia, hyperkalemia, acidosis, dehydration, vascular collapse, and, if untreated, death. About 50% of patients have the first salt-losing adrenal crisis at between 6 and 14 days of age; the"crisis" usually occurs as early as 6 to 12 weeks of age, usually in association with a concomitant stress. Masculinization of the external genitalia and urogenital sinuses in affected females tends to be more severe in complete CYP21 deficiency than in simple
TABLE 22-20  Incidence of Classic Congenital Virilizing Adrenal Hyperplasia (CYP21 Deficiency) After Screening

<table>
<thead>
<tr>
<th>Population</th>
<th>Number of Newborns Screened</th>
<th>Newborns Affected/Live Births</th>
<th>Incidence by Case Survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alaska</td>
<td>1,131</td>
<td>1/282</td>
<td>1/480</td>
</tr>
<tr>
<td>La Réunion, France</td>
<td>31,472</td>
<td>1/3,147</td>
<td></td>
</tr>
<tr>
<td>Rome, Italy</td>
<td>22,400</td>
<td>1/5,600</td>
<td></td>
</tr>
<tr>
<td>Lille (Lyon), France</td>
<td>199,624</td>
<td>1/11,090</td>
<td>1/23,000</td>
</tr>
<tr>
<td>Illinois</td>
<td>357,825</td>
<td>1/11,928</td>
<td>1/15,000 Wisconsin</td>
</tr>
<tr>
<td>Sweden</td>
<td>370,000</td>
<td>1/12,758</td>
<td></td>
</tr>
<tr>
<td>Portugal</td>
<td>100,000</td>
<td>1/14,285</td>
<td></td>
</tr>
<tr>
<td>Emilia-Romagna, Italy</td>
<td>73,000</td>
<td>1/14,600</td>
<td></td>
</tr>
<tr>
<td>Scotland</td>
<td>119,960</td>
<td>1/17,137</td>
<td>1/20,307</td>
</tr>
<tr>
<td>Washington</td>
<td>255,527</td>
<td>1/18,251</td>
<td>1/15,000 Wisconsin</td>
</tr>
<tr>
<td>New Zealand</td>
<td>168,965</td>
<td>1/18,773</td>
<td></td>
</tr>
<tr>
<td>Japan</td>
<td>585,000</td>
<td>1/20,892</td>
<td>1/43,674</td>
</tr>
</tbody>
</table>


CYP21. Without specific therapy, death can result from hyperkalemia, dehydration, hypoglycemia, and shock. In the affected male whose genitalia are normal, the differential diagnosis includes septal, pyloric stenosis, gastroenteritis, congenital heart disease, congenital adrenal hypoplasia, pseudohypopaldosteronism, and isolated defects in mineralocorticoid biosynthesis.

A newly identified abnormality in patients with classic 21-hydroxylase deficiency (especially the salt-losing form) or 11-hydroxylase deficiency is chronic adrenomedullary hypofunction as initially detected by subnormal plasma concentrations of epinephrine and free metanephrine. Adrenomedullary function correlates highly with the genotype in 21-hydroxylase deficiency. The concentration of plasma total metanephrine and especially free metanephrine was lowest in the salt-losing form, and it correlated with the expected 21-hydroxylase activity according to the genotype. It seems that determination of plasma free metanephrine is a useful indicator of the severity of 21-hydroxylase deficiency/cyan inexpensive biochemical marker that is available in commercial laboratories. More recently, children with these adrenal disorders and adrenomedullary hypofunction were reported to have increased serum leptin levels and evidence of insulin resistance. The adrenomedullary hypofunction is attributed to a low concentration of intra-adrenal cortisol and developmental defects in the formation of the adrenal medulla. Although the possibility of potential adverse effects of the adrenomedullary hypofunction and its documented biochemical consequences has been raised, these effects remain to be documented.

Nonclassic CYP21 Deficiency.

Studies of families affected with 21-hydroxylase deficiency have revealed heterogeneity in the biochemical and clinical manifestations of this condition, including asymptomatic as well as symptomatic “late-onset” disease. Affected females have no genital ambiguity at birth, in contrast to those with classic CYP21 deficiency, but can manifest symptoms of androgen excess, such as premature development of pubic hair in early childhood, accelerated linear growth, and bone maturation with resulting short stature, cystic acne, male-pattern baldness, hirsutism, menstrual abnormalities, and infertility. Symptoms and findings in women may be similar to those in women with polycystic ovary disease.

In affected males, premature development of pubic hair, beard growth, growth spurt, and phallic maturation can occur prepubertally as a result of the increased adrenal androgen production. Oligospermia and decreased fertility have also been attributed to late-onset CYP21 deficiency. Mostly, affected males are entirely asymptomatic, having been detected serendipitously as a result, for example, of family studies.

In studies of families with classic CYP21 deficiency, some individuals have hormonal criteria of this disease (i.e., elevated basal and/or corticotropin-stimulated 17-hydroxyprogesterone, androstenedione, and testosterone levels) but no clinical signs of androgen excess. These patients have been designated as having “cryptic” CYP21 deficiency. Symptoms of hyperandrogenism may wax and wane in these patients, who therefore may not be truly asymptomatic over a period of time.

New and co-workers, using corticotropin-induced rises in 17-hydroxyprogesterone, have defined hormone reference data in the form of a nomogram. Their studies provide a means of distinguishing patients with classic CYP21 deficiency from those with milder variant forms, heterozygotes, and normal individuals. These hormonal data in conjunction with linkage studies with human leukocyte antigen (HLA) typing, and in some instances, genotyping, indicated that all three variants classic CYP21 deficiency and the cryptic and late-onset variants resulted from a defect in the same gene. The features of the cryptic and late-onset forms can occur in families with the classic disease and can be present in the same patient at different times of life. The cryptic and late-onset forms arise from allelic variants of the gene that causes classic CYP21 deficiency. Genotypically, patients may be compound heterozygotes with a classic mutation and a variant allele, or they may have two variant alleles. Using hormonal data and linkage studies, Spelzer and co-workers estimated that nonclassic CYP21 deficiency is the most common autosomal recessive disorder in humans, the disease frequency being 0.1% to 0.2% in all ethnic groups. The gene frequency is higher in Ashkenazi Jews (3%-4%), and in Hispanics and Yugoslavs (1%-2%).

Molecular Genetics of CYP21 Deficiency.

CYP21, a member of the cytochrome P450 superfamily of mono-oxygenases, is a heme-containing enzyme that is bound to endoplasmic reticulum and receives electrons from NADPH by way of a flavoprotein, P450 reductase. The gene for this enzyme is located within the major HLA locus on the short arm of chromosome 6 (6p21.3). HLA types are co-dominantly inherited and can be used in informative families (those containing an affected child and the parents) to distinguish homozygotes, heterozygotes, and unaffected individuals. Although a wide variety of HLA antigens and haplotypes is found in affected patients, genetic disequilibrium has been found for certain specific types and haplotypes. In particular, HLA-Drw7 has a high degree of association with the salt-losing form of CYP21 deficiency, and HLA-DR14 occurs more frequently in patients with the nonclassic form of the disease.

Two CYP21 genes are located on each chromosome 6 between HLA-B and HLA-DR (Fig. 22-57) (Figure Not Available), a functional CYP21 gene and a nonfunctional CYP21P (CYP21A1P) pseudogene. The CYP21 gene, the gene for the fourth component of complement (C4), and a gene for a large extracellular matrix protein called tenasin are duplicated in tandem in the sequence C4A, CYP21P, XA, C4B, CYP21, and X8. This locus is one of the most complex and polymorphic in the human genome. The CYP21 genes and X genes overlap one another on opposite strands of DNA. The CYP21 and
In affected patients, plasma 17-hydroxyprogesterone values usually range from 90 to 1200 nmol/L (3000 to 40,000 ng/dL), depending on age and the severity of the disease. Infants who are sick, and especially premature infants, may have elevated androstenedione and 17-hydroxyprogesterone levels that can confound the diagnosis. Measurement of plasma 17-hydroxyprogesterone and androstenedione establishes the diagnosis in affected infants and children. The level of plasma 17-hydroxyprogesterone normally decreases within 2 to 3 weeks after birth and can be followed as a guide to the adequacy of the medical treatment. The karyotype should be obtained. Pelvic ultrasonography or MRI is useful to determine the nature of internal genital structures. An elevated concentration of plasma androstenedione also establishes the diagnosis in affected females. The level of plasma androstenedione normally decreases within 2 to 3 weeks after birth and can be followed as a guide to the adequacy of the medical treatment. The karyotype should be obtained. Pelvic ultrasonography or MRI is useful to determine the nature of internal genital structures.

Suspected CYP21 deficiency includes measurement of plasma electrolytes and glucose, 17-hydroxyprogesterone, androstenedione, and testosterone levels. An elevated concentration of plasma androstenedione and elevation of electrolytes and glucose establishes the diagnosis in affected females. The level of plasma androstenedione normally decreases within 2 to 3 weeks after birth and can be followed as a guide to the adequacy of the medical treatment. The karyotype should be obtained. Pelvic ultrasonography or MRI is useful to determine the nature of internal genital structures. An elevated concentration of plasma androstenedione also establishes the diagnosis in affected females. The level of plasma androstenedione normally decreases within 2 to 3 weeks after birth and can be followed as a guide to the adequacy of the medical treatment. The karyotype should be obtained. Pelvic ultrasonography or MRI is useful to determine the nature of internal genital structures.

Newborn screening programs for CYP21 deficiency have been instituted in a number of regions and countries by measurement of heel-blot 17-hydroxyprogesterone levels. A blood heel-blot specimen is obtained at 3 to 5 days of life in conjunction with screening for hypothyroidism and a variety of other inborn metabolic diseases. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Furthermore, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world. Nevertheless, inclusion of CAH in populationwide newborn screening programs is not universal throughout North America or in many developed countries around the world.
17-hydroxyprogesterone (the “gold standard”), androstenedione, and 21-deoxycortisol levels identifies affected infants. 

In the past, diagnosis of CYP21 deficiency was based on assessment of the excretion of urinary 17-ketosteroids and pregnanetriol. The excretion of 17-ketosteroids varies with age, and in the first few days of life in unaffected infants it can be as high as 7 to 14 μmol/day (2 to 4 mg/24 hr). After 1 month of age, urinary 17-ketosteroid levels decrease to an upper limit of approximately 1.6 μmol/day (0.5 mg per year of age) until the onset of adrenarche. An elevated level of pregnanetriol, the metabolite of 17-hydroxyprogesterone, is a hallmark of 21-hydroxylase deficiency. However, in the neonatal period the excretion of urinary pregnanetriol may be within the normal range in affected infants. Thereafter, the levels rise and are useful diagnostically.

Infants with salt wasting usually have clinical evidence of frank or incipient adrenal insufficiency or crisis after the sixth day of life and especially during the second week. Early diagnosis of the salt-losing form of CAH is usually based on the clinical findings of poor feeding, weight loss, vomiting, hypotension, hyperkalemia, and often renal acidosis. The plasma concentration and excretion of aldosterone are low, and plasma renin activity is high. Mild salt losers may have normal electrolytes under basal conditions but exhibit elevated plasma renin activity and hypertension, hyperkalemia, and inappropriate natriuresis with salt restriction.

Prenatal Diagnosis and Therapy. 

The striking elevation in the level of plasma 17-hydroxyprogesterone is such a distinctive marker of 21-hydroxylase deficiency that prenatal diagnosis has been attempted by determining its concentration in amniotic fluid at risk. 

The success of prenatal diagnosis has led to attempts at prenatal treatment. Dexamethasone crosses the placenta and suppresses the fetal adrenal gland if given in sufficient doses; unlike cortisol or prednisolone, the placenta does not convert the 11-hydroxy group of dexamethasone to the inactive 11-keto group. Dexamethasone administration to pregnant women early in gestation (starting at 4 to 6 weeks) can decrease the virilization of the external genitalia in the majority of affected female infants.

The use of prenatal dexamethasone therapy in pregnancies with a fetus at risk for CYP21 deficiency is not without problems and remains controversial. 

Striking improvement in mortality and morbidity of patients with CAH has occurred over the past 50 years following the introduction of glucocorticoid and mineralocorticoid therapy. 

Treatment.
Steroid replacement treatment is usually initiated for 7 to 10 days with doses in excess of maintenance to suppress more rapidly elevated ACTH concentrations and consequent hyperandrogenism and to reduce the size of the hyperplastic adrenal glands. Lack of recognition of the true cortisol secretion rate in infants (estimated daily cortisol production about 8 mg/m² per day), and of the efficient absorption of orally administered hydrocortisone has also contributed to excess treatment. Consequently, for the initiation of glucocorticoid treatment (as opposed to maintenance therapy) we use high doses of hydrocortisone, 50 mg/m² daily (12.5 mg/day for the neonate) divided into an every-6-hour dose for five to seven days, together with fludrocortisone, 100 to 200 μg daily depending on the concentration of serum Na⁺ and the blood pressure. Salt is added to the feedings (17 to 34 mEq of sodium chloride to 2 g). Hydrocortisone is the preferred preparation in tablet form, which can be crushed or split before administration. The oral suspension of hydrocortisone is unstable and has been discontinued.

Ongoing treatment needs to be monitored using a combination of clinical indices such as growth velocity, signs of glucocorticoid excess, and skeletal maturation as determined by bone age measurements, but also biochemical indices to help modulate treatment on a shorter-term basis. Because 17-OH-progesterone is a sensitive index of 21-hydroxylase deficiency, the steroid has also been used as a monitor of control. However, it is subject to a profound diurnal variation in secretion, so that it is important to standardize the hour of sampling. The development of filter paper blood spot and saliva assays to measure 17OH-progesterone has led to the use of daily profiles to monitor treatment, the samples often conveniently collected at home by the family. However, the availability of blood spot and saliva assays for profiling and routine use is limited (e.g., it is not generally available in North America) and its cost effectiveness has not been established. Serial plasma androstenedione and testosterone measurements are also useful to monitor control in children and women, although testosterone cannot be used in boys from puberty onward in view of the predominance of testicular testosterone secretion. With careful monitoring, especially during growth and skeletal maturation, it is usually possible to maintain control during childhood with hydrocortisone doses in the range of 10 to 15 mg/m² per day divided into three equal doses and even less in the milder forms of CAH. The cortisol secretory rate in premature and term infants is 7 to 9 mg/m² per day and in normal children and adolescents is about 6 mg/m² per day. Longer-acting preparations of glucocorticoids such as prednisolone (2 to 4 mg/m² per day about 1/5 the dose of hydrocortisone) or dexamethasone (0.25 mg/m² per day as a single dose) are generally only used after linear growth has ceased because of the greater risk of complications of glucocorticoid excess such as suppressed growth velocity, obesity, and other cushingoid features. These preparations may be particularly useful in the postpubertal female with amenorrhea or irregular menses.

All salt-losers should be treated with fludrocortisone at the time of diagnosis in infancy because it decreases ACTH and vasopressin secretion and reduces the maintenance dose of hydrocortisone. The adequacy of mineralocorticoid replacement is best monitored by plasma renin determinations and blood pressure. Excess mineralocorticoid therapy can result in hypertension and congestive heart failure. Although there is a tendency for salt-losing to improve with age, this may only have occurred owing to the patient compensating by increasing dietary salt. Even though, by definition, aldosterone deficiency is not clinically apparent in the simple virilizing form of 21-hydroxylase deficiency, in a majority of patients it is useful to add fludrocortisone to the glucocorticoid regimen, especially in patients difficult to control on glucocorticoid therapy alone. The addition of mineralocorticoid often leads to a reduction in the dose of glucocorticoid. Increased circulating levels of progesterone and 17-hydroxyprogesterone are mineralocorticoid antagonists and their action on the renal mineralocorticoid receptor is blocked by the potent fludrocortisone.

Long-term follow-up studies on the effects of glucocorticoid and mineralocorticoid replacement in patients with CAH indicate that the mean adult height of both males and females was less than that of unaffected siblings and less than the normal mean adult height. More recent studies of final height indicate values more likely to be within the normal population range but still below target height. A meta-analysis of data from 561 patients with classic CAH (from 18 centers) found a mean final height of -1.4 SD less reduction in adult height than generally thought. Noteworthy, a randomized prospective crossover study showed that CAH patients treated with 15 mg/m² per day of hydrocortisone (in the lower dosage range) had a greater growth velocity compared with those given doses of 25 mg/m² per day, an observation that supports the use of the lower dose to avoid the effect of excessive glucocorticoid on growth, despite higher plasma concentrations of 17-hydroxyprogesterone. The aim is to reduce plasma 17-hydroxyprogesterone plasma levels to between 400 and 1000 ng/dL reductions to the normal range require excessive doses of glucocorticoid, especially because final height does not correlate well with the magnitude of suppression of adrenal androgens. It is likely that circulating bone and cartilage estradiol concentrations arise from conversion of adrenal androgens to estrogen by aromatase and have a role in the rate of skeletal maturation and growth. Early diagnosis with the avoidance of excessive glucocorticoid treatment and careful monitoring of compliance can lead to an adult height within the normal range for target height.

A retrospective study reported that the prevalence of learning disorders was increased in children with the salt-losing form, probably related to unrecognized hypoglycemia and to electrolyte derangements. Retrospective studies in adult women have described a high prevalence of an inadequate introitus, lack of or decreased interest in sexual activity, and a below-average proportion who were married or had sexual partners. More recent observational data suggest improvement in surgical outcome, but long-term outcome studies of the improved surgical techniques are not yet available.

The fertility rate among heterosexual sexually active women, especially those with the salt-losing form in one of these studies, was low. However, more recent studies of women who have had the advantage of improved medical, surgical, and psychological treatment and guidance report an improvement in fertility rate. An apparent increase in the frequency of homosexual and bisexual fantasies and an increased propensity for homosexual and bisexual behavior has been reported. Gender identity remains female. Although fertility can occur in untreated adult males with CAH, those who discontinue therapy or are noncompliant are at risk for tumors from adrenal rests (see Chapter 24). The testicular adrenal rest tumors are usually responsive to glucocorticoid treatment. Some rests contain LH receptors, which may be a cause of unresponsiveness to glucocorticoids.
adrenal rests can be reversed when glucocorticoid therapy is reinstituted. In some patients, surgical ablation of enlarged adrenal rests has been indicated.

Noncompliant patients are at increased risk for adrenal adenoma or carcinoma and adrenal incidentalomas, which may occur in more than 70% of affected adults, including those with the nonclassic form of CAH and even heterozygotes.

It is recommended that all patients receive treatment throughout life with a glucocorticoid and, if indicated, a mineralocorticoid (Fig. 22-60 : Table 22-21 ).

With the improved fertility rate with modern medical and surgical management of affected females, one can look forward to an increased number of successful pregnancies. When possible, these patients should be managed in a tertiary center with the facilities, personnel, and experience to care for high-risk pregnancies. It is important to use glucocorticoids that are metabolized by placental 11-hydroxysteroid dehydrogenase II (e.g., hydrocortisone, prednisone, prednisolone, not dexamethasone). Table 22-22 contains guidelines for the care of women with classic CAH during pregnancy and delivery.

Genital surgery should not be undertaken until the parents provide informed consent after discussion and thorough review of clinical status; issues surrounding the need for and the type of surgery, the experience and skill of the surgical team, and the surgery to be undertaken; and the options. Emerging clinical evidence suggests that the recommended time of genitoplasty is at age 2 to 6 months, especially for infants with a high proximal junction of the vagina and urethra. Citroplasty is the procedure of choice, not clitoridectomy, and it merits thorough consideration. If clitoridoplasty is undertaken, it is essential, as indicated by long-term outcome studies, to preserve the neurovascular bundle, the glans, and surrounding skin. In a retrospective study of cosmetic and anatomic outcomes in a selected group of women (which included those with CAH) who had feminizing genital surgery in childhood, nearly half had a poor cosmetic result and nearly all the patients needed further treatment for tampon use or intercourse. The time of revision of the vaginoplasty in adolescence is the decision of the patient and, if undertaken before 18 years of age, the patient and the family. The patient and parents must be reassured that with appropriate treatment and compliance the child will grow and develop into a normal, functional adult. Fertility in males and feminization, menstruation, and fertility in females can be expected in the adequately treated patient.

Labor and Delivery

Stress Dose Glucocorticoid Therapy

A soluble hydrocortisone ester (up to 50100 mg IV every 8 hr) should be given at the initiation of active labor and continued until after delivery, followed by a rapid taper to previous maintenance doses.

Cesarean versus Vaginal Delivery

Android pelvic characteristics may increase the risk for cephalopelvic disproportion.

Elective cesarean section should be considered in all patients, especially those who have had reconstructive genital surgery.

Evaluation of the Infant

Examine the infant for ambiguous genitalia. Female pseudohermaphroditism may be a consequence of either maternal hyperandrogenism or, if the father is a carrier, fetal 21-hydroxylase deficiency. (Male infants may have enlarged external genitalia.)

If the external genitalia are ambiguous, appropriate laboratory studies in the infant should be carried out to exclude 21-hydroxylase deficiency.


Suppression

Use a glucocorticoid that is metabolized by placental 11-hydroxysteroid dehydrogenase-II (e.g., hydrocortisone, cortisol acetate, prednisone, methylprednisolone).

Assess clinical status, serum electrolytes, and serum androgen levels regularly to determine the need for increased glucocorticoid and/or mineralocorticoid therapy. Excessive nausea, salt craving, and poor weight gain suggest adrenal steroid insufficiency. In select cases, measurement of plasma renin activity may be helpful.

Serum testosterone and free testosterone levels should be measured every 6 weeks in the first trimester and every 6 to 8 weeks thereafter. Target free testosterone levels to the high normal range for pregnancy; however, management must be individualized for each patient. Avoid inducing cushingoid effects from too high a dose of glucocorticoids.

Fetal gender determination by ultrasonography may be helpful in guiding treatment goals, because maternal androgen excess will have minimal effects on the male fetus.

TABLE 22-22 -- Suggested Guidelines for the Management of Women with Classic Congenital Adrenal Hyperplasia Owing to 21-Hydroxylase Deficiency during Pregnancy and Delivery

<table>
<thead>
<tr>
<th>Gestational Management</th>
<th>Adrenal Steroid Replacement and Adrenal Androgen Suppression</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Use a glucocorticoid</strong></td>
<td><strong>Assess clinical status, serum electrolytes, and serum androgen levels regularly to determine the need for increased glucocorticoid and/or mineralocorticoid therapy.</strong></td>
</tr>
<tr>
<td>that is metabolized by</td>
<td><strong>Excessive nausea, salt craving, and poor weight gain suggest adrenal steroid insufficiency. In select cases, measurement of plasma renin activity may be helpful.</strong></td>
</tr>
<tr>
<td>placental 11-hydroxysteroid</td>
<td><strong>Serum testosterone and free testosterone levels should be measured every 6 weeks in the first trimester and every 6 to 8 weeks thereafter. Target free testosterone levels to the high normal range for pregnancy; however, management must be individualized for each patient. Avoid inducing cushingoid effects from too high a dose of glucocorticoids.</strong></td>
</tr>
<tr>
<td>dehydrogenase-II (e.g.,</td>
<td><strong>Fetal gender determination by ultrasonography may be helpful in guiding treatment goals, because maternal androgen excess will have minimal effects on the male fetus.</strong></td>
</tr>
<tr>
<td>hydrocortisone, cortisol</td>
<td><strong>Labor and Delivery</strong></td>
</tr>
<tr>
<td>acetate, prednisone,</td>
<td><strong>Stress Dose Glucocorticoid Therapy</strong></td>
</tr>
<tr>
<td>methylprednisolone).</td>
<td>A soluble hydrocortisone ester (up to 50100 mg IV every 8 hr) should be given at the initiation of active labor and continued until after delivery, followed by a rapid taper to previous maintenance doses.</td>
</tr>
</tbody>
</table>

**Cesarean versus Vaginal Delivery**

Android pelvic characteristics may increase the risk for cephalopelvic disproportion.

Elective cesarean section should be considered in all patients, especially those who have had reconstructive genital surgery.

**Evaluation of the Infant**

Examine the infant for ambiguous genitalia. Female pseudohermaphroditism may be a consequence of either maternal hyperandrogenism or, if the father is a carrier, fetal 21-hydroxylase deficiency. (Male infants may have enlarged external genitalia.)

If the external genitalia are ambiguous, appropriate laboratory studies in the infant should be carried out to exclude 21-hydroxylase deficiency.


*In an affected mother with 21-hydroxylase deficiency, we estimate the risk of having an affected female infant with fetal congenital adrenal hyperplasia to be approximately 1 in 240 based on an estimated 1 in 60 incidence for heterozygous individuals (see text for further details).
advancement in skeletal maturation with premature epiphyseal closure. An effective intermediate dose cannot consistently be achieved in some patients, especially those with the severest form of CAH associated with null mutations. Adolescents and adults with this disorder who have been treated from infancy are being assessed currently in many clinics, and despite the improvement in survival, the overall results, even with acceptable compliance with the treatment, are disappointing in terms of growth, obesity, bone mineral density, and, in women, fertility, especially in those with severe CYP21 deficiency. These considerations have led to a reassessment of therapy and the proposal to test and assess surgical and novel medical treatments.

Adrenoleukopenia by laparoscopy is increasingly being used in a small number of centers for females with null CYP21 gene mutations, on the basis that physiologic glucocorticoid and mineralocorticoid substitution therapy may be easier in the absence of ACTH-induced hypersecretion of abnormal amounts of androgens and estrogen. Histologically, adrenal tissue appears to be normal, and this controversial procedure is now safely undertaken by laparoscopy and normal menses, and body composition can be restored in poorly controlled adult females. Recurrence of virilization has been reported after adrenalec-tomy in females, the source of androgens being ovarian or ectopic adrenal tissue.

A second experimental approach involves the use of pharmacologic agents to improve skeletal growth. Trials are in progress to study the effects of third-generation aromatase inhibitors to reduce the conversion of androgens to estrogens in conjunction with an antiandrogen. The usefulness of aromatase inhibitors such as letrozole or anastrozole to block estrogen production in the periphery was suggested by the finding of delayed skeletal maturation in a man with estrogen resistance and men and women with severe generalized aromatase (CYP19) deficiency, indicating a critical role for estrogen in epiphyseal maturation and fusion. Pharmacologic blockade of estrogen production and androgen action may make it possible to reduce the dose of glucocorticoid to a physiologic level. A 2-year interim analysis of a study of children with CAH given testolactone (a weak aromatase inhibitor) and flutamide (an antiandrogen) showed a reduction in the dose of hydrocortisone (8.6 mg/m² per day versus 13.5 mg in treated and control groups, respectively) and improved short-term growth and skeletal maturation.

The long-term outcome of this approach and the safety of such polypharmacy for the treatment of CAH will require careful evaluation and monitoring. Experience with LHRH agonists in true precocious puberty has led to a trial in a small cohort of short CAH patients of treatment with hGH with or without an LHRH agonist; this approach improved growth velocity, and predicted final height and long-term data are awaited with interest.

Other potential future therapies may involve the application of corticotropin-releasing factor (CRF or CRH) receptor antagonists that already are showing promising results in the treatment of psychiatric disorders. Because receptors for CRF are widely expressed, the challenge will be to use antagonists selective for the role of CRF in pituitary-adrenal function.

The management of single gender disorders such as CAH can benefit from the advances in gene therapy. Restoration of normal adrenal steroidogenesis was reported in homozygous 21-hydroxylase-deficient mice given intra-adrenal injections of an adrenoviral vector encoding the genomic sequence of human CYP21. There are clearly issues to resolve not only about efficient, safe, stable, tissue-specific gene delivery systems but also about the practicalities of serial intraadrenal injections in patients with CAH. Nevertheless, the adrenal may be a "privileged site" for gene therapy.

CYP11B1 (11-hydroxylase) Deficiency: Virilization with Hypertension

Cytochrome P450 c11 is located in the mitochondrial inner membrane. It serves as a terminal oxidase of an electron transport chain that includes adrenodoxin reductase and adrenodoxin. There are two distinct 11-hydroxylase enzymes. CYP11B1 (P450 c11) and CYP11B2 (P450 c11a), encoded by two genes located in tandem at 8p22-22.1. The genes contain nine exons and are 93% identical. CYP11B1 encodes the 11-hydroxylase, which converts 11-deoxycorticosterone to corticosterone and 11-deoxycorticisol to cortisol. CYP11B2 encodes the enzyme aldosterone synthetase, which catalyzes the conversion of deoxycorticosterone to corticosterone and corticosterone to aldosterone. CYP11B1 has little 18-hydroxylation and 18-oxidation activity.

CAH resulting from CYP11B1 deficiency was first described by Eberlein and Bongiovanni and accounts for 5% to 8% of patients with CAH, occurring in about 1 of 10,000 births of European extraction. However, there is an increased incidence of 11-hydroxylase deficiency among Sephardic Jews from Morocco (1 in 5,000 to 1 in 7,000) and Arabs. In its classic form, CYP11B1 deficiency impairs conversion of 11-deoxycorticisol to cortisol and of deoxycorticosterone (DOC, a mineralocorticoid) to corticosterone in the zona fasciculata of the adrenal gland, resulting in accumulation of these steroid precursors. The cortisol deficiency results in increased corticotropin secretion and leads to increased secretion of 11-deoxycortisol, DOC, corticosterone, and androgen by the adrenal gland. Hypertension, a hallmark of 11-hydroxylase deficiency, occurs in approximately two thirds of patients, sometimes as early as the first few years of life. Hypertension is presumably a consequence of the excess DOC secretion, with resultant salt and water retention, and volume expansion. Classically, plasma renin activity levels are low (low renin hypertension); hypokalemia is uncommon. The presence of hypertension does not always correlate with the plasma concentration of DOC and, contrariwise, transient salt loss has been reported in rare infants with apparent classic 11-hydroxylase deficiency.

Excess androgen secretion in utero by the stimulated fetal adrenal gland masculinizes the external genitalia of the female fetus and causes female pseudohermaphroditism. Postnatally, untreated males and females experience progressive virilization and rapid somatic growth and skeletal maturation. Prepubertal gynecomastia can occur in untreated patients and regresses with glucocorticoid therapy.

Putative mild, late-onset, and even cryptic forms of 11-hydroxylase deficiency have been reported. These patients are born with normal genitalia and develop signs and symptoms of androgen excess in childhood, at adolescence, or as adults, but they usually do not manifest hypertension.

Over 31 mutations in the CYP11B1 gene have been reported associated with 11-hydroxylase deficiency. Most are substitutions with a small number of deletions or insertions. There appears to be some clustering in exons 6 to 8 of the gene mutations. In Moroccan Jews, all affected patients are homozygous for the same mutation, Arg448His, consistent with a founder effect. This mutation is also found in other ethnic groups. Another mutation, Arg448Cys, occurs in codon 448, suggesting that this codon is a "hot spot" for mutations. Although CYP11B1 and CYP11B2 are closely linked homologues, both genes are functional, and gene conversions are not a cause of impaired enzyme activity. As with 21-hydroxylase deficiency, the nonsense mutations (i.e., those resulting in less enzymatic activity) would be expected to result in the more severe phenotypic manifestations. However, as in 21-hydroxylase deficiency, heterogeneity in phenotypic manifestations with the same genotype can occur.

Mutations in the CYP11B2 gene impair the conversion of DOC to aldosterone and result in hyponatremia, hyperkalemia, and failure to thrive but do not impair sex differentiation or gonadal function. Two forms exist: so-called corticosterone methyl oxidase I (CMO1) deficiency, a defect in the 18-hydroxylation of corticosterone, and short corticosterone methyl oxidase II (CMO2) deficiency, a defect in the 18-oxidation of 18-hydroxycorticosterone to aldosterone. Neither of these mutations affects cortisol synthesis or results in excess androgen secretion. Glucocorticoid-suppressible hyperaldosteronism and hypertension result from the fusion of the 5' end of the CYP11B1 gene and the 3' end of the CYP11B2 gene; as a consequence of this mutation, synthesis of aldosterone in the zona fasciculata is mediated by corticopin. There is no defect in 11-hydroxylation of deoxycorticisol; hence, virilization does not occur.

Diagnosis.
DOC, and corticosterone levels in plasma or measurement of tetrahydrodemobolates in urine. The diagnosis should be suspected in patients who have levels of these steroids that are at least threefold higher than the 95th percentile for age. The increased secretion of the mineralocorticoid DOC produces salt and water retention and consequently results in the suppression of plasma renin activity and aldosterone. Steroidogenic enzymes for CYP11B1 activity do not differ from normal individuals in the response of 11-deoxycorticosteroids, DOC, or corticosterone to the administration of corticotropin. The disorder can be detected prenatally, and the masculinization of the external genitalia of affected female fetuses can be prevented by prenatal dexamethasone treatment.

The treatment of CYP11B1 deficiency is similar to that of the non-salt-losing form of CYP21 deficiency. Cortisol suppresses corticotropin secretion and, as a consequence, corrects the increased secretion of adrenal androgens and DOC. Replacement therapy usually arrests virilization and alleviates the hypertension. However, in patients with long-standing hypertension, adjunctive therapy with antihypertensive agents may be necessary to control the hypertension. The principles concerning surgical repair of the external genitalia are similar to those relating to CAH caused by 21-hydroxylase deficiency.

### 3-Hydroxysteroid Dehydrogenase-14,15-Dioxygenase Deficiency

3-Hydroxysteroid dehydrogenase type 1 and type 2 (HSD3B1 and HSD3B2) are enzymes that convert cholesterol to androgens and estrogens. Mutations in these genes can lead to a variety of clinical manifestations, including virilization in females and ambiguous genitalia in males. Inheritance is typically autosomal recessive. The diagnosis is made by detecting elevated basal or corticotropin-induced 11-deoxycortisol, DOC, or corticosterone across the age span. The most useful steroid synthesis from 5-17-hydroxypregnenolone is the 17-steroid synthesis from 5-17-hydroxypregnenolone. Aldosterone is synthesized from 11-deoxycortisol, which is converted to aldosterone by the enzyme aldosterone synthase. Aldosterone is a mineralocorticoid that regulates blood pressure and sodium and water balance. The increased secretion of aldosterone results in sodium and water retention, leading to hypertension. The diagnosis should be suspected in patients who have levels of aldosterone that are at least threefold higher than the 95th percentile for age. The increased secretion of aldosterone produces hypertension and is associated with the non-salt-losing phenotype.

Inheritance of HSD3B2 gene is expressed in the placenta, adrenals, testes, and ovaries. The diagnosis should be suspected in patients who have levels of aldosterone that are at least threefold higher than the 95th percentile for age. The increased secretion of aldosterone produces hypertension and is associated with the non-salt-losing phenotype.
stereoidogenesis. In contrast to the rat and the pig, the human and bovine P450{	extsubscript{C17}} enzymes preferentially convert 17-hydroxyprogrenenolon to DHEA (19, 21, 22). In contrast to the rat and the pig, the human and bovine P450{	extsubscript{C17}} enzymes preferentially convert 17-hydroxyprogrenenolon to DHEA (19, 21, 22). They have very poor 17,20-lyase activity (22). The human 17,20-lyase is about 50 times more efficient than the 17,20-lyase. This enzyme also has significant 16-hydroxylase activity. The enzyme is bound to the smooth endoplasmic reticulum where it accepts electrons from a specific flavoprotein, NADPH450 oxidoreductase. The gene CYP17, which encodes the enzyme, is approximately 13 kb long, contains eight exons, is located on chromosome 10 at 10q24.25, and is expressed in the adrenals and gonads, but not in the placenta or in ovarian granulosa cells.

Two types of enzymatic deficiency causing this rare form of 17-hydroxylase and 17,20-lyase deficiency. The combined form is most common (discussed in the section on male pseudohyperplasmosis). However, rare patients with putative isolated 17,20-lyase deficiency have been described. Women with 17,20-lyase deficiency would be expected to have sexual infantilism with lack of adrenarche and elevated gonadotropins at puberty resulting from an inability to synthesize both androgens and estrogens in the adrenals and gonads. In contrast to 17-hydroxylase deficiency, there should be no defect in cortisol synthesis and no mineralocorticoid (DOC) excess and hypertension. Studies on a patient with combined 17-hydroxylase and 17,20-lyase deficiency demonstrated another defect: the absence of 16-ene-synthetase activity, which resulted in impaired conversion of pregnenolone to C{	extsubscript{19}}-steroid sex pheromone precursor and suggests that CYP17 affects 16-ene-synthetase activity in the gonads.

17-Hydroxylase deficiency is a rare autosomal recessive disorder that occurs in approximately 1 in 50,000 individuals. More than 130 cases have been reported. It was initially reported by Biglieri and colleagues in 46,XX females who had low renin hypertension, hypokalemia, and sexual infantilism. Subsequently, this defect was described in 46,XY male infants, children, and adults with pseudohyperplasmosis. A defect in 17-hydroxylase in both the adrenal cortex and gonads results in impaired synthesis of 17-hydroxyprogestrone and 17-hydroxyprogrenenolone and thus of cortisol, androgens, and estrogens. Decreased cortisol synthesis causes increased corticotropin secretion, which results in excessive secretion of 17-deoxysteroids by the adrenal cortex, including the mineralocorticoid DOC, corticosterone, and 18-hydroxycorticosterone. Excess DOC secretion leads to hypertension, hypokalemia, suppression of the renin-angiotensin system, and, secondarily, diminished aldosterone synthesis and secretion in most reported patients. Corticosterone is a weak glucocorticoid, the high plasma concentrations in this disorder prevent the signs and symptoms of cortisol deficiency and modulate the secretion of corticotropin.

The phenotypic manifestations are a consequence of the biochemical defects in adrenal and gonadal steroid biosynthesis. 17-Hydroxylase deficiency is usually recognized at the time of expected puberty in the female because of the presence of hypertension and/or hypokalemia associated with hyperadrenocorticism. Affecte 46,XX females have normal female internal and external genitalia, but the ovaries cannot secrete estrogens at puberty, resulting in sexual infantilism and hypogonadism with elevated plasma FSH and LH levels. In addition, the lack of adrenal and ovarian androgens can result in little or no growth of pubic and axillary hair. In affected 46,XY individuals, the ovaries have a high proportion of atretic follicles and some ovaries contain an increased number of enlarged follicular cysts. (For manifestations in 46,XY males, see Figure 22-63).

In contrast to the rat and the pig, the human and bovine P450{	extsubscript{C17}} enzymes preferentially convert 17-hydroxyprogrenenolon to DHEA (19, 21, 22). In contrast to the rat and the pig, the human and bovine P450{	extsubscript{C17}} enzymes preferentially convert 17-hydroxyprogrenenolon to DHEA (19, 21, 22). They have very poor 17,20-lyase activity (22). The human 17,20-lyase is about 50 times more efficient than the 17,20-lyase. This enzyme also has significant 16-hydroxylase activity. The enzyme is bound to the smooth endoplasmic reticulum where it accepts electrons from a specific flavoprotein, NADPH450 oxidoreductase. The gene CYP17, which encodes the enzyme, is approximately 13 kb long, contains eight exons, is located on chromosome 10 at 10q24.25, and is expressed in the adrenals and gonads, but not in the placenta or in ovarian granulosa cells.

The CYP17 gene has been studied in many patients, and more than 22 different mutations have been identified in its coding region. Complete absence of 17-hydroxylase/17,20-lyase activity has resulted from a variety of mutations, including single base-pair changes resulting in missense mutations, duplications, deletions, and premature translational termination. These mutations, except for one reported in Mennonite kindreds and one in Micronesian patients, appear to be random. The most common mutation is the four base-pair duplication in exon 8, shared by Mennonites and individuals in the Friesland region of the Netherlands, which is attributed to a founder effect originating in Friesland. The mutations in the CYP17 gene that result in complete loss of 17-hydroxylase activity provide strong evidence that only one enzyme has 17-hydroxylase activity.

Patients with partial combined deficiencies of both activities have been analyzed. A 46,XX female with homozgyous deletions of the phenylalanine codon at amino acid 53 or 54 is a deletion of phenylalanine at codon 53 or 54. Subsequently, is a deletion of 518 nucleotides and an insertion of 469 nucleotides. His120 + 7nt is a seven-nucleotide duplication at codon 120 (histidine). +ILe112 is a duplication of isoleucine at codon 112. GC300,301 is a deletion of two nucleotides, guanine and cytosine, at codon 300 and 301. Leu480 + 4nt is a four-nucleotide duplication (cytosine-adenine-thymidine-cytosine) at codon 480. Asp487, Ser487, and Phe489 indicate a deletion of aspartic acid (codon 487), serine (codon 488), and phenylalanine (codon 489). All of these mutations cause 17-hydroxylase deficiency. Missense mutations at codons 347 and 358 (indicated by the box) have been associated with "isolated" 17,20-lyase deficiency.

The mutations in the CYP17 gene that result in complete loss of 17-hydroxylase activity provide strong evidence that only one enzyme has 17-hydroxylase activity.

Patients with partial combined deficiencies of both activities have been analyzed. Patients with partial combined deficiencies of both activities have been analyzed. A 46,XY male with the same mutation had only hypoplasdas and cryptorchidism. Another 46,XY male with ambiguous genitalia was a compound heterozygote with a stop codon (TGA) at amino acid position 239 in exon 4 (a null mutation) on one allele and a missense mutation on the other allele that changes a proline to threonine at amino acid 342 (Pro342Thr) in exon 6. This patient had 20% of 17-hydroxylase activity in transfected cells. Analysis of these mutations suggests that 5% of normal activity in a 46,XX female is necessary to allow estrogen production with normal secondary sexual characteristics and irregular menses, whereas more than 25% of normal activity appears to be necessary to achieve normal virilization of the external genitalia of affected 46,XY males.

Diagnosis.

17-Hydroxylase/17,20-lyase deficiency should be considered in all patients with ambiguous genitalia and hypergonadotropic hypogonadism and in all phenotypic females with or without sexual infantilism, including absent adrenarche, who have hypertension and hypokalemic alkalosis. Elevated levels of 17-deoxy-C{	extsubscript{19}}-steroids such as progesterone, pregnenolone, DOC, and corticosterone in plasma and increased urinary excretion of their metabolites establish the diagnosis. The basal plasma concentrations of DOC, corticosterone, 18-hydroxycorticosterone, and 18-hydroxy-DOC and their response to a corticotropin challenge can be used to discriminate among homogonadotrophic, heterogonadotrophic, and unaffected individuals.

Glucocorticoid therapy for 21-hydroxylase deficiency suppresses DOC and corticosterone secretion. With suppression of the excess circulating mineralocorticoids, the blood pressure and serum potassium level return to normal. At puberty, both affected males and affected females usually require gonadal steroid replacement.

This autosomal recessive form of CAH is associated with severe glucocorticoid and mineralocorticoid deficiency, in which no C{	extsubscript{19}}-C{	extsubscript{21}}, or C{	extsubscript{21}}-steroids are elaborated by the adrenal glands or gonads because of failure to convert cholesterol to pregnenolone. This disorder is the most severe genetic defect in steroidogenesis can result in little or no mineralocorticoid production and absent mineralocorticoid derivatives. Females with this disorder have normal internal and external genital development. Clinical manifestations of adrenal insufficiency, including hyponatremia, hypokalemia, acidosis, dehydrad, and hypoglycemia, usually become apparent in the first few weeks of life, but for survival for months without therapy has been described. Hypogonadism is common, and respiratory distress occurs in about one fourth of neonates. On ultrasonography, CT, or MRI, markedly enlarged, lipid-laden adrenals displace the kidneys downward.

Most of the more than 80 patients with SIAD deficiency are of Japanese and Korean origin; it is seen in 50% of patients with SIAD deficiency in prevalence in Japan and Korea. There is an unexpected 3:1 male/female sex ratio. Many affected individuals...
die in infancy (approximately one third survive with replacement therapy); we have cared for one patient for more than 30 years. In contrast to the severe fetal and postnatal testosterone deficiency in affected 46,XY individuals, surviving affected 46,XX females can enter puberty and menstruate; they later develop hirsutism and may develop gonadal tumors. The prolonged survival described in a few patients is the result of adrenal insufficiency and the puberty and menses described in females have been perplexing. As proposed by Bose, two separate events seem responsible for these phenomena: The first event is the loss of steroid hormone synthesis, which is dependent on StAR in steroid-producing cells. The second event is the accumulation of cholesterol, which cannot be converted to pregnenolone by the cell. Eventually this accumulation engorges the cell and results in disruption of the structural and functional integrity of the cell. It is assumed that the functional activity of the cell in question mediates the time course to functional and structural disruption. Hence, postnatal survival for a period of time may reflect the relatively low level of activity of the definitive adrenal glands prenatally. Postnatally, the zona glomerulosa and fasciculata can make a limited amount of steroids independent of StAR until they become engorged and dysfunctional. The ovaries, in contrast to the testes, remain relatively quiescent through fetal life and childhood. At puberty, estrogen synthesis independent of StAR can occur, leading to feminization and menses. Progressive gonadal failure due to cholesterol engorgement of steroidogenic cells then results (see Color Plates).

In support of the two-hit hypothesis, no surviving XY patient has had evidence of testicular function at the expected age of puberty. All patients are markedly pigmented. The SIAR knockout mouse has a similar phenotype to patients with congenital lipoid adrenal hyperplasia, consistent with the two-hit hypothesis model.

The transfer of cholesterol from the outer to the inner mitochondrial membrane is the rate-limiting step in acute or rapid steroid synthesis. A 30-kd mitochondrial protein in adrenal cells rapidly increases in response to corticotropin stimulation and is inhibited by cycloheximide; it is also present in the gonads.

The CDNA for this factor has been cloned, and the protein for this gene was named the steroidogenic acute regulatory (StAR) protein. As with other steroid hormone hydroxylases, the SIAR gene is transcriptionally regulated by SF1. The stimulation of SIAR by corticotropin and by LH is mediated through a cyclic adenosine monophosphate (AMP)/protein kinase A-dependent pathway and involves the phosphorylation of the SIAR protein.

The stimulation of cholesterol transport by SIAR from the outer to the inner mitochondrial membrane, the site of the cholesterol side-chain cleavage complex, apparently does not require the import of SIAR across the mitochondrial membrane. Human SIAR is encoded by a gene on chromosome 8p11.2 and is expressed in the adrenal gland and gonad, but not in the placenta or the cerebral nervous system. Fifty-seven patients with congenital lipoid adrenal hyperplasia (studied from 10 countries) had mutations in the gene encoding SIAR (Fig. 22-65). In 80% of affected alleles from affected Japanese and Korean individuals, a Gln285Stop mutation was detected, whereas an Arg182Leu mutation was present in 78% of alleles from affected Arabs. Study of congenital lipoid adrenal hyperplasia provided the decisive evidence of the critical role of SIAR in steroid hormone biosynthesis in humans. Unlike CYP11A1, the gene encoding the SIAR protein is not expressed in the human placenta, and mutations in SIAR do not impair progesterone synthesis by the placenta.

In patients with SIAR deficiency, little or no C19→C21 steroids are detectable in plasma or urine, even after corticotropin stimulation. In 46,XX females the differential diagnosis includes congenital adrenal hyperplasia. Demonstration of greatly enlarged adrenals in StAR deficiency by imaging techniques readily differentiates these two entities, even when affected with a frameshift mutation affecting small adrenal glands. Imaging. Affected males are raised as females, and their functionless testes are reduced to make the risk of malignant transformation and for cosmetic reasons. Therapy requires replacement with glucocorticoids and mineralocorticoids and the addition of estrogen when gonadal failure ensues.

In obligate heterozygotes with a SIAR mutation, unlike heterozygotes with other forms of CAH, steroid responses to corticotropin are normal. Prenatal diagnosis of SIAR deficiency was successfully demonstrated in a family in which two previously affected children. Amniotic fluid levels of progesterone and pregnenolone were 30% and 50% of normal, respectively, but the concentrations of steroids such as 17-hydroxypregesterone, cortisol, DHEA, androstenedione, and estradiol were low or undetectable; androstenedione and subsequent failure to synthesize fetal adrenal precursors for transformation to estrogens by the placenta result in low maternal plasma and urinary estradiol values.

The clinical manifestations of each form of CAH are summarized in Table 22-23.
TABLE 22-23 -- Clinical Manifestations of Various Types of Congenital Adrenal Hyperplasia

<table>
<thead>
<tr>
<th>Gene</th>
<th>Enzymatic defect</th>
<th>Chromosomal sex</th>
<th>External genitalia</th>
<th>Postnatal virilization</th>
<th>Addisonian crises</th>
<th>Hypertension</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAR</td>
<td>(no defect)</td>
<td>XX XY</td>
<td>Female</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>HSD3B2</td>
<td>3-HSD 2</td>
<td>XX XY</td>
<td>Female</td>
<td>-</td>
<td>±</td>
<td>-</td>
</tr>
<tr>
<td>CYP17</td>
<td>P-450&lt;sub&gt;17&lt;/sub&gt;</td>
<td>XX XY</td>
<td>Female or ambiguous</td>
<td>-</td>
<td>±</td>
<td>-</td>
</tr>
<tr>
<td>CYP11B1</td>
<td>(17-Hydroxylase)</td>
<td>XX XY XX XY</td>
<td>Ambiguous</td>
<td>-</td>
<td>± in 80%</td>
<td>-</td>
</tr>
<tr>
<td>CYP 21</td>
<td>(21-Hydroxylase)</td>
<td>XX XY XY XY</td>
<td>Male or male</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 22-24 -- Clinical Features of CYP19 (P450<sub>arom</sub>) Deficiency in the Female

<table>
<thead>
<tr>
<th>Karyotype:</th>
<th>46,XX</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inheritance:</td>
<td>Autosomal recessive</td>
</tr>
<tr>
<td>Maternal history:</td>
<td>Virilization of mother during pregnancy</td>
</tr>
<tr>
<td>Genitalia:</td>
<td>Ambiguous or female with clitoromegaly</td>
</tr>
<tr>
<td>Wolffian duct derivatives:</td>
<td>Absent</td>
</tr>
<tr>
<td>Müllerian duct derivatives:</td>
<td>Present</td>
</tr>
</tbody>
</table>

AFFECTED FEMALES ARE BORN WITH CLITOROMEgALy, VARYING DEGREES OF POSTERIOR FUSION, SRCRALIZATION OF THE LABIOSCROTAL FOLDS, AND, IN SOME INFANTS WITH A UROGENITAL SINUS, A SINGLE
perineal orifice (Table 22-24). Müllerian structures are normal. 

During infancy, basal and LHRH-induced FSH and LH are elevated. The history of the ovary in infancy is normal, but under increased FSH stimulation in the absence of ovarian CYP19, multiple enlarged follicular cysts develop. All pubertal affected females have hypergonadotropin hypergonadism, fail to develop female secondary sexual characteristics, and exhibit progressive virilization. 

Plasma androstenedione and testosterone are elevated, and estrone and estradiol levels are low or unmeasurable. The ovaries enlarge and develop multiple cysts at puberty, in one affected female, polycystic ovaries were detected in infancy. The hypergonadotropism and the multiple ovarian cysts respond to estrogen replacement therapy.

The affected 2 adult men had normal sex differentiation and pubertal maturation with macro-orchidism (in one male) and elevated concentrations of FSH, LH, and testosterone in plasma. They also had osteoporosis, hyperinsulinemia, and abnormal plasma lipids, similar to the findings in a tall man with a null mutation in the estrogen receptor. These observations suggest that estrogens as well as testosterone and inhibin play a role in the regulation of gonadotropin secretion in males and females and that estrogen deficiency in males can be associated with insulin resistance and hyperinsulinemia and an abnormal plasma lipid profile.

The finding of apparently normal psychosexual development in the three aromatase-deficient adolescent or adult patients and in the man with an estrogen receptor defect suggests that estrogen does not play a critical role in sex differentiation of the human brain, as has been reported in nonprimate mammals. The detection of severe defects in critical regions of the gene encoding CYP19 that lead to generalized aromatase deficiency is strong evidence that survival of the conceptus can occur in the absence of estrogen synthesis by the implanting blastocyst, the fetus, and the fetal compartment of the placenta.

Analysis of the CYP19 gene in nine affected individuals has revealed 10 different mutations (Fig. 22-67) (Table 22-25). The patient of Shouz and colleagues was homozygous for a mutation (GT GC) in the consensus 5' splice acceptor sequence in the gene. This mutation resulted in a CYP19 protein with a 29-amino acid insert and less than 0.3% of normal enzyme activity. A second patient was a compound heterozygote with two missense mutations. Assay of the expressed mutated proteins showed that one allele had 1.1% of the activity of the wild-type CYP19, whereas the other had no activity. The male and female siblings reported by Morishima and associates resulted in a single base change that resulted in an amino acid substitution. Expression of this mutant cDNA showed that it had 0.2% of the wild-type aromatase activity. The patient of Mullis and co-workers had a cytosine deletion in one allele (codon 408, proline) that corresponds to the point mutation (GT GC) in the consensus 5' splice acceptor sequence in the gene. This mutation resulted in a CYP19 protein with a 29-amino acid insert and less than 0.3% of normal enzyme activity. A second patient was a compound heterozygote with two missense mutations. Assay of the expressed mutated proteins showed that one allele had 1.1% of the activity of the wild-type CYP19, whereas the other had no activity. The male and female siblings reported by Morishima and associates resulted in a single base change that resulted in an amino acid substitution. Expression of this mutant cDNA showed that it had 0.2% of the wild-type aromatase activity. The patient of Mullis and co-workers had a cytosine deletion in one allele (codon 408, proline) that corresponds to the

### TABLE 22-25 — Mutations of CYP19: Relation between Aromatase Activity in In Vitro Expression Systems and Virilization of the Mother

<table>
<thead>
<tr>
<th>Study</th>
<th>Sex of Affected Child</th>
<th>Genetic Defect</th>
<th>Activity (% of Normal)</th>
<th>Mother</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kanazawa, 1992</td>
<td>F</td>
<td>Splice junction defect (exon VI)</td>
<td>0.3</td>
<td>Virilized</td>
</tr>
<tr>
<td>San Francisco, 1993</td>
<td>F</td>
<td>Arg435Cys</td>
<td>1.1</td>
<td>Not virilized</td>
</tr>
<tr>
<td>New York, 1995 (sibs)</td>
<td>F</td>
<td>Cys437Ty (exon X)</td>
<td>0</td>
<td>Virilized</td>
</tr>
<tr>
<td>Lyon, 1996</td>
<td>F</td>
<td>Arg457X (exon X)</td>
<td>0</td>
<td>Virilized</td>
</tr>
<tr>
<td>Bern, 1997</td>
<td>F</td>
<td>Exon III (splice site)</td>
<td>0</td>
<td>Virilized</td>
</tr>
<tr>
<td>Modena, 1997</td>
<td>M</td>
<td>Arg385Gln (exon IX)</td>
<td>0.4</td>
<td>?</td>
</tr>
<tr>
<td>Bonn, 1998</td>
<td>M</td>
<td>Val370Met (exon IX)</td>
<td>0</td>
<td>Virilized</td>
</tr>
<tr>
<td>Bern, 1999</td>
<td>M</td>
<td>C base deletion frameshift stop codon (exon V)</td>
<td>0</td>
<td>Virilized</td>
</tr>
</tbody>
</table>


**ND**, not determined.

*Presumed lack of activity because of stop codon.

Glucocorticoid Receptor Gene Mutation

Autosomal dominant forms of glucocorticoid resistance, a rare disorder, are a consequence of a heterozygous mutation in the gene encoding the glucocorticoid receptor, usually in the ligand-binding domain of this transactivation factor. End-organ insensitivity to glucocorticoids in the heterozygote is heterogeneous in its manifestations but characteristically there is excess ACTH secretion, with elevated cortisol levels without Cushing syndrome and usually with few symptoms. Nevertheless, the hypersecretion of ACTH increases adrenal steroidogenesis and can lead not only to increased cortisol secretion but also to increased...
mineralocorticoids (as a consequence, hypertension and hypokalemia may occur) and androgen precursors. The latter steroids can produce acne, hirsutism, adolescent hyperandrogenism, menstrual abnormalities in the female, and sexual precocity in the male.

A Brazilian girl has been reported who had female pseudohemophrditism owing to a novel homozygous inactivating mutation in the glucocorticoid receptor gene. Born of consanguineous clinically unaffected parents, she had a large clitoris, posterior labioscrotal fusion, and a urogenital sinus. When she was 9 years of age, studies showed low renin hypertension and hypokalemia with increased plasma DOC and cortisosterone, high concentrations of plasma ACTH and cortisol, and impaired suppression of plasma cortisol during a dexamethasone test. Consistent with her advanced bone age and progressive virilization, the concentrations of plasma testosterone, androstenedione, and 17-Hydroxyprogesterone were elevated. Molecular analysis identified a homozygous Val571Ala mutation involving the ligand-binding region. Mutant receptors had less than one sixth the binding affinity and one tenth to one fifth of the trans-activation activity of the wild-type receptor. After repair of the hypokalemia, high-dose dexamethasone therapy (6 mg/day) led to impressive clinical improvement.

**Maternal Androgens and Pregestagens**

Masculinization of the external genitalia of female infants has been observed after maternal ingestion of testosterone or synthetic progestational agents during the first trimester of pregnancy. If the exposure occurs after the 12th week of gestation, the labioscrotal folds do not fuse, although the clitoris may enlarge. Severe masculinization of the external genitalia of a female fetus may be caused, for example, by methylnorchisosterone in dosages as low as 3 mg daily, even though androgenic effects are not noticeable in the mother. Placenta aromatase may be unable to efficiently aromatize synthetic androgens such as methyltestosterone.

Because progesterone itself is only slightly active when administered orally, various synthetic derivatives that may be taken orally were prescribed in the past for women with habitual or threatened abortion. Most of these progestagens are 19-nortestosterone derivatives; they are intrinsically androgenic to some degree and can cause virilization of female fetuses in experimental animals. Principal among the offenders have been norethindrone and ethisterone and, less commonly, norethynodrel and the C17-steroid medroxyprogesterone acetate. Ishizuka and co-workers reported some degree of masculinization of the external genitalia in 2.75% of female infants whose mothers received synthetic progestagens of various types during pregnancy. This consequence of synthetic progestagen administration to the pregnant female is dose and time dependent.

Danazol, the 2,3-disoxazole derivative of 17-ethinyltestosterone, a progestagen, is used for the treatment of endometriosis. Danazol crosses the placenta and can cause virilization of the external genitalia of the fetus in a manner similar to other androgenic compounds. Several instances of female pseudohemaphroditism are believed to be the consequence of maternal ingestion of danazol. The mechanism of masculinization is unknown but may be related to inhibition of 3-HSD by stilbestrol or its metabolites.

Masculinization of the female fetus occurs on occasion if the mother has a virilizing ovarian tumor (usually arrhenoblastoma or Krukenberg’s tumor) or adenral tumor, a virilizing form of CAH, or virilization of some other cause during pregnancy. An increasing number of women with CAH are becoming pregnant as a result of improved treatment and it is reassuring to note that female offspring of such pregnancies are not virilized (with a single exception), despite higher than usual maternal testosterone levels. Luteoma of pregnancy, an ovarian pseudotumor composed of hyperplastic luteinized thecal cells that regress after delivery, has been associated with masculinization of the external genitalia of female infants, especially in the presence of maternal virilization.

**Ovarian lutein cysts in pregnancy (hyperreactio luteinalis), considered by some to be a cystic form of luteoma, are less frequently associated with maternal virilization and only rarely with fetal masculinization.** Placental aromatization of androgens such as testosterone and androstenedione protects the mother and the female fetus from virilization unless the placental CYP19 (P450arom ) activity is insufficient for the androgen steroid load or unless the synthetic androgen is not a substrate for CYP19.

Some of the rare cases of female pseudohemaphroditism of undetermined origin may have resulted from a luteoma of pregnancy that regressed spontaneously after delivery or an undiagnosed placental aromatase deficiency. In these patients a history of maternal ingestion of androgenic steroids is lacking and the postpartum course of the mother is inconsistent with a virilizing neoplasm, but the clinical features are most compatible with fetal exposure to androgens. The absence of virilism in the mother does not exclude a maternal source of androgen in these children, because the amounts of androgen required to masculinize the external genitalia of a female fetus may be less than those required to cause overt manifestations in the mother.

Female pseudohemaphroditism caused by the transfer of androgenic steroids from the mother to the fetus is the most easily treated of all types of androgens in terms of successful development. No hormone therapy is necessary, postnatal virilism does not occur, and female secondary sexual characteristics can be expected to emerge at the usual age of adolescence. Surgical correction of the external genitalia in infancy or childhood is rarely indicated.

**Malformations of the Intestine and the Urinary Tract (NonAndrogen-Induced Female Pseudohemaphroditism)**

Genital abnormalities are frequently associated with imperforate anus, renal agenesis or dysplasia, and other congenital malformations of the lower intestine and urinary tract. Carpenter and Potter reviewed the findings in such infants and suggested the term "nonspecific female pseudohemaphroditism." Some of these anomalies are incompatible with life. Renal failure, often accompanied by pyelonephritis, is common and may confuse the clinical picture with that of adrenal insufficiency. In contrast to other forms of female pseudohemaphroditism, the female müllerian derivatives may also be malformed. The findings in these patients may be bizarre; persistence of a primitive cloaca, imperforate anus, and fistulae are not infrequent. The pathogenesis of these anomalies is different from that of other types of androgenic development and should be considered in the context of other forms of developmental field defects. Familial occurrence of nonadrenal female pseudohemaphroditism with multiple anomalies has been reported. Citrormegalmy may sometimes be the presenting feature of neurofibromatosis and may be more common in this neurologic disorder than is realized. The clitoris is richly innervated, and biopsies have confirmed the presence of neuromas when citrormegalmy occurs with neurofibromatosis. Surgical excision is not curative because of the tendency for recurrence of the neuromas.
Male Pseudohermaphroditism

Male pseudohermaphroditism is a heterogeneous condition in which the gonads are exclusively testes but the genital ducts and/or external genitalia are incompletely masculinized. The clinical spectrum varies from individuals with female external genitalia to those with mild impairment of masculinization of the external genitalia, as represented by hypospadias, cryptorchidism, and minimal ambiguity of the external genitalia.

With the advances in the knowledge of pathogenesis, systems of nomenclature based on eponyms and phenotype have become less important. There are at least six major etiologic categories of male pseudohermaphroditism, with many subtypes, all of which are associated with incomplete masculinization of the fetal genital tract and/or incomplete regression of the müllerian ducts.

In this section, forms of male pseudohermaphroditism in 46,XY individuals with relatively normal embryonic differentiation of the testes are discussed. In such patients, defective male development must be ascribed to a more specific failure of the fetal testes to overcome the inherent tendency toward feminization of the somatic sex structures. This failure may stem either from a secretory failure of the testes during the critical period of sex differentiation or from a failure of target tissues to respond normally to androgen stimulation or to AMH. Table 22-26 reflects an attempt to classify the many forms of male pseudohermaphroditism on the basis of cause, insofar as that is known.

The ability of the testes to virilize at adolescence is in many ways a recapitulation of their capacity to masculinize the external genitalia in utero. The greater the development of the phallus in an infant, the greater likelihood that male secondary sexual characteristics will emerge at the time of expected puberty. Individuals with ambiguous genitalia may remain eunuchoid, exhibit mild virilism, or develop breast enlargement and other female secondary sexual characteristics. Those with an external female phenotype usually either feminize or remain sexually infantile. These are only approximate guides, however, and the development of male sexual characteristics at adolescence may occur, especially in patients with partial androgen resistance, 17-hydroxysteroid dehydrogenase-3 deficiency, or 5-reductase-2 deficiency.

Male pseudohermaphroditism can result from (1) testicular unresponsiveness to hCG and LH and consequent Leydig cell aplasia or hypoplasia; (2) a specific enzyme defect in testosterone biosynthesis; (3) familial end-organ resistance to androgen caused by abnormalities in the cytosolic receptor for testosterone and DHT or by an enzyme defect in the intracellular metabolism of testosterone; (4) aberrations in testicular organogenesis (dysgenetic male pseudohermaphroditism); (5) defective synthesis, secretion, or response to AMH; (6) administration of progestagens during pregnancy; and (7) putative environmental exposures. Apart from dysgenetic male pseudohermaphroditism and the persistent müllerian duct syndrome, all other forms of male pseudohermaphroditism are characterized by the absence of müllerian duct derivatives. Except for some variants of dysgenetic male pseudohermaphroditism and the maternal ingestion of progestagens, virtually all forms of male pseudohermaphroditism are familial and characterized by genetic heterogeneity. No doubt many subtypes will be defined and characterized by molecular, genetic, and biochemical techniques. Although dysgenetic male pseudohermaphroditism and the group of disorders associated with defective organogenesis of the testes has already been discussed, it is included under male pseudohermaphroditism because this category of intersexuality must be considered by the clinician in the differential diagnosis of male pseudohermaphroditism. In Table 22-4 and Figure 22-15 genes implicated in sex determination (gonadogenesis) are shown; Table 22-19 lists genes involved in sex differentiation.

Defects in Testosterone Biosynthesis and Metabolism

Testicular Unresponsiveness to hCG and LH, LH/hCG Resistance (Leydig Cell Agenesis or Hypoplasia)

The production of testosterone by fetal Leydig cells is critical to male sexual differentiation of the Wolffian ducts and the
Increased gonadotropins postpubertally, decreased testosterone levels with decreased or absent response to hCG stimulation, decreased or absent hypoplasia and hypergonadotropic hypogonadism, and a sexually mature 46,XX female with elevated plasma LH levels and amenorrhea.

Phenotypically, the external genitalia vary, from those of a normal-appearing female to those of a male with micro penis and hypoplastic external genitalia. Male pseudohermaphroditism results from an LH/LH-receptor defect in fetal testes and includes a complete absence of female differentiation. The LH-receptor is responsible for the biosynthesis of testosterone, and it is also necessary for male sexual differentiation.

Phenotypic features of Leydig cell unresponsiveness to hCG/LH (Leydig cell aplasia or hypoplasia) include:

- Aplasia/hypoplasia of the testis (recognizable in the first trimester of pregnancy)
- Aplasia/hypoplasia of the seminiferous tubules
- Absent or decreased number of Leydig cells
- Absent or diminished binding of labeled hCG and LH to Leydig cells
- Absent or diminished testicular production of testosterone
- Absent or diminished hCG/LH response to LHRH-evoked responses
- Absent or diminished hCG/LH response to hCG/LH resistance

Male pseudohermaphroditism may result from an LH-receptor defect, which is responsible for the biosynthesis of testosterone in male testes and is also necessary for male sexual differentiation. The LH-receptor is responsible for the biosynthesis of testosterone, and it is also necessary for male sexual differentiation.

The LH-receptor is responsible for the biosynthesis of testosterone, and it is also necessary for male sexual differentiation. The LH-receptor is responsible for the biosynthesis of testosterone, and it is also necessary for male sexual differentiation.
Homozgyous mutation (Ser616Tyr) in the seventh trans-membrane domain of the LH receptor gene. The mutant receptor did not bind hCG. Homozygous mutations in eight 46,XY males analyzed resulted in fetal and postnatal testosterone deficiency and a spectrum in appearance of the external genitalia extending from female external genitalia to a micro penis and in hypogenital dystrophic hypogonadism. Wu and associates described a compound heterozygous loss of function mutation with a nonsense mutation (Cys545X) in exon 11 of the LH receptor in one allele and a 3 base-pair insertion in exon 1 on the other allele in two 46,XY females with Leydig cell hypoplasia and LH/hCG resistance. In 10 patients with clinical features of Leydig cell hypoplasia, only 2 patients had mutations in the coding region of the LH/hCG receptor, which suggests that mutations may occur in other regions of the gene or a mutation in other Leydig cell specific genes may produce the same phenotype. For example, in the mouse Desert hedgehog (DHH) a signaling protein and its receptor Patched 1 (Pch1) are critical factors in the differentiation of fetal Leydig cells in the mouse but do not affect cell migration from the mesonephros nor the contribution of interstitial cells from the coelomic epitheum. An XY DH-null mouse bred on a mixed strain background exhibited severely impaired fetal Leydig cell differentiation and female external genitalia. A homozygous mutation in DHH was reported in a 46,XY male pseudohermaphrodite with partial gonadal dysgenesis and polyneuropathy. These observations support the genetic heterogeneity of the syndrome and the importance of searching for genes involved in Leydig cell differentiation and function. Null mutations of the LH receptor (LH-R) would not be expected to result in feminization of 46,XY females because female secondary sexual characteristics are maintained in the absence of testicular androgens. In contrast, females with complete androgen insensitivity (Xg7,8) and with mutations of the androgen receptor have normal expression of female secondary sexual characteristics.

In patients with testicular unresponsiveness to hCG/LH, fetal testosterone deficiency impairs masculinization of the external genitalia, but müllerian duct regression is complete because the secretion of AMH by the fetal Sertoli cells is intact. Of interest is the paradoxical finding of Wolffian derivatives, which are testosterone dependent, in some patients with no or minimal masculinization of the external genitalia (only posterior labial fusion). One explanation, supported by the variation in masculinization of the external genitalia, is that the defect in the hCG/LH receptor is of variable severity. A second, more likely possibility is that during the early fetal period, sufficient testosterone may have been secreted locally and autonomously independently of circulating hCG to induce male duct development, but the concentration of testosterone in the fetal circulation was too low to evoke normal male differentiation of the external genitalia and urogenital sinus. hCG is necessary to sustain Leydig cell differentiation and growth and testosterone secretion by the fetal testes, at least by about the 10th week of gestation, but, as discussed previously, it may not be essential for initiation of these functions at week 8. Variation in the magnitude of hCG/LH resistance of the unresponsive embryonic and fetal Leydig cells would result in variable degrees of fetal testosterone deficiency and thus a variable degree of failure to develop normal male external genitalia. Therapy depends on the age at diagnosis and the degree of virilization. In the severe form of testicular unresponsiveness to hCG/LH with female external genitalia, sex assignment is usually female. The gonads are removed, and estrogen replacement therapy is instituted at the time of expected puberty. In the less extreme forms with predominantly male external genitalia, testosterone therapy augments phallic development and virilizes the patient at puberty. In the human male fetus, deficient fetal pituitary gonadotropin secretion associated with anencephaly, hypothalmic hypopituitarism, and isolated gonadotropin deficiency (including Kallmann syndrome) is not associated with ambiguous external genitalia (with one possible exception of the male/female ratio in this disorder appears to be 3:1. Enzyme defects common to adrenal and testicular biosynthesis include 3-HSD, CYP17, cholesterol side-chain cleavage, as well as abnormalities in the intracellular steroid transport protein, StAR. These have already been reviewed in the section on CAH. A fourth defect involves the biosynthesis of cholesterol (7-cholesterol reductase deficiency) and results in multisystem metabolic malformation syndrome.

Enzyme defects affecting both adrenal and testicular biosynthesis include CAH and StAR Deficiency (Congenital Lipoid Adrenal Hyperplasia). A single step involving enzymes, and the third, StAR, a protein that affects both glucocorticoid and gonadal steroid biosynthesis; these errors in steroid biosynthesis are discussed, in part, in the section on CAH. A fourth defect involves the biosynthesis of cholesterol (7-cholesterol reductase deficiency) and results in multisystem metabolic malformation syndrome.

Enzyme Defects of Testosterone Biosynthesis Affecting Both Adrenal Steroid and Testosterone Biosynthesis (Variants of Congenital Adrenal Hyperplasia)

Enzyme defects common to adrenal and testicular biosynthesis include 3-HSD, CYP17, cholesterol side-chain cleavage, as well as abnormalities in the intracellular steroid transport protein, StAR. These have already been reviewed in the section on CAH. Some additional features are described for the 17,20-lyase variant of CYP17 deficiency in view of the primary effect on testicular steroid biosynthesis and, hence, male pseudohyperpladioid.

Defects in testosterone biosynthesis (Fig. 22-70) have been described, one at each of the enzymatic steps required for the conversion of cholesterol to testosterone[916] (Fig. 22-71). Three of the defects (3-HSD, CYP17, CYP11A1 [P450,]) involve enzymes, and the third, StAR, a protein that affects both glucocorticoid and gonadal steroid biosynthesis; these errors in steroid biosynthesis are discussed, in part, in the section on CAH. A fourth defect involves the biosynthesis of cholesterol (7-cholesterol reductase deficiency) and results in multisystem metabolic malformation syndrome.

StAR Deficiency (Congenital Lipoid Adrenal Hyperplasia)

Infants with this defect (Table 22-28; see previous discussion in the section on CAH) present with severe adrenal insufficiency and accumulation of lipid in the cells of both the adrenal cortex and the gonads. The second most common form of CAH in Japan (the GnRH6X mutation was found in over 80% of Japanese patients studied) and the male/female ratio in the United States and Europe is 1:1. Affected females have female external genitalia with a blind vaginal pouch and hypoplastic male genital ducts but no uterus or fallopian tubes; the genitalia of affected females are normal. In males, the testes may be abdominal, inguinal, or in the labia. All reported patients are diffusely pigmented; glucocorticoid and mineralocorticoid insufficiency is severe; and adrenal crises in infancy can lead to death if untreated. Three male pseudohermaphrodites

---

Figure 22-70 Diagram of male sex determination and differentiation showing the consequences of an enzymatic block in biosynthesis of testosterone that results in male pseudohermaphroditism. Solid bar indicates defect (see legend for Figure 22-43).

Survived the perinatal period without therapy and presented at 6 weeks, 12 weeks, and 10 months of age; (see discussion of StAR and its mutations in the section on female pseudohermaphroditism). The patient reported by Haufla and co-workers (and in the previous editions of this chapter) is more than 35 years old and well maintained on glucocorticoid and mineralocorticoid replacement therapy. Sexual hair is absent unless small doses of testosterone are given; female secondary sexual characteristics are induced by estrogen replacement. As described in the section on CAH, no secondary sexual characteristics (either male or female) develop at the age of puberty in affected males, in contrast to females with this disorder. Wu and Pedigree analysis of families and DNA analysis of affected patients and parents indicate autosomal recessive transmission. The usual ratio of affected patients is yet to be explained and may represent, at least in part, ascertainment bias. The molecular genetics of StAR deficiency are illustrated in Figure 22-45 and 22-46.

The diagnosis of StAR deficiency should be suspected in patients with male pseudohermaphroditism, including all phenotypic female infants with evidence of adrenal insufficiency. The diagnosis can be confirmed by documentation of low or absent mineralocorticoids, glucocorticoids, and gonadal steroids and their metabolites in plasma and urine and an absent steroid response to corticotropin and hCG administration. The adrenals are large and lipid laden and displace the kidneys caudal on ultrasonography, CT, or MRI; the enlargement can
Persist for months after glucocorticoid therapy. Therapy requires replacement doses of glucocorticoid and mineralocorticoid from the time of diagnosis. All affected 46,XY males have been reared as females. Estrogen replacement therapy to induce female sexual characteristics at puberty and low-dose testosterone treatment to elicit development of sexual hair are indicated; prophylactic orchidectomy is appropriate.


P450c17 is the enzyme that converts cholesterol to pregnenolone, the first step in steroidogenesis. As noted in the section on CAH and female pseudohemaphroditism, null mutations in both alleles of this gene would lead to failure of placental production of progesterone, which in the human (unlike, for example, the rabbit) would be incompatible with continued gestation beyond the 8th to 10th week. Tajima and his associates described a hyperpigmented 4-year-old XY male pseudohemaphrodit with mild virilization of the external genitalia as evidenced by citoromegaly but no labial fusion, with a blind vaginal pouch, who had had bilateral inguinal testes removed. It was not until age 4 years that she experienced life-threatening adrenal insufficiency which with a low concentration of plasma cortisol, a high plasma ACTH and plasma renin activity, but normal concentrations of serum sodium and potassium.

**TABLE 22-28**  -- Clinical Features of SIAR Deficiency in 46,XY Males

| Karyotype: | 46,XY |
| Genitalia: | Female |
| Wolffian duct derivatives: | Absent hypoplastic |
| Müllerian duct derivatives: | Absent |
| Gonads: | Testes |
| Habitus: | Severe adrenal insufficiency in infancy, no virilization at puberty |
| Hormone profile: | Decreased or absent glucocorticoids, mineralocorticoids, and gonadal steroids in plasma and urine; increased plasma LH and FSH; increased plasma renin |

LH, luteinizing hormone; FSH, follicle-stimulating hormone.

*46,XY male with heterozygous mutation of P450c17. Genitalia: Clitoromegaly separate vaginal and urethral openings, no labial fusion, no müllerian ducts, childhood onset of adrenal insufficiency.

**TABLE 22-29**  -- Clinical Features of 3-HydroxySteroid Dehydrogenase II Deficiency in 46,XY Males

| Karyotype: | 46,XY |
| Inheritance: | Autosomal recessive; mutations in 3HSD2 gene |
| Genitalia: | Ambiguous; hypospadiac male |
| Wolffian duct derivatives: | Normal |
| Müllerian duct derivatives: | Absent |
| Gonads: | Testes |
| Habitus: | Severe adrenal insufficiency in infancy; poor virilization at puberty with gynecomastia. Mild form: no mineralocorticoid deficiency, premature adrenarche mild virilization |

This rare autosomal recessive disorder is a consequence of mutations in the CYP11A1 genean in-frame insertion of Gly and Asp between Asp 271 and Val 272was identified that lacked P450c17 enzymatic activity. The suggestion that the P450 haplo-insufficiency eventually led to a late-onset type of congenital lipid hyperplasia is consistent with the two-stage model of pathogenesis proposed for SIAR deficiency. The discrepancy between the compromised function of the fetal testes and the resultant male pseudohemaphroditism and the relatively late onset of symptomatic adrenal insufficiency is incompletely understood. In contrast to the absent Müllerian duct derivatives in this patient, the male pseudohemaphrodit with adrenal insufficiency caused by haploinsufficiency of the SFT gene due to Gly35Glu mutation had a uterus as did the male pseudohemaphrodit with a homozygous mutation (arg92Gin).

3-HydroxySteroid Dehydrogenase II (3HSD2) Deficiency

This rare autosomal recessive disorder is a consequence of mutations in the HSD3B2 gene encoding the 3-HSD2, isomerase type 2 isozyme, which is expressed mainly in the adrenals and gonads. This enzyme catalyzes a crucial step in the biosynthesis of all steroid hormones, the conversion of 4,5-steroids. The type 1 isozyme is expressed predominantly in the placenta and in peripheral tissues (e.g., skin, breast), has 93% homology in structure with the type II isozyme, is about five times as active, and is closely linked to the type II isozyme (both are encoded by genes on chromosome 1p13). The type 1 isozyme is not associated with CAH, and mutations in the coding region of type I are probably lethal because they prevent or compromise progesterone synthesis by the placenta and, hence, survival of the fetus.

Males with 3-HSD2 deficiency have CAH associated with male pseudohemaphroditism with or without adrenal insufficiency (salt loss) (Table 22-29). The external genitalia in males are usually ambiguous, with a small phalic structure, second-or third-degree hypospadias, partial fusion of the labia, a urogenital sinus, and a blind vaginal pouch; rarely, the external genitalia are female in appearance. Wolffian duct differentiation is normal. The testes are usually in the scrotum or lower inguinal region, and, as with other blocks in testosterone biosynthesis and defects in androgen action, Müllerian structures are absent.
Newborns with severe deficiencies of the 3-HSD II (e.g., null mutations) have severe CAH and exhibit signs or symptoms of both glucocorticoid and mineralocorticoid deficiency after the first 5 to 7 days of life. In those infants with severe deficiency of the enzyme, adrenal insufficiency may result in death if the diagnosis is missed and therapy is delayed.\textsuperscript{1093} 46,XY males with partial deficiency of 3-HSD II are not salt losers; the mutant gene in these cases encodes an enzyme with 2% to 10% of residual 3-HSD II enzymatic activity in the adrenals and gonads, as assessed in intact transfected cells.\textsuperscript{1094,1095} This amount of enzymatic activity is apparently sufficient to prevent salt wasting and adequate for aldosterone synthesis but not enough to provide sufficient fetal testosterone to virilize the external genitalia.\textsuperscript{1096}

Gynecomastia can occur at puberty in both affected males and affected females.\textsuperscript{1097,1098} Presumably by peripheral conversions of $^{5}$-C$_{19}$-steroids to $^{4}$-C$_{19}$-steroids by the 3-HSD 1 isoform expressed principally in peripheral tissues and the subsequent aromatization of androgens to estrogens. Menses in treated females and fertility in males have been reported.\textsuperscript{1099} A 46,XY male had a null mutation in the HSD3B2 gene caused by a premature stop codon (Tyr171X) in one allele and a frameshift in the other allele.\textsuperscript{1100} This patient, who had male pseudhermaphroditism and adrenal insufficiency requiring replacement therapy for survival, virilized at puberty and developed gynecomastia and was sufficiently fertile to father two children.\textsuperscript{1101,1102} Paternity was confirmed in one child by DNA analysis.\textsuperscript{1103} The source of the 3-HSD enzymatic activity to enable the synthesis of testosterone to facilitate spermatogenesis was undetermined. However, the 3-HSD 1 enzyme in peripheral tissues and in small amounts in the testis is a likely candidate.\textsuperscript{1104,1105}

Simard and co-workers reviewed the molecular genetics of 3-HSD 2 deficiency in 17 patients, 10 of whom were salt losers.\textsuperscript{1106} They described eight different point mutations in the 10 latter patients. All of the mutations led to profound impairment of enzymatic activity (to <1%\textsuperscript{1107}) ; they were nonsense, missense, and frameshift mutations that clustered in the fourth exon of the gene (see Fig. 22-62).\textsuperscript{1108} Seven different mutations were found in patients with six families with the so-called nonsalt-losing form of 3-HSD deficiency (see Fig. 22-62).\textsuperscript{1109} All were missense mutations except for one mutation that produced a splice error.\textsuperscript{1110} These mutations resulted in 1% to 10% of the wild-type 3-HSD enzymatic activity in transfected cells.\textsuperscript{1111,1112} Additional studies of the molecular genetics and enzymatic activity of the mutant genes are reported.\textsuperscript{1113,1114} No mutations in either HSD3B1 or HSD3B2 have been found in patients with the putative “late onset,” “attenuated,” or nonclassic form of 3-HSD deficiency.\textsuperscript{1115}

Mendonca and colleagues described four patients in two families with an arginine-to-threonine substitution at amino acid 82 (Arg82Thr) in the HSD3B2.\textsuperscript{1116} Two males in one family were pseudohermaphrodites with differing degrees of mild salt loss. A 46,XX phenotypic female sibling homozygous for the same mutation had mild hypospadias and cryptorchidism.\textsuperscript{1117} No mutations in either HSD3B1 or HSD3B2 have been found in patients with the putative “late onset,” “attenuated,” or nonclassic form of 3-HSD deficiency.\textsuperscript{1118} Some individuals develop severe salt losses in early childhood. Patients with severe 3-HSD deficiency exhibit clinical findings that are consistent with severe glucocorticoid deficiency. A more subtle presentation may result in delayed diagnosis and inappropriate replacement therapy, which may precipitate episodes of adrenal crisis.\textsuperscript{1119}

Additional studies of the molecular genetics and enzymatic activity of the mutant genes are reported.\textsuperscript{1120,1121} No mutations in either HSD3B1 or HSD3B2 have been found in patients with the putative “late onset,” “attenuated,” or nonclassic form of 3-HSD deficiency.\textsuperscript{1115}

TABLE 22-30 -- Clinical Features of CYP17 Mutations with Both 17-Hydroxylase and 17,20-Lyase Deficiencies in 46,XY Males

| Karyotype: | 46,XY |
| Inheritance: | Autosomal recessive; CYP17 (P450C17) gene mutations |
| Genitalia: | Female ambiguous hypoplastic male; blind vaginal pouch |
| Wolffian duct derivatives: | Absent hypospadias |
| Müllerian duct derivatives: | Absent |
| Gonads: | Testes |
| Habitus: | Absent or poor virilization at puberty, gynecomastia |
| Hormone and metabolic profile: | Decreased plasma testosterone; increased plasma LH and FSH levels; increased plasma deoxy corticosterone, corticosterone, and progesterone concentrations; decreased plasma renin activity |

| Low renin hypertension with hypokalemic alkalosis |

LH, luteinizing hormone; FSH, follicle-stimulating hormone.

during the first few weeks of life and by the peripheral conversion of $^{5}$-3-hydroxysteroids to $^{4}$-3-ketosteroids by the type I enzyme in peripheral tissues. Newborn males with severe 3-HSD II deficiency can have elevated levels of 17-hydroxyprogesterone in infancy secondary to the peripheral conversion of 17-hydroxyprogesterone.\textsuperscript{1122} These infants usually present in the first week of life. In contrast, another unrelated 46,XX female from another family with the same mutation had acne and premature pubarche as well as elevated plasma 17-hydroxyprogesterone and DHEA levels.\textsuperscript{1123} Hence, as in patients with 21-hydroxylase deficiency, the same mutations may be “cryptic” in some individuals and severe clues in others result.

The diagnosis of 3-HSD II deficiency should be suspected in all 46,XY males with ambiguous genitalia and adrenal insufficiency. The hormonal characteristics are increased levels of $^{5}$-3-hydroxy C$_{19}$- and C$_{19}$-steroids (e.g., 17-hydroxyprogesterone and DHEA and their sulfates) and their metabolites in plasma and urine and striking increases in ACTH stimulated $^{17}$-17-hydroxyprogrenolone (>50 SD above controls) and $^{17}$-17-hydroxyprogrenolone to cortisol ratios (>30 SD above controls). However, the diagnosis in early infancy can be confounded by the elevated concentrations of $^{5}$-3-hydroxy C$_{19}$- and C$_{19}$-steroids normally seen in premature and full-term infants.
than 25% of normal enzymatic activity is necessary for normal fetal masculinization of the external genitalia. The tests may be intra-abdominal, in the inguinal canal, or in the labioscrotal folds. Inguinal hernias are commonly present. In one affected 46,XY patient, no gonads were found at laparotomy. Müllerian structures are absent, and wolfian derivatives are usually hypoplastic. Bone age is retarded during childhood and can lead to tall stature. Excessive secretion of DOC and corticosterone, the consequence of the failure of 17-hydroxylolation of the C21-steroids, usually leads to hypertension, hypokalemia, and alkalosis. The adrenal zona fasciculata is the source of the increased plasma concentration of DOC, corticosterone, 18-hydroxycorticosterone, and 18-hydroxydeoxycorticosterone. Salt and water retention, volume expansion, and hypertension suppress renin and, consequently, aldosterone secretion in the classic form (although aldosterone concentrations are normal or elevated in some patients). This process is reversible with cortisol therapy. Because gonadal sex steroid secretion is low, severely affected patients fail to develop secondary sexual characteristics, including pubic and axillary hair. Plasma and urinary FSH and LH levels are increased. One patient with a partial deficiency of 17-hydroxylase activity developed prominent gynecomastia and incomplete male secondary sexual characteristics at the expected time of puberty. As previously discussed in the section on CAH, mutations in the CYP17 gene (see Fig. 22-63) that are associated with less than 25% of normal 17-hydroxylase activity in intact transfected cells result in female (complete deficiency) or ambiguous genitalia in affected 46,XY males, whereas activity equal to at least 25% of that of the normal enzyme is associated with normal male genitalia. Smith-Lemli-Opitz Syndrome (7-Dehydrocholesterol Reductase Deficiency)

The Smith-Lemli-Opitz syndrome has a broad phenotypic spectrum but typically includes microcephaly, mental retardation, ptosis, upturned nose, micrognathia, polydactyly, syndactyly of toes, severe hypospadias, microopenis, and growth failure. The abnormalities of the external genitalia in 71% of XY patients vary from hypospadias to complete failure of masculinization, resulting in a female phenotype. It has now been established that the syndrome is caused by a deficiency of 7-dehydrocholesterol reductase (sterol 7-reductase, DHCR7), the phylogenetically conserved sterol-sensing domain-containing enzyme required for the last step in the biosynthetic path-way from acetaldehyde to cholesterol biosynthesis. This is the first characterized metabolic malformation syndrome; the defect in cholesterol metabolism acts as a multisystem teratogen perhaps related in part to defective signaling of Hedgehog proteins. The diagnosis is confirmed biochemically by demonstrating low plasma levels of cholesterol and elevated levels of 7-dehydrocholesterol; the plasma cholesterol concentration correlates negatively with severity. The gene encoding this enzyme (DHCR7) maps to 11q1213 and more than 70 different mutations have now been described. The Smith-Lemli-Opitz syndrome has a broad phenotypic spectrum, but typically includes microcephaly, mental retardation, ptosis, upturned nose, micrognathia, polydactyly, syndactyly of toes, severe hypospadias, microopenis, and growth failure. The abnormalities of the external genitalia in 71% of XY patients vary from hypospadias to complete failure of masculinization, resulting in a female phenotype. It has now been established that the syndrome is caused by a deficiency of 7-dehydrocholesterol reductase (sterol 7-reductase, DHCR7), the phylogenetically conserved sterol-sensing domain-containing enzyme required for the last step in the biosynthetic path-way from acetaldehyde to cholesterol biosynthesis. This is the first characterized metabolic malformation syndrome; the defect in cholesterol metabolism acts as a multisystem teratogen perhaps related in part to defective signaling of Hedgehog proteins. The diagnosis is confirmed biochemically by demonstrating low plasma levels of cholesterol and elevated levels of 7-dehydrocholesterol; the plasma cholesterol concentration correlates negatively with severity. The gene encoding this enzyme (DHCR7) maps to 11q1213 and more than 70 different mutations have now been described. The Smith-Lemli-Opitz syndrome has a broad phenotypic spectrum, but typically includes microcephaly, mental retardation, ptosis, upturned nose, micrognathia, polydactyly, syndactyly of toes, severe hypospadias, microopenis, and growth failure. The abnormalities of the external genitalia in 71% of XY patients vary from hypospadias to complete failure of masculinization, resulting in a female phenotype. It has now been established that the syndrome is caused by a deficiency of 7-dehydrocholesterol reductase (sterol 7-reductase, DHCR7), the phylogenetically conserved sterol-sensing domain-containing enzyme required for the last step in the biosynthetic path-way from acetaldehyde to cholesterol biosynthesis. This is the first characterized metabolic malformation syndrome; the defect in cholesterol metabolism acts as a multisystem teratogen perhaps related in part to defective signaling of Hedgehog proteins. The diagnosis is confirmed biochemically by demonstrating low plasma levels of cholesterol and elevated levels of 7-dehydrocholesterol; the plasma cholesterol concentration correlates negatively with severity. The gene encoding this enzyme (DHCR7) maps to 11q1213 and more than 70 different mutations have now been described. The Smith-Lemli-Opitz syndrome has a broad phenotypic spectrum, but typically includes microcephaly, mental retardation, ptosis, upturned nose, micrognathia, polydactyly, syndactyly of toes, severe hypospadias, microopenis, and growth failure. The abnormalities of the external genitalia in 71% of XY patients vary from hypospadias to complete failure of masculinization, resulting in a female phenotype. It has now been established that the syndrome is caused by a deficiency of 7-dehydrocholesterol reductase (sterol 7-reductase, DHCR7), the phylogenetically conserved sterol-sensing domain-containing enzyme required for the last step in the biosynthetic path-way from acetaldehyde to cholesterol biosynthesis. This is the first characterized metabolic malformation syndrome; the defect in cholesterol metabolism acts as a multisystem teratogen perhaps related in part to defective signaling of Hedgehog proteins. The diagnosis is confirmed biochemically by demonstrating low plasma levels of cholesterol and elevated levels of 7-dehydrocholesterol; the plasma cholesterol concentration correlates negatively with severity. The gene encoding this enzyme (DHCR7) maps to 11q1213 and more than 70 different mutations have now been described. The Smith-Lemli-Opitz syndrome has a broad phenotypic spectrum, but typically includes microcephaly, mental retardation, ptosis, upturned nose, micrognathia, polydactyly, syndactyly of toes, severe hypospadias, microopenis, and growth failure. The abnormalities of the external genitalia in 71% of XY patients vary from hypospadias to complete failure of masculinization, resulting in a female phenotype. It has now been established that the syndrome is caused by a deficiency of 7-dehydrocholesterol reductase (sterol 7-reductase, DHCR7), the phylogenetically conserved sterol-sensing domain-containing enzyme required for the last step in the biosynthetic path-way from acetaldehyde to cholesterol biosynthesis. This is the first characterized metabolic malformation syndrome; the defect in cholesterol metabolism acts as a multisystem teratogen perhaps related in part to defective signaling of Hedgehog proteins. The diagnosis is confirmed biochemically by demonstrating low plasma levels of cholesterol and elevated levels of 7-dehydrocholesterol; the plasma cholesterol concentration correlates negatively with severity. The gene encoding this enzyme (DHCR7) maps to 11q1213 and more than 70 different mutations have now been described. The Smith-Lemli-Opitz syndrome has a broad phenotypic spectrum, but typically includes microcephaly, mental retardation, ptosis, upturned nose, micrognathia, polydactyly, syndactyly of toes, severe hypospadias, microopenis, and growth failure. The abnormalities of the external genitalia in 71% of XY patients vary from hypospadias to complete failure of masculinization, resulting in a female phenotype. It has now been established that the syndrome is caused by a deficiency of 7-dehydrocholesterol reductase (sterol 7-reductase, DHCR7), the phylogenetically conserved sterol-sensing domain-containing enzyme required for the last step in the biosynthetic path-way from acetaldehyde to cholesterol biosynthesis. This is the first characterized metabolic malformation syndrome; the defect in cholesterol metabolism acts as a multisystem teratogen perhaps related in part to defective signaling of Hedgehog proteins. The diagnosis is confirmed biochemically by demonstrating low plasma levels of cholesterol and elevated levels of 7-dehydrocholesterol; the plasma cholesterol concentration correlates negatively with severity. The gene encoding this enzyme (DHCR7) maps to 11q1213 and more than 70 different mutations have now been described. The Smith-Lemli-Opitz syndrome has a broad phenotypic spectrum, but typically includes microcephaly, mental retardation, ptosis, upturned nose, micrognathia, polydactyly, syndactyly of toes, severe hypospadias, microopenis, and growth failure. The abnormalities of the external genitalia in 71% of XY patients vary from hypospadias to complete failure of masculinization, resulting in a female phenotype. It has now been established that the syndrome is caused by a deficiency of 7-dehydrocholesterol reductase (sterol 7-reductase, DHCR7), the phylogenetically conserved sterol-sensing domain-containing enzyme required for the last step in the biosynthetic path-way from acetaldehyde to cholesterol biosynthesis. This is the first characterized metabolic malformation syndrome; the defect in cholesterol metabolism acts as a multisystem teratogen perhaps related in part to defective signaling of Hedgehog proteins. The diagnosis is confirmed biochemically by demonstrating low plasma levels of cholesterol and elevated levels of 7-dehydrocholesterol; the plasma cholesterol concentration correlates negatively with severity. The gene encoding this enzyme (DHCR7) maps to 11q1213 and more than 70 different mutations have now been described.

The diagnosis of 17-hydroxylase deficiency should be suspected in male pseudohermaphrodites, including 46,XY phenotypic females, who have hyporeninemic hypertension and hypokalemia as well as low concentrations of testosterone and hydrocortisone. Plasma concentrations of corticotropin, DOC, corticosterone, and progesterone are high, and those of aldosterone, 17-hydroxyprogesterone, cortisol, and gonadal steroids are low. Replacement therapy with physiologic doses of glucocorticoids suppresses DOC and corticosterone secretion and causes return of serum potassium levels, blood pressure, and plasma renin and aldosterone levels to normal. At puberty, appropriate gonadal steroid replacement therapy is indicated, and gonadectomy should be performed in 46,XY patients who have been assigned a female sex of rearing.

The topology of -5-sterol reductase. Dark circles designate amino acid substitutions due to missense mutations in DHCR7. The two arrows indicate site of stop codons. (From Witsch-Baumgartner M, Loffler J, Utermann G. Mutations in the human DHCR7 gene. Hum Mol Hum Mutat 2001; 17:172182.)
activity can be demonstrated in the human testis or in cultured human theca cells. The conversion of 17-hydroxyprogrenolone to DHEA is about 50 times more efficient than the conversion of 17-hydroxyprogesterone to androstenedione. The control of 17,20-lyase activity in the adrenal appears to be independent of that in the gonad; adrenal 17,20-lyase activity is age dependent, as illustrated by adrenarche. Data suggest that the ratio of 17-hydroxylation to 17,20-lyase activity of the CYP17 enzyme is a function of the molar ratio of its electron transfer (redox) partners P450 oxidoreductase and of cytochrome b5. Increasing the amount of either NADPH-P450 reductase or cytochrome b5 increases the activity of 17,20-lyase sevenfold. Furthermore, the CYP17 enzyme undergoes post-translational modification through phosphorylation of serine and threonine residues by cyclic AMP-dependent protein kinase A. Phosphorylation of the enzyme increases 17,20-lyase activity, and dephosphorylation reduces or eliminates it. Both these mechanisms appear to play a role in control of 17,20-lyase activity in the adrenal and gonads. There have been about 18 case reports of putative isolated 17,20-lyase deficiency, despite the fact that one gene encodes a single enzyme with both 17-hydroxylase and 17,20-lyase activities. Isolated 17,20-lyase deficiency, however, is exceedingly rare; there are only two patients in whom the diagnosis has been established unequivocally by demonstration of the mutation and detailed functional and structural studies.

Zachmann and colleagues initially reported two first cousins with a familial form of male pseudohermaphroditism ascribed to a partial deficiency of 17,20-lyase in both the adrenals and testes. The patients had ambiguous genitalia, inguinal or intra-abdominal testes, and a 46,XY sex chromosome constitution. Both cousins had severe hypospadias with a male-type urethra and male duct development. Only urinary steroids were examined. A sample of testicular tissue from one cousin studied in vitro exhibited a defect in the conversion of C_{19}-steroids to C_{17,20}-steroids. Subsequent studies of the cousins at ages 12 and 15 years disclosed a partial defective partial in the conversion of 3α,5α-C_{19}-steroids to C_{17,20}-steroids. Analysis of the CYP17 gene in one cousin detected compound heterozygosity with two different mutant alleles. Transfection studies indicated combined 17-hydroxylation and 17,20-lyase deficiencies despite the clinical findings. Further study of the steroid pattern in this patient revealed that although 17-hydroxylation activity had been putatively normal in childhood, it was decreased in adulthood.

Two Brazilian male pseudohermaphrodites are described from consanguineous marriages with presumed isolated 17,20-lyase deficiency. Both had 46,XY karyotypes, microphthalmia, perineal hypospadias, bifid scrotum, a blind vaginal pouch, and cryptorchidism. Basal LH and FSH concentrations were elevated in the postpubertal case, and testosterone levels were low. The administration of hCG resulted in a marked rise in plasma 17-hydroxyprogesterone with a paucity of response in DHEA, androstenedione, and testosterone, consistent with isolated 17,20-lyase deficiency. Molecular modeling and site-directed mutagenesis indicated that the Arg347 Ala mutation selectively ablates 17,20-lyase activity in the rat or human enzyme while leaving 17-hydroxylase activation intact. Analyses of CYP17 in these patients showed one to be homozygous for an Arg347His mutation and the other to have an Arg356Gln mutation. These two mutations are located in the redox partner binding site of the CYP17 enzyme and therefore cause a decrease in 17,20-lyase activity by reducing electron transfer. These are the first patients with isolated 17,20-lyase deficiency defined by molecular analyses and the first example of prediction of the specific location for a mutation by site-directed mutagenesis and modeling. The role of redox partners in the regulation of 17,20-lyase is illustrated by a description of a male pseudohermaphrodite with congenital methemoglobinemia due to a mutation in cytochrome b5.

Depending on the degree of impairment in 17,20-lyase activity and its effect on fetal testosterone production during gestation, the external appearance may vary from female to ambiguous to hypoplastic male. The testes may be intra-abdominal, in the inguinal region, or in the scrotum. As with other defects in testosterone synthesis, the Wolffian duct derivatives are either hypoplastic or normal, depending on the severity of the testosterone deficiency, and müllerian duct derivatives are absent. Gynecomastia can occur at the age of puberty. In the 46,XX females, putative isolated 17,20-lyase deficiency leads to failure of pubertal development and elevated gonadotropin levels.

The diagnosis of 17,20-lyase deficiency should be considered in male pseudohermaphrodites with absent müllerian derivatives and in 46,XX females who have no abnormality in glucocorticoid or mineralocorticoid synthesis but fail to develop secondary sexual characteristics at the expected time of puberty and have elevated concentrations of FSH and LH. In prepubertal male pseudohermaphrodites, 17,20-lyase deficiency must be distinguished from the partial form of androgen resistance, 5-reductase deficiency, and 17,20-HSD type 3 deficiency. In the prepubertal patient, both corticotropin and hCG stimulation may be useful in unmasking the defect. Prenatal diagnosis is possible by the measurement of androgenic hormone (C_{19}-steroids to testosterone (C_{17,20}-steroids) in placental liver and endometrial microsomes and primarily oxidizes (inactivates) both androgens and estrogens. The 17-HSD 3, a microsomal enzyme, utilizes NADPH as a cofactor and it is encoded by a gene on chromosome 9q22 that is 23% homologous to the genes for 17-HSD 1 and 17-HSD 2 and is expressed primarily in the testes, where it favors the reduction of androstenedione to testosterone; male pseudohermaphroditism is a consequence of mutations in the gene encoding 17-HSD 3. The 17-HSD 3 gene is expressed in multiple tissues. 17-HSD 3 is encoded by a gene on chromosome 10p14 and favors the reduction of androstenedione to testosterone and of DHEA to androst-5-ene-3,17-diol in peripheral tissues and the ovary. 17-HSD 7 is expressed in the placenta, mammary gland, and kidney and efficiently catalyzes the conversion of E_{1} (DHEA) to E_{2} (Estradiol).

Male pseudohermaphrodites caused by 17-HSD 3 deficiency (also called 17-hydroxysteroid oxidoreductase or 17-ketosteroid reductase) was first reported by Saenz and colleagues (Table 22-32). Many patients have been described, including a cohort of 68 subjects from a highly inbred population in the Gaza Strip. Except for a few 46,XY individuals with ambiguous genitalia at birth, most affected 46,XY males have predominantly female external genitalia, testes (usually located in the inguinal canal), male Wolffian duct derivatives (epididymides, vas deferens, seminal vesicles, and ejaculatory ducts), and a blind vaginal pouch. Because of unambiguous female genitalia at birth, such individuals are usually assigned a female sex, raised

<table>
<thead>
<tr>
<th>Karyotype:</th>
<th>46,XY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inheritance:</td>
<td>Autosomal recessive; mutations in HSD17B3 gene</td>
</tr>
<tr>
<td>Genitalia:</td>
<td>Female ambiguous; blind vaginal pouch</td>
</tr>
<tr>
<td>Wolffian duct derivatives:</td>
<td>Hypoplastic</td>
</tr>
<tr>
<td>Müllerian duct derivatives:</td>
<td>Absent</td>
</tr>
<tr>
<td>Gonads:</td>
<td>Testes (usually undesended)</td>
</tr>
<tr>
<td>Habitus:</td>
<td>Virilization at puberty (phallus enlargement, deepening of voice, and development of facial and body hair); gynecomastia variable</td>
</tr>
</tbody>
</table>
As females, and in some cases mistakenly assumed to have the complete androgen resistance syndrome. However, at the age of puberty, gonadotropin levels and plasma concentrations of androstenedione, estrone, and testosterone increase. The levels of testosterone in some cases approach the normal male range, and some virilization invariably occurs. **Hormone profile:**
- Increased plasma estrone and androstenedione; decreased ratio of plasma testosterone/androstenedione and estradiol/estrone after hCG stimulation test; increased plasma FSH and LH levels.

**hCG, human chorionic gonadotropin; FSH, follicle-stimulating hormone; LH, luteinizing hormone**

Figure 22-74 Diagram of the gene encoding 17-hydroxyxandrostenedione dehydrogenase type 3 with selected mutations reported to cause 17-HSD deficiency. The exons are the numbered black boxes and the missense mutations causing amino acid substitutions in the enzyme are indicated by the three-letter abbreviation for the wild-type amino acid, followed by the amino acid number in the enzyme and the three-letter abbreviation for the substituted amino acid. The position of the amino acid of the mutation from the amino terminus is indicated by the number. **The mutations included a frameshift, 3 splice site abnormalities, and 10 missense mutations. Expression of 8 of the 9 missense mutations revealed complete absence of 17-HSD 3 enzymatic activity.**

The Arg20Gln mutation in the Ghana Strip Arab patients resulted in an enzyme with partial activity. **Only mutations in the HSD17B3 gene that encodes the testis-specific enzyme cause male pseudohormonaphroditism. The gene is located on chromosome 9q22 and is composed of 11 exons.** The **gene that encodes the testis-specific enzyme cause male pseudohormonaphroditism. The gene is located on chromosome 9q22 and is composed of 11 exons.**

Ala56Thr, on kinetic analysis showed a 20-fold decrease in NADPH cofactor affinity and a 6-fold decrease in affinity for the androstenedione substrate.

**The presence of wolffian duct derivatives in these patients with homozygous mutations of the 17-HSD 3 isoenzyme that result in complete absence of enzymatic activity is unexplained.**

Analyses of 17 patients with classic 17-HSD 3 deficiency, including 4 from San Francisco, revealed 14 mutations in the HSD17B3 gene. **Twelve patients had homozygous mutations, 4 were compound heterozygotes, and 1 was a presumed heterozygote (see Fig. 22-74).** The mutations included a frameshift, 3 splice site abnormalities, and 10 missense mutations. Expression of 8 of the 9 missense mutations revealed complete absence of 17-HSD 3 enzymatic activity. **The Arg20Gln missense mutation in the Ghana Strip Arab patients resulted in an enzyme with partial activity.**

Little or no HSD 3 is expressed in the human ovary. **Women with homozygous or compound heterozygous mutations of 17-HSD 3 are asymptomatic.** A putative late-onset form of 17-HSD deficiency causing gynecomastia and hypogonadism in males has been reported but has not been confirmed by DNA analyses of the coding sequence of the HSD17B3 gene.

In the past, patients reared as females treatment involved castration followed by estrogen substitution therapy at puberty, but this approach needs to be reexamined, especially in view of the dramatic advances in neonatal diagnosis and DNA analysis. Patients with ambiguous genitilia in whom 17-HSD deficiency is detected in infancy should probably be reared as males; testosterone therapy to augment phallic size and genitoplasty are indicated in infancy. As noted previously, male pseudohormonaphroditism is caused by 17-HSD deficiency and is relatively common among Arabs of the Gaza Strip. The natural history in this isolated virilization at puberty; furthermore, a change in gender role behavior from female to male is the rule. Because of this, Gros and colleagues proposed that these patients should be given male gender assignment at diagnosis. They described seven young affected 46,XY males with female external genitilia who after biochemical confirmation were treated with testosterone enanthate, 25 to 50 mg each month for 3 months. Most patients received two or three courses of testosterone therapy, which resulted in an increase in phallic length into the normal range for age.

Although 17-HSD 3 activity appears to be completely deficient in infant, a progressive rise occurs in plasma testosterone from puberty to adulthood. This apparent "recovery" of 17-HSD 3 enzymatic activity is undoubtedly a result of the increase with puberty in gonadotropin and androstenedione secretion as well as the extraglandal activity of the other 17-HSD isozymes in converting androstenedione to testosterone. **However, the patients described from the Gaza Strip had a less severe enzyme block, with 15% to 20% of normal 17-HSD 3 activity and evidence of synthesis of testosterone by the testes at the expected time of puberty.**

The presence of wolffian duct derivatives in these patients with homozygous mutations of the 17-HSD 3 isoenzyme that result in complete absence of enzymatic activity is unexplained. Anderson and colleagues suggested that because the androgen receptor in the wolffian duct appears to be identical to the mature androgen receptor, there must be an alternate pathway for testosterone synthesis by 17-HSD in utero to induce wolffian duct stabilization. The diagnosis of 17-HSD 3 deficiency should be considered in (1) male pseudohormonaphroditism who have no abnormality in adrenal steroid biosynthesis, absent müllerian ducts, and normal wolffian duct structures and (2) male pseudohormonaphroditoids who virilize at puberty either with or without gynecomasia. The absence of müllerian duct derivatives generally distinguishes patients with defective testosterone biosynthesis or androgen resistance from those with dysgenetic male pseudohermaphroditism. In the prepubertal or young adolescent patient, basal androstenedione and estrone levels may not be elevated for age. However, at any age the defect in testosterone biosynthesis can be demonstrated best by a prolonged hCG stimulation test. In response to hCG, there is a disproportionate rise in plasma androstenedione and estrone levels, with testosterone and estradiol concentrations.

The ratio of plasma testosterone to androstenedione concentrations after hCG stimulation usually well distinguishes androgen insensitivity from 17-HSD 3 deficiency. A ratio less than 0.8 is typically found in 17-HSD deficiency; although similarly low ratios can be encountered in testicular dysgenesis, the absolute amount of plasma C19 and C18 steroids after the administration of hCG is strikingly different.

Hormone profile: Increased plasma estrone and androstenedione; decreased ratio of plasma testosterone/androstenedione and estradiol/estrone after hCG stimulation test; increased plasma FSH and LH levels.

**hCG, human chorionic gonadotropin; FSH, follicle-stimulating hormone; LH, luteinizing hormone**
Defects in Androgen-Dependent Target Tissues

A defect at any step in the mechanism of action of androgens on their target cells (see Fig. 22-29) (Figure Not Available) 5-reduction of testosterone, binding of DHT to the receptor, nuclear localization of the steroid receptor complex to hormone response elements on DNA and subsequent transcription, or translation can lead to impaired androgen action and result in male pseudohermaphroditism. Two major forms have been identified: end-organ resistance to androgenic hormones (androgen receptor defects) and errors in testosterone metabolism in target tissues (5-reductase deficiency).

End-Organ Resistance to Androgens (Androgen Receptor Defects)

The spectrum of phenotypes in 46,XY individuals with resistance to androgens varies from patients with normal female external genitalia through those with genital ambiguity to those with a normal male phenotype who have a small phallus and are fertile. The syndromes of androgen insensitivity are a paradigm of clinical disorders resulting from resistance to the action of normal or increased circulating concentrations of a hormone. 

Complete Androgen Insensitivity Syndrome and Its Variants (Androgen Resistance, Testicular Feminization, Feminizing Testes).

The term "testicular feminization," coined by Morris but no longer used, was applied to a highly distinctive X-linked disorder in which affected males are phenotypic females and develop female secondary sexual characteristics at puberty but fail to menstruate (Table 22-33). Affected individuals are genetic males as shown by the 46,XY karyotype and have testes. The prevalence of this disorder is estimated at between 1 in 20,000 and 1 in 60,000 live male births.

The syndromes of androgen insensitivity represent the most common identifiable cause of male pseudosexual disorder. Phenotypically, these patients have unambiguous female external genitalia, hypoplastic labia majora; a blind vaginal pouch; absent or, rarely, vestigial Müllerian structures (uterus and tubes); testes located in the labia, the inguinal canal, or the abdomen; and absent or vestigial Wolffian derivatives. Histologically, the testes are difficult to distinguish from normal before puberty. After puberty, there are Sertoli-cell-only seminiferous tubules within nodules, spermatogenesis are sparse, and spermatogenesis is absent. The Leydig

<table>
<thead>
<tr>
<th>Karyotype:</th>
<th>46,XY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inheritance:</td>
<td>X-linked recessive; mutations in AR gene</td>
</tr>
<tr>
<td>Genitalia:</td>
<td>Female with blind vaginal pouch</td>
</tr>
<tr>
<td>Wolffian duct derivatives:</td>
<td>Usually absent; less commonly, rudimentary or hypoplastic</td>
</tr>
<tr>
<td>Müllerian duct derivatives:</td>
<td>Absent or vestigial</td>
</tr>
<tr>
<td>Gonads:</td>
<td>Testes</td>
</tr>
<tr>
<td>Hormone and metabolic profile:</td>
<td>Increased plasma LH and testosterone concentration; increased estradiol (for men); FSH levels often normal or slightly increased</td>
</tr>
<tr>
<td>Androgen receptor studies:</td>
<td>Genetic heterogeneity; mutations can lead to low or undetectable amount of normal receptor (receptor-negative), unstable receptor (thermolabile, partial receptor deficiency), or the receptor-positive form</td>
</tr>
<tr>
<td></td>
<td>LH, luteinizing hormone; FSH, follicle-stimulating hormone.</td>
</tr>
</tbody>
</table>

Table 22-33 – Clinical Features of Complete Androgen Resistance

cells are hyperplastic and tend to form adenomatous clumps. The testes are predisposed to malignant transformation. Carcinoma in situ and seminoma have been reported, especially in patients with the partial form of androgen insensitivity. The overall risk of a testicular neoplasm in the affected adult has been estimated at 4% to 9%; however, the risk appears to be significantly less in those younger than 20 years of age.

At birth and in childhood the diagnosis should be suspected in phenotypic females with an inguinal hernia (particularly if bilateral) and a testis-like mass in the inguinal region or in the labia. It has been estimated that 1% to 2% of phenotypic females with inguinal hernias have the syndrome. Nevertheless, it is debatable whether a routine karyotype should be performed in all female infants with inguinal hernias. At adolescence, female secondary sexual characteristics develop and include normal breasts and female body habitus but no menses. Pubic and axillary hair is usually sparse and is completely lacking in about one third of patients. This was the case in a family with complete deletion of the androgen receptor, representing the null phenotype of this syndrome.

Pathophysiology and Hormone Profile

The Androgen Receptor

In 1950 Wilkins first suggested that failure of androgenization of the male fetus and the development of female rather than male secondary sexual characteristics at puberty could be explained by end-organ unresponsiveness to androgen based on lack of virilization with 50 mg of methyltestosterone administered daily . Studies by subsequent workers supported this contention by failing to demonstrate a clinical or metabolic response to testosterone administration in patients with the complete form of this syndrome.
A Japanese patient with the syndrome was reported to have a normal gene on sequencing and absence of a specific co-activator for transcription factors. ARA 70 cofactor, which appears to bind relatively specifically to the androgen receptor, was screened for mutations in a group of patients with partial androgen resistance.  

The role of co-regulator protein dysfunction in androgen action may be relevant in some forms of androgen insensitivity. Targeted disruption of the mouse androgen receptor gene led to a more functional interpretation of androgen-binding results. The lack of strict correlation between phenotype and receptor binding, as well as the undetectable or low amount of androgen receptor activity was subsequently demonstrated in cultured fibroblasts from the genital skin of karyotypic males with the syndrome. These observations were amply confirmed by others. The lack of androgen binding in genital skin fibroblasts from patients with this disorder provided an explanation for the observed failure of androgen action.

Extensive studies of the parameters of androgen binding in genital skin fibroblasts using DHT and synthetic androgens showed a range of quantitative and qualitative defects. Typically, there was absent specific androgen binding in patients with the complete androgen insensitivity. In contrast, there was detectable binding in the partial form of the syndrome, but this was qualitatively abnormal as based on characteristics such as decreased binding affinity, thermolability of binding, and increased dissociation kinetics. In some instances, there was curiously normal or increased androgen binding that was only adequately explained once the androgen receptor had been cloned. Knowledge of the functional domains of the androgen receptor protein as a result of characterizing the androgen receptor gene led to a more functional interpretation of androgen-binding results. The lack of strict correlation between phenotype and receptor binding, as well as the apparently normal binding found in patients with androgen resistance, was clarified by studies of the molecular biology of the androgen receptor and of mutations in the gene encoding the receptor.

A large number of androgen receptor gene mutations have been described in patients with androgen insensitivity, and they are recorded on an international database (see http://www.mcgill.ca/androgendb/). About 400 mutations distributed throughout the gene are now described. The vast majority are missense/nonsense nucleotide substitutions located predominantly in exons encoding the hormone-binding domain. There are also examples of splice site intronic mutations, small deletions and insertions, and, rarely, complete deletions of the gene. Mental retardation was an associated feature.

**Molecular Biology of the Androgen Receptor (see Hormonal Sex Differentiation)**

The androgen receptor gene is located at Xq11-q12, comprises eight exons, and encodes a 110- to 114-kd protein that varies in length from 910 to 919 amino acids. Exons 1 through 4 encode the DNA-binding region of the androgen receptor. Part of exon 4 (Fig. 22-77) encodes the hinge region of the androgen receptor, which contains a nuclear localization signal. The 3' end of exon 4 through exon 8 encodes the ligand (androgen)-binding region. A mutational hot spot is located in exon 5 and can cause both CAR and PAR. Not shown are nonsense, frameshift, and splice junction mutations and deletions that can cause either CAR or PAR. (Redrawn from Quigley CA, De Bellis A, Marschke KB, et al. Androgen receptor defects: historical, clinical and molecular perspectives. Endocr Rev 1995; 16:271321. © 1995, The Endocrine Society.)

Asterisks indicate mutations that have been found to cause both complete and partial androgen resistance. Each mutation is indicated by the three-letter abbreviation for the wild-type amino acid.

![Diagram of the androgen receptor gene (AR) with missense mutations that cause complete androgen resistance (CAR) and partial androgen resistance (PAR).](image)

A large number of androgen receptor gene mutations have been described in patients with androgen insensitivity, and they are recorded on an international database (see http://www.mcgill.ca/androgendb/). About 400 mutations distributed throughout the gene are now described. The vast majority are missense/nonsense nucleotide substitutions located predominantly in exons encoding the hormone-binding domain. There are also examples of splice site intronic mutations, small deletions and insertions, and, rarely, complete deletions of the gene. Mental retardation was an associated feature.

**Molecular Biology of the Androgen Receptor (see Hormonal Sex Differentiation)**

The androgen receptor gene is located at Xq11-q12, comprises eight exons, and encodes a 110- to 114-kd protein that varies in length from 910 to 919 amino acids.

Asterisks indicate mutations that have been found to cause both complete and partial androgen resistance. Each mutation is indicated by the three-letter abbreviation for the wild-type amino acid.

![Diagram of the androgen receptor gene divided into its eight exons. Exon 1 codes for the NH2-terminal domain and regulates transcription. Exons 2 and 3 code for two zinc fingers. Exons 4 through 8 code for the androgen-binding domain of the receptor.](image)

The 3' end of exon 4 through exon 8 encodes the ligand (androgen)-binding region. A mutational hot spot is located in exon 5 and can cause both CAR and PAR. Not shown are nonsense, frameshift, and splice junction mutations and deletions that can cause either CAR or PAR. (Redrawn from Quigley CA, De Bellis A, Marschke KB, et al. Androgen receptor defects: historical, clinical and molecular perspectives. Endocr Rev 1995; 16:271321. © 1995, The Endocrine Society.)

Asterisks indicate mutations that have been found to cause both complete and partial androgen resistance. Each mutation is indicated by the three-letter abbreviation for the wild-type amino acid.

![Diagram of the androgen receptor gene divided into its eight exons. Exon 1 codes for the NH2-terminal domain and regulates transcription. Exons 2 and 3 code for two zinc fingers. Exons 4 through 8 code for the androgen-binding domain of the receptor.](image)

The 3' end of exon 4 through exon 8 encodes the ligand (androgen)-binding region. A mutational hot spot is located in exon 5 and can cause both CAR and PAR. Not shown are nonsense, frameshift, and splice junction mutations and deletions that can cause either CAR or PAR. (Redrawn from Quigley CA, De Bellis A, Marschke KB, et al. Androgen receptor defects: historical, clinical and molecular perspectives. Endocr Rev 1995; 16:271321. © 1995, The Endocrine Society.)

Asterisks indicate mutations that have been found to cause both complete and partial androgen resistance. Each mutation is indicated by the three-letter abbreviation for the wild-type amino acid.

**Molecular Biology of the Androgen Receptor (see Hormonal Sex Differentiation)**

The androgen receptor gene is located at Xq11-q12, comprises eight exons, and encodes a 110- to 114-kd protein that varies in length from 910 to 919 amino acids.

Asterisks indicate mutations that have been found to cause both complete and partial androgen resistance. Each mutation is indicated by the three-letter abbreviation for the wild-type amino acid.

![Diagram of the androgen receptor gene divided into its eight exons. Exon 1 codes for the NH2-terminal domain and regulates transcription. Exons 2 and 3 code for two zinc fingers. Exons 4 through 8 code for the androgen-binding domain of the receptor.](image)

The 3' end of exon 4 through exon 8 encodes the ligand (androgen)-binding region. A mutational hot spot is located in exon 5 and can cause both CAR and PAR. Not shown are nonsense, frameshift, and splice junction mutations and deletions that can cause either CAR or PAR. (Redrawn from Quigley CA, De Bellis A, Marschke KB, et al. Androgen receptor defects: historical, clinical and molecular perspectives. Endocr Rev 1995; 16:271321. © 1995, The Endocrine Society.)

Asterisks indicate mutations that have been found to cause both complete and partial androgen resistance. Each mutation is indicated by the three-letter abbreviation for the wild-type amino acid.

![Diagram of the androgen receptor gene divided into its eight exons. Exon 1 codes for the NH2-terminal domain and regulates transcription. Exons 2 and 3 code for two zinc fingers. Exons 4 through 8 code for the androgen-binding domain of the receptor.](image)

The 3' end of exon 4 through exon 8 encodes the ligand (androgen)-binding region. A mutational hot spot is located in exon 5 and can cause both CAR and PAR. Not shown are nonsense, frameshift, and splice junction mutations and deletions that can cause either CAR or PAR. (Redrawn from Quigley CA, De Bellis A, Marschke KB, et al. Androgen receptor defects: historical, clinical and molecular perspectives. Endocr Rev 1995; 16:271321. © 1995, The Endocrine Society.)

Asterisks indicate mutations that have been found to cause both complete and partial androgen resistance. Each mutation is indicated by the three-letter abbreviation for the wild-type amino acid.

![Diagram of the androgen receptor gene divided into its eight exons. Exon 1 codes for the NH2-terminal domain and regulates transcription. Exons 2 and 3 code for two zinc fingers. Exons 4 through 8 code for the androgen-binding domain of the receptor.](image)

The 3' end of exon 4 through exon 8 encodes the ligand (androgen)-binding region. A mutational hot spot is located in exon 5 and can cause both CAR and PAR. Not shown are nonsense, frameshift, and splice junction mutations and deletions that can cause either CAR or PAR. (Redrawn from Quigley CA, De Bellis A, Marschke KB, et al. Androgen receptor defects: historical, clinical and molecular perspectives. Endocr Rev 1995; 16:271321. © 1995, The Endocrine Society.)

Asterisks indicate mutations that have been found to cause both complete and partial androgen resistance. Each mutation is indicated by the three-letter abbreviation for the wild-type amino acid.

![Diagram of the androgen receptor gene divided into its eight exons. Exon 1 codes for the NH2-terminal domain and regulates transcription. Exons 2 and 3 code for two zinc fingers. Exons 4 through 8 code for the androgen-binding domain of the receptor.](image)

The 3' end of exon 4 through exon 8 encodes the ligand (androgen)-binding region. A mutational hot spot is located in exon 5 and can cause both CAR and PAR. Not shown are nonsense, frameshift, and splice junction mutations and deletions that can cause either CAR or PAR. (Redrawn from Quigley CA, De Bellis A, Marschke KB, et al. Androgen receptor defects: historical, clinical and molecular perspectives. Endocr Rev 1995; 16:271321. © 1995, The Endocrine Society.)

Asterisks indicate mutations that have been found to cause both complete and partial androgen resistance. Each mutation is indicated by the three-letter abbreviation for the wild-type amino acid.

![Diagram of the androgen receptor gene divided into its eight exons. Exon 1 codes for the NH2-terminal domain and regulates transcription. Exons 2 and 3 code for two zinc fingers. Exons 4 through 8 code for the androgen-binding domain of the receptor.](image)

The 3' end of exon 4 through exon 8 encodes the ligand (androgen)-binding region. A mutational hot spot is located in exon 5 and can cause both CAR and PAR. Not shown are nonsense, frameshift, and splice junction mutations and dele
of some hormone resistance syndromes being caused by disorders in nuclear co-regulators.  

Germ line mutations have been reported in male breast cancer; the mutations affected adjacent arginine residues in the second zinc finger part of the DNA-binding domain in all three reported patients. Somatic mutations are described in prostate cancer, a disease that is androgen dependent in its initial stages. A common mutation is Thr877Ala, which is also found in the human prostate LNCAP cell line. A unique inactivating somatic mutation (Ty649Cys) has been identified that is transcriptionally inactive and appears to sequester the SRC1 co-activator protein, a pathophysiologic feature analogous to what occurs as a result of the increased triplet repeats in Kennedy’s disease.

Hormone Profile

The hormone profile is similar in all variants of androgen insensitivity but is best characterized in the complete form. The hallmark at puberty and in the adult is an elevated concentration of plasma LH and testosterone in the absence of virilization. However, plasma LH and testosterone levels were strikingly low in 9 of 10 neonates with the complete form of androgen resistance at 30 and 60 days of age as was the LH response to the administration of LHRH. In contrast, 5 similarly studied neonate with partial androgen resistance had plasma testosterone levels in the high normal range and exhibited a normal neonatal LH surge. The suggestion is that the normal postnatal range of testosterone and LH in male infants is abolished or attenuated by the lack of or very functionally defective androgen receptors in the hypothalamus whereas the less functionally defective androgen receptor abnormalities in the partial androgen resistance syndrome are sufficient for the expression of the characteristic surge.

The pattern of LH and testosterone at puberty and in the adult is much the same in the partial form of the syndrome. Indeed, elevated LH in the face of normal to increased adult male-related testosterone concentrations in men with reduced sperm counts may be a marker of mild androgen insensitivity secondary to an androgen receptor gene mutation. Estradiol levels are elevated, both as a direct effect of testosterone secretion and peripheral aromatization of androgens to estrogens.  

Feminization takes place at puberty amplified by estrogen action unopposed by androgens. Plasma FSH concentrations are generally within the normal range. When gonadectomy is performed, there is a further elevation of plasma LH and a rise in FSH concentration that suggests that both estradiol and inhibit play a role in the negative feedback of gonadotropins in patients with androgen resistance.

Sex hormone binding globulin (SHBG) levels are higher in adult females compared with males as a result of an estrogen effect. SHBG levels are in the female range in androgen insensitivity. The lack of SHBG response to the rise in testosterone concentrations after hCG stimulation has been proposed as a useful biochemical screen for androgen insensitivity before puberty. Furthermore, the absence of the androgen-steroid binding protein fails to suppress SHBG levels in complete androgen insensitive patients, whereas there was an intermediate response in patients with the partial form as compared with normal male controls. However, this bioassay is not useful in newborns because of the normal decrease in the concentration of SHBG in the neonate. Few confirmatory data have been published in older patients; an increased plasma level of AMH may be a marker of androgen resistance during the first year of life and after the onset of puberty.

Diagnosis

The diagnosis in the complete form of androgen insensitivity is relatively straightforward and can usually be established on clinical grounds. This is particularly so when there is a positive family history. It is not unusual to establish a history of an older "sister" having had surgery to repair bilateral inguinal hernias. The main differential in infancy is 17-HSD 3 deficiency and Leydig cell hypoplasia (LH/ICG resistance), hence the recommendation to perform an hCG stimulation test. Later, the differential is mainly with complete XY gonadal dysgenesis, where typically there is absent or poor breast development and the retention of müllerian structures.

The full form of the syndrome is much more of a diagnostic challenge; the disorders that can give rise to a similar phenotype have already been discussed. As yet there is no readily available, clinically practical, in vivo or in vitro assay of the trans-activational capacity of the androgen receptor; but new approaches are promising. It is mandatory to assess the androgen response to hCG stimulation, ensuring that androstenedione, testosterone, and DHT concentrations are measured in plasma. There is sometimes a response in growth of the phallicus after the hCG stimulation test, particularly if prolonged. Furthermore, a short course of testosterone injections is recommended as a diagnostic test. However, a response does not necessarily exclude partial androgen insensitivity because it is now known from trans-activation studies of mutant receptors in vitro that certain androgen receptor gene mutations are compatible with male sex of rearing. The phallic response to testosterone is an important assessment in the recommendation for sex assignment in the affected infant. In addition, the phallic response has been used as a predictive test of androgen responsiveness in utero. If the phallic growth, in some cases the results have been equivocal and a neonatal response has not invariably followed by a response to testosterone at puberty.

Management

The sex of rearing is female in the complete form of androgen insensitivity, and the gonads need to be removed, owing to a later risk of malignancy. The tests may be left in place (especially if they are not located in the labia major) until late adolescence to provide a natural source of estrogen. The patient may then elect to undergo orchidectomy, having undergone spontaneous pubertal feminization. Using laparoscopy it is feasible to repair an inguinal hernia in childhood without removing the testis. The status of the intra-abdominal part of the appendix is not removed in infancy or early childhood. There is also the question of a modest decrease in bone mineralization in patients with complete androgen insensitivity, although the risk does not seem to be related to when gonadectomy is performed. The limited data available on bone mineral density in partial androgen insensitivity indicate less of a deficit in bone mineralization.

A potential surgical problem in the complete form of androgen insensitivity is the shortened vagina. There are numerous techniques described to re-vagina a fashion, and the details are beyond the scope of this chapter. Most patients do not require surgery if graded vaginal dilators are used on a regular basis in adolescence. Estrogen treatment is required after gonadectomy; a transdermal patch is widely used but is expensive. It is generally thought that the lack of a uterus precludes the necessity for concomitant treatment with a progestogen. Outcome studies indicate that psychosexual development and sexual function are generally satisfactory in adult life. The practice of not disclosing full information to the patient about the nature of her medical condition is no longer acceptable. The authors recommend that with consent of the parents and the establishment of an appropriate psychosocial support system, the diagnosis, pathophysiology, quality of life issues, and in our view on a need-to-know basis. A potentially painful if not disasterous scenario is one in which the adult patient inadvertently discovers the diagnosis. Carrier detection for genetic counseling purposes is often requested and can be undertaken once the gene mutation within a family has been identified. The polymorphic CAG repeat sequence in exon 1 is useful to analyze segregation of the two X chromosomes in a carrier female. Furthermore, this linked marker can be useful to exclude the possibility of partial androgen insensitivity in familial hypospadias of unknown cause.

Management of partial androgen insensitivity syndrome is complicated by the major issue of predicting the likely response to androgens at puberty. The disorder is not different to other causes of severe undermasculinization (e.g., in 45X/46,XY karyotype) where decisions are often based on the response of the penis to a course of intramuscular testosterone. The surgical issues concerning creation a functional phallicus, which one. The surgery has been an important consideration, are no longer compelling in light of the remarkable advances in the surgical correction of severe degrees of hypospadias. There is evidence that male sex of rearing is increasingly chosen despite severe undermasculinization at birth. Some patients respond at least partially to high-dose androgen therapy. Outcome studies are sparse, but studies of males with microphoria reported heterosexual orientation and normal male sexual function in the majority. If the parents select a female sex assignment, it is necessary to inform them that a variable degree of masculinization as well as feminization will occur at puberty. In patients in whom the
assigned female sex and their gender identity are congruent, orchidectomy is usually advisable by early puberty.

Partial (Incomplete) Form of Androgen Resistance and Its Variants (Reifenstein’s Syndrome)

A heterogeneous group of 46,XY individuals have partial androgen resistance. The external genitalia range from perineoscrotal hypospadias with cryptorchidism and micropenis to clitoromegaly with partial labial fusion. The patients described in the past by Lubbs, Gilbert, Dreyfus, Reifenstein, Rosewater, Walker, and their associates quite likely had partial androgen resistance. The variable degree of masculinization of affected males within and between kindreds is well illustrated by one family studied by Wilson and colleagues. Eleven males were affected; two had a relatively mild defect in masculinization of the external genitalia (small penis and bifid scrotum), eight had peripheral hypospadias, and one had hypospadia, a urogenital sinus with a blind vaginal pouch, and an absent vas deferens. All lacked müllerian structures. In contrast, families with the complete form of androgen resistance exhibit little variability in expression of the mutant gene. The most common presentation in infancy is that of an apparent male with third-degree hypospadias (the urethral orifice located at the base of the phallus), a small penis, and, often, cryptorchidism. Müllerian duct derivatives are absent; Wolffian duct derivatives are usually present but hypoplastic. At puberty, pubic and axillary hair and gynecomastia usually develop, male secondary sexual characteristics are incompletely developed, and the testes remain small and exhibit azospermia because of germinal cell arrest beyond the primary spermatocyte stage. Less severely affected men may exhibit a bifid scrotum, infertility, and poor virilization at puberty. More severely affected males may have ambiguous genitalia, a blind vaginal pouch, and poorly developed Wolffian structures. As in other patients with androgen insensitivity, the concentrations of plasma LH and testosterone are elevated and the high LH levels are resistant to suppression by exogenous androgens. Estradiol and testosterone production rates are increased. However, the degree of feminization at puberty, despite elevated estradiol secretion, is less than in the complete form of androgen resistance.

Analysis of a large kindred with Reifenstein’s syndrome (partial androgen resistance syndrome) led to the detection of two phenotypically normal men who were infertile and lacked the clinical features of androgen resistance. These infertile males could not be distinguished endocrinologically or by androgen receptor studies from their more severely affected relatives. Subsequently, Aiman and co-workers reported infertility in three unrelated men with uninformative family histories and a quantitatively defective androgen receptor. Two of these men had a normal adult male phenotype, and one had mild feminine habits. In a study of the androgen receptor in 28 unrelated, phenotypically normal men with idiopathic azoospermia or oligospermia, Aiman and Griffin estimated the frequency of androgen receptor abnormalities in men with idiopathic infertility. Aiman and Griffin also studied 28 unrelated, phenotypically normal men with idiopathic azospermia or oligospermia. Using genital skin fibroblasts, they observed a partial deficiency of the androgen receptor. Further studies suggested the existence of both quantitative and qualitative abnormalities in infertile males.

Molecular analysis of the androgen receptor in transfected cells demonstrated a decrease in hormone and metabolic profile: Increased plasma LH and testosterone concentrations; increased estradiol (for men); FSH levels may be normal or slightly increased. Partial resistance to androgenic and metabolic effects of testosterone. Genetic heterogeneity; partial deficiency of normal receptor; mutations lead to qualitatively abnormal receptor.

| Karyotype: 46,XY |
| Inheritance: X-linked recessive; mutations in \( \alpha \) subunit gene |
| External genitalia: Ambiguous with blind vaginal pouch hypoplastic male normal male with infertility normal fertile male |
| Wolffian duct derivatives: Rudimentary hypoplastic normal |
| Müllerian duct derivatives: Absent |
| Gonads: Testes (usually undescended) |
| Habitus: Decreased to normal auxiliary and pubic hair, beard growth, and body hair; gynecomastia common at puberty |
| Hormone and metabolic profile: Genetic heterogeneity; partial deficiency of normal receptor; mutations lead to qualitatively abnormal receptor |

Androgen Resistance in Infertile Men

It was postulated that infertility in otherwise normal men was the most consistent and most subtle clinical manifestation of quantitative and qualitative defects in the androgen receptor. However, families have now been described that include men with normal male genitalia, postpubertal gynecomastia, and poor virilization in spite of elevated plasma testosterone levels. Studies of the androgen receptor in affected individuals detected several qualitative abnormalities, including receptor instability, failure of up-regulation, increased dissociation of receptor-synthetic androgen complexes, and thermal instability. Molecular analysis of the AR gene in one patient revealed a single nucleotide substitution in the androgen-binding domain, a leucine to phenylalanine change at amino acid 789 (Leu789Phe). Expression of the mutant gene in transfected cells demonstrated a decrease in trans-activational activity with this substitution compatible to that seen in this mild form of androgen resistance. Hence, subtle undervirilization (phallic length, 5 to 6 cm in two patients) with or without gynecomastia, rather than infertility, may represent the extreme of the phenotypic spectrum of subtle defects in androgen receptor function.

Defects in Testosterone Metabolism by Peripheral Tissues: Steroid 5-Reductase Type 2 Deficiency (Male Pseudohermaphroditism with Virilization at Puberty)

In 1961, Nowakowski and Lenzi described a familial type of male pseudohermaphroditism, which they called "pseudovaginal perineoscrotal hypospadias," that was transmitted as an autosomal recessive trait. The patients resemble those with other forms of male pseudohermaphroditism by having a 46,XY karyotype, normally differentiated testes, male internal ducts, and ambiguous external genitalia. At puberty, striking but selective signs of masculinization appear.

In 1974, Walsh and associates and Imperato-McGinley and colleagues reported a defect in the conversion of testosterone to its 5-reduced metabolite DHT in patients with this syndrome. Imperato-McGinley described a genetic isolate from villages in the southwestern part of the Dominican Republic. The classic clinical features of this form of male pseudohermaphroditism in infancy include a clitoris-like, hypospadic phallicus bound in chordee of variable degree, a bifid scrotum, and a urogenital sinus that opens on the perineum. A blind vaginal pouch opens either into the urogenital sinus or onto the perineum behind the
and seminal vesicle) are well differentiated; the ejaculatory ducts usually terminate in the blind vaginal pouch. If a vaginal pouch is not present, the Wolffian ducts terminate on the perineum next to the urethra. The prostate is hypoplastic. At puberty, plasma testosterone levels increase into the adult male range, whereas DHT levels remain disproportionately low but measurable. Affected males virilize to a variable degree without gynecomasia; the voice deepens, muscle mass increases; the phallus, although bound in chordee of variable severity, enlarges to 4 to 8 cm in length; libido ensues; and penile erections occur. The bifid scrotum becomes rugated and pigmented, and the testes enlarge and descend into the labioscrotal folds. However, none of the postpubertal affected males have acne, more than sparse facial or body hair, temporal hair recession, or enlargement of the prostate, nor do they develop gynecomasia. Histologic examination of the adult testes in affected males shows Leydig cell hyperplasia and decreased spermatogenesis. In general, spermatogenesis is either absent or profoundly impaired, which appears to be due to the cryptorchidism. Three male siblings in a Swedish family affected with 5-reductase-2 deficiency had successful hypospadias repair in infancy and two brothers were demonstrably fertile. All were compound heterozygotes for mutations in the SRD5A2 gene.

As with other forms of male pseudohermaphroditism, phenotypic variability has been described both within and between cohorts. Approximately 55% of patients have had a pseudovagina; the rest have a urogenital sinus, a hypospadiac phallus, or even a microphallus with a penile urethra. Eighteen of 19 46,XY patients with 5-reductase-2 deficiency from the Dominican cohort who were raised "unambiguously" as females changed their gender identity and gender role behavior to male after the onset of puberty. Similar observations were reported in 19 of 40 families with 5-reductase-2 deficiency, as well as in male pseudohermaphrodites with 17-HSD3 deficiency or XO/XY mosaicism. This phenomenon appears to be particularly prevalent, although not exclusively so, in 46,XY male pseudohermaphrodites whose gonads are retained and who produce testosterone that results in masculinization at puberty. An isolated case of 5-reductase deficiency was reported where the child was raised as a girl but on virilization at puberty the gender was changed to male. The parents were first cousins of Pakistani origin. These patients raise provocative questions about the effect of sex of rearing, social and cultural factors, and of learning and prenatal androgen exposure on the brain and Y-bearing genes on psychosexual development. A genetic isolate of individuals from New Guinea deficient in 5-reductase-2 was described by Herdt and Davidson. As in the Dominican cohort, a third category of sex was identified in this cultural isolate, the so-called Turnim man. However, sex reversal at puberty has not been as common in this cultural isolate as in others, where the sex reversal appears to be in part a consequence of social and cultural pressure rather than solely a sex hormone effect on behavior. Another cluster of cases has been described from Southern Lebanon. Again, affected individuals were unambiguously female at birth and remained so until puberty, when there was a change in gender role.

Females homozygous for 5-reductase-2 deficiency are phenotypically normal and undergo normal pubertal maturation except for delayed menarche. They have an absence of hair on the arms and legs and decreased axillary and pubic hair, which suggests an important effect of DHT on the growth of body hair. Fertility is normal, and two of the three homozygous females studied in the Dominican kindred gave birth to nonidentical twins, which suggests a role for DHT in the regulation of ovarian follicular maturation; the human ovary has 5-reductase-2 activity.

In infancy and childhood, patients with 5-reductase-2 deficiency have normal to elevated plasma concentrations of testosterone and decreased levels of DHT after the administration of hCG. In affected postpubertal patients, the testosterone/DHT ratio in peripheral blood is increased from 12 ± 3.1 (mean ± SD) to 35 to 84. Postpubertally, plasma LH levels are either normal or slightly elevated; plasma FSH levels are elevated in about 50% of patients. Studies of estrogen and androgen synthesis demonstrate normal male androgen and estrogen production; this explains the lack of gynecomasia postpubertally in these patients, compared with patients with partial androgen resistance. Additional biochemical features of 5-reductase-2 deficiency are a diminished ratio of urinary 5 to 5-reduced C19 and C18 steroids. Analysis of 5-reduced C19-steroid urinary products can be a useful diagnostic test if the gonads have already been removed. Deficient or abnormal 5-reductase-2 activity in fibroblasts cultured from genital skin is typically seen, but the range of activity in normal individuals is very wide and varies between cell passage number. This is no longer used for diagnosis now that molecular studies of the SRD5A2 gene are possible. Heterozygotes for 5-reductase-2 deficiency have
no clinical manifestations and have intermediate ratios of urinary 5-reduced to 5-reduced C17-steroids (e.g., androsterone/eiocholanolone).  

Genetics.  

The disorder is transmitted as an autosomal recessive trait. There are two steroid 5-reductase enzymes. Both isoforms, which share 50% homology, and catalyze the conversion of androsterone to the more androgentic DHT. Steroid 5-reductase-1 is encoded by a gene located on chromosome 5p15.23; the gene, SRD5A1, has five exons and four introns and encodes a protein with a neutral to basic pH optimum. It is expressed at birth in the liver and nongenital skin. Although the expression of the type 1 enzyme persists in the liver throughout postnatal life, its expression decreases in skin to unmeasurable levels after 2 to 3 years of age and remains low until puberty, when it is again present in nongenital skin, especially the sebaceous glands of the scalp. No mutations in the type 1 gene have as yet been described. Targeted disruption of the SRD5A1 gene in mice causes a defect in parturition that is caused by impaired cervical ripening.  

Steroid 5-reductase-2 is encoded by the SRD5A2 gene on chromosome 2p23, which, like the type 1 gene, contains five exons and encodes a 254-amino acid protein. The type 2 isozyme has a lower Michaelis constant (Km) for testosterone than does the type 1 isozyme; steroid 5-reductase-2 has an acidic pH optimum and is more sensitive to inhibition by finasteride. Most of the 5-reductase activity in the early fetus is caused by the type 2 enzyme. It is expressed in the primordia of the prostate and external genitalia before their differentiation, but it is not expressed in the embryonic Wolffian duct until after the differentiation of the epithelium, vas deferens, and seminal vesicles, which are induced by testosterone and not by DHT. Similar observations were found in the human male reproductive tract based on immunohistochemistry.  

Type 2 expression increases in liver and nongenital skin at birth. It persists in the liver throughout life but diminishes to unmeasurable levels in nongenital skin after 3 years of age. A pseudogene has been mapped to the long arm of the X chromosome at band q24-qter.  

Steroid 5-reductase-2 deficiency is inherited as an autosomal recessive trait and is genetically heterogeneous and more than 30 mutations have been detected in the gene (Fig. 22-82). Most are missense mutations and distributed throughout the coding region of the gene. In general, mutations involving the 3' end of the gene and the carboxyl terminus of the enzyme are associated with a decrease in the affinity for the cofactor NADPH, whereas mutations that affect the binding of testosterone involve exons that encode either the NH2 or COOH-terminal ends of the enzyme. About 35% of the mutation-positive cases are compound heterozygotes and there are relative "hot spots" in exons 4 and 5.  

Consanguinity has been described in approximately 40% of patients. The occurrence of the disorder in three genetic isolates in the Dominican Republic (Arg246Trp), the Sambia tribe of the New Guinea highlands (deletion of the SRD5A2 gene), and cohorts in Turkey is probably the result of a "founder effect." The severity of undermasculinization does not appear to be closely related to the type of mutation, especially in nongenetic isolates. The two mutations reported to cause either 17HSD3 or 5-reductase-2 deficiency are very similar. Consequently, the report of mutations in both HSD17B3 and SRD5A2 in a large isolated Turkish kindred with male pseudohermaphroditism is noteworthy. The results of molecular analysis were complex, with some members homozygous for the SRD5A2 mutation but heterozygous for the HSD17B3 mutation, and vice versa for other affected family members. There was some phenotypic distinction as judged by the presence of mild gynecomastia if an individual was homozygous for the HSD17B3 mutation. There is evidence that 5-reductase-2 deficiency may result from uniparental disomy (UPD). This is based on a report of two unrelated patients with the enzyme deficiency whose parents were heterozygous carriers for two different but identical mutations in the two families (Glu197Asp and Pro212Arg). One patient was a compound heterozygote for the two mutations, but the other was a homozygote for the paternal mutation (Glu197Asp). The reduction to homozygosity for this mutation suggested not only the first example of 5-reductase deficiency resulting from UPD but also the first case of paternal (as opposed to maternal) UPD involving chromosome 2.  

The finding of two 5-reductase genes with different tissue distributions and different temporal expressions has clarified the nature of the pubertal masculinization in the human male. The increase in plasma DHT levels at puberty and the chronic effect of adult circulating DHT levels are responsible for the masculinization at puberty.  

The incidences of 5-reductase-2 deficiency are consistent with a frequency of at least 1 in 20,000 live male births. The finding of two 5-reductase genes with different tissue distributions and different temporal expressions has clarified the nature of the pubertal masculinization in the human male. The increase in plasma DHT levels at puberty and the chronic effect of adult circulating DHT levels are responsible for the masculinization at puberty.  

The diagnosis of 5-reductase-2 deficiency can be difficult, especially before the age of puberty. It should be suspected in all prepubertal male pseudohermaphrodites, especially those with perineoscrotal hypospadias with or without a blind vaginal pouch, in males with hypospadias or microphallus, and in male pseudohermaphrodites who virilize at puberty without evidence of gynecomastia. Virilization at puberty and the absence of gynecomastia in male pseudohermaphrodites are not unique to 5-reductase-2 deficiency. For example, patients with 17-HSD3 deficiency or partial androgen resistance may present in this manner, but can be distinguished biochemically or by DNA analysis from patients with 5-reductase-2 deficiency. The diagnosis of 5-reductase-2 deficiency can be confirmed prepubertally and postpubertally by demonstration of an abnormally high testosterone/DHT ratio in peripheral blood before and/or after hCG administration. The testosterone/DHT ratio under basal conditions in postpubertal affected males is 35 to 84, whereas the ratio in normal men is 12 ± 3.1. In normal male infants, when there is active testicular steroidogenesis, the testosterone/DHT ratio ranges from 1.7 to 17 (mean ± SD, 4.9 ± 2.9). In view of the low levels of testosterone and DHT in prepubertal males, it is usually necessary to administer hCG (1500 U/m2 intramuscularly every 24 hours three times) to demonstrate the defect. Normal infants, when there is active testicular steroidogenesis, the testosterone/DHT ratio ranges from 1.7 to 17 (mean ± SD, 4.9 ± 2.9). Using both basal and hCG-induced increases in testosterone level and urinary analyses of 5α- and 5β-hydroxy metabolites, Imperato-McGinley and co-workers detected 5-reductase-2 deficiency in three infants between the ages of 1 and 3 months. Mutations in the SRD5A2 gene have been reported rarely in boys with isolated hypospadias. Two of the mutations (Ala49Thr and Leu113Val) have not been reported in patients with typical 5-reductase deficiency. A positive family history of hypospadias is not necessarily a pointer to a mutation being more likely.  

Early diagnosis of 5-reductase-2 deficiency is important because of its bearing on the assignment of sex in the affected infant. Although the majority of missense mutations in the SRD5A2 gene are associated with less than 0.4% of normal activity, mutations with 3% to 15% residual activity have been reported in 46,XY individuals with sufficient masculinization of the external genitalia at birth to be assigned a male gender. Masculinization in utero and the plasma testosterone/DHT ratio in early infancy correlate with the degree of residual 5-reductase-2 activity. The natural history of patients with this deficiency is that, in the propensity for patients to change to male gender role behavior and for virilization at puberty makes male assignment of neonatally diagnosed patients the recommendation of choice, especially in affected individuals with ambiguous or hypoplastic male genitalia. Therapy with DHT should increase phallic length into the normal range for age and enable repair of hypospadias. Carpenter and co-workers described a 9-month-old infant with 5-reductase-2 deficiency who had been assigned a male gender at birth. The genitalia exhibited penoscrotal hypospadias with a phallus 1.9 cm in length and bound down in choordee. Therapy was instituted with DHT, 25 mg/day (2% by weight in a cold cream base), applied to the patient's abdomen. Four months of therapy resulted in an increase of stretched phallic length from 1.8 to 3.8 cm. No advancement in bone maturation was noted. Hypospadias repair was undertaken, and a second course of DHT was given without consequence. Two affected siblings with very small phallices and a bifid scrotum containing palpable
gonads were treated with topical DHT cream applied to the external genitalia. There was significant phallic growth in both siblings, which also aided subsequent hypospadias corrective surgery. In adults with 5-reductase-2 deficiency, supraphysiologic doses of testosterone have resulted in normal DHT levels and partial masculinization; the conversion to DHT is mediated by the type 1 5-reductase isozyme. However, Mendonca and colleagues[246] report that treatment of late adolescents or adults with high doses of testosterone and/or DHT induced an increase in phallic size that remained more than 2 SD below the normal mean and usually did not exceed 5 cm. Should the patient inadvertently be assigned a female sex role or the parents elect to rear the infant or child as a female, we suggest that female genitaloplasty and orchidectomy not be performed until the age of puberty and only with the consent of the adolescent as well as the parents.

### Dysgenetic Male Pseudohermaphroditism (Ambiguous Genitalia Resulting from Dysgenetic Gonads)

Ambiguous development of the genital ducts, urogenital sinus, and external genitalia occurs in patients with dysgenic gonads. They usually present with evidence of AMH deficiency as well as androgen deficiency and therefore have müllerian duct derivatives and ambiguous external genitalia. Mutations and deletions of any and all of the genes involved in the testes determination and differentiation cascade (see Fig. 22-19 and Table 22-4) have been implicated in the etiology of dysgenetic male pseudohermaphroditism. The differential diagnosis encompasses a spectrum of abnormalities of the Y chromosome as well as

**TABLE 22-36** -- Clinical Features of DAX1 Duplication in 46,XY Males

<table>
<thead>
<tr>
<th>Karyotype:</th>
<th>46,XY dup Xp21</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inheritance:</td>
<td>X-linked duplication of DAX1 gene</td>
</tr>
<tr>
<td>Geneticia:</td>
<td>Female ambiguous (rare)</td>
</tr>
<tr>
<td>Wolffian duct derivatives:</td>
<td>Absent hypoplastic (rare)</td>
</tr>
<tr>
<td>Müllerian duct derivatives:</td>
<td>Normal hypoplastic (rare)</td>
</tr>
<tr>
<td>Gonads:</td>
<td>Ovaries hypoplastic testes (rare)</td>
</tr>
<tr>
<td>Habitus:</td>
<td>No somatic abnormalities associated with DAX1. Duplications including segments contiguous to DAX1 may be associated with delayed psychomotor development and growth and dysplastic facies</td>
</tr>
<tr>
<td>Hormone profile:</td>
<td>Consistent with functional integrity of gonad</td>
</tr>
</tbody>
</table>

mosaicism involving a 45,X chromosome cell line and a cell line with Y chromosome DNA, which we have classified as "abnormalities of gonadal differentiation" (see sections on X-chromatin negative variants of the syndrome of gonadal dysgenesis and familial and sporadic XY gonadal dysgenesis). These patients, all of whom have in common a defect in testes differentiation, can present with the clinical syndrome of "dysgenetic male pseudohermaphroditism"[1378] (Fig. 22-83) .

Certain abnormalities of the X chromosome or an autosome are associated with dysgenetic male pseudohermaphroditism (see Fig. 22-19 and Table 22-4 ). Duplications of the Xp21.3 region that contains the DAX1 gene can cause dysgenic male pseudohermaphroditism as well as other extragenital anomalies and mental retardation (Table 22-36). Mutations in DAX1 cause X-linked congenital adrenal hypoplasia and hypogonadotropic hypogonadism. DAX1 encodes a protein with three and one-half repeats of a motif that may be a DNA-binding domain. The gene is expressed in the ovaries, testes, hypothalamus, and pituitary gland and has a steroidogenic factor (SF1) response element in its promoter. Mutations in DAX1 in XY males have no apparent effect on either the differentiation of the testes or male differentiation of the external genitalia. DNA analysis of 46,XY phenotypic females has failed to detect an abnormality in the DAX1 gene. Deletion of DAX1 does not impair testicular determination and differentiation, but duplications of the DAX1 region impair testicular differentiation in 46,XY individuals. This observation has led to the suggestion that DAX1 may function as a repressor of male differentiation; the conversion to DHT is mediated by the type 1 5-reductase isozyme. However, DAX1 expression is also detected in other tissues, such as the brain, suggesting a more complex role for this gene.

46,XY dysgenetic male pseudohermaphroditism has been associated with mental retardation and thalassemia, the "ATRX syndrome." Mutations in a gene called XH2, located at Xq13.3, that encodes a DNA helicase are described in patients with an atypical form of this syndrome. Terminal deletions at chromosome 10q (10q26-qter) and at 9p24-pter are associated with dysgenetic male pseudohermaphroditism and dysmorphic features. The putative gene on the long arm of chromosome 10 autosomes involved in testes development has not been ascertained. However, accumulating evidence suggests that haploinsufficiency of a gene related to double sex in Drosophila, DMRT1, in chromosome 9p24.3 leads to defective testis differentiation. Less common is the association of dysgenetic male pseudohermaphroditism with

**TABLE 22-37** -- Clinical Features of Denys-Drash Syndrome in 46,XY Males

<table>
<thead>
<tr>
<th>Karyotype:</th>
<th>46,XY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inheritance:</td>
<td>Autosomal dominant: heterozygous mutations in exon 9 of the WT1 gene on 11p13</td>
</tr>
<tr>
<td>Geneticia:</td>
<td>Phenotypic female (40%) ambiguous genitalia hypoplastic male</td>
</tr>
<tr>
<td>Wolffian duct derivatives:</td>
<td>Absent hypoplastic</td>
</tr>
<tr>
<td>Müllerian duct derivatives:</td>
<td>Present hypoplastic</td>
</tr>
<tr>
<td>Gonads:</td>
<td>Streak dysgenetic testes</td>
</tr>
<tr>
<td>Other features:</td>
<td>Renal failure in first year of life due to focal or diffuse mesangial sclerosis. Wilms' tumor in the first decade of life (75%), gonadoblastoma (5%)</td>
</tr>
<tr>
<td>Hormone profile:</td>
<td>Elevated LH, FSH levels; decreased testosterone values</td>
</tr>
</tbody>
</table>

Anomalies of the urinary tract are common in patients with abnormalities of genital differentiation. Less common is the association of dysgenetic male pseudohermaphroditism with
congenital or early-onset renal disease (diffuse mesangial sclerosis) and the development of Wilms’ tumor in the first decade (i.e., the Denys-Drash syndrome) or the childhood onset of renal disease and gonadal tumors (the Frasier syndrome).

XY individuals with Denys-Drash syndrome usually present in the newborn period with ambiguous genitalia, although both normal male and normal female genitalia have been reported (Table 22-37). The karyotype is 46,XY, albeit affected 46,XX females with renal disease and normal genitalia have been reported. Gonadal development in males varies from streak gonads to dysgenetic testes. The differentiation of the müllerian ducts varies depending on the functional status of the Sertoli cells of the gonads. Diffuse mesangial sclerosis, leading to renal failure, is seen on renal biopsy. Wilms’ tumor occurs in the first decade of life, and 4% of patients with Denys-Drash syndrome develop a gonadoblastoma. Frasier and associates described a pair of 46,XY monozygotic twins with streak gonads and gonadoblastoma, one of whom developed renal failure: these patients appear to represent a component of the spectrum of the Denys-Drash syndrome (Table 22-38). DNA analysis of patients with Denys-Drash syndrome has revealed heterozygous mutations of the Wilms’ tumor suppressor gene (WT1) located on 11p13. The WT1 gene encodes a transcription factor with four CysCys/HisHis zinc fingers that is expressed in the fetal kidney, gonad, and genital ridge. Heterozygous mutations, mostly missense mutations, are most common in exon 9 of the WT1 gene, with Arg394Trp being the most frequent in patients with Denys-Drash syndrome. The majority of WT1 mutations in Denys-Drash syndrome are de novo and appear to act as dominant negative mutations.

The Frasier syndrome, a variant of the Denys-Drash syndrome, is characterized in XY individuals by XY complete gonadal dysgenesis, with streak gonads resulting in female external and internal genital structures, late-onset glomerulopathy, focal glomerular sclerosis with renal failure occurring in the second decade, and predisposition to the development of a gonadoblastoma rather than Wilms’ tumor. The syndrome is associated with a heterozygous mutation in the WT1 gene at the donor splice site of intron 9. DNA analysis of patients with Denys-Drash syndrome has revealed heterozygous mutations of the Wilms’ tumor suppressor gene (WT1) located on 11p13.

Habitus:

TABLE 22-38 — Clinical Features of Frasier Syndrome in 46,XY Males.

| Karyotype: | 46,XY |
| Inheritance: | Autosomal dominant: mutation in splice donor site in intron 9 of the WT1 gene on 11p13 affecting the inclusion of 3 amino acids (KTS) and thus altering the normal balance of ± KTS isoforms. |
| Genitalia: | Female (hypoplastic male with cryptorchidism, rare) |
| Wolffian duct derivatives: | Absent (hypoplastic, rare) |
| Müllerian duct derivatives: | Present hypoplastic |
| Gonads: | Streak dysgenetic hypoplastic testes |
| Other features: | Late-onset renal disease from focal and segmental sclerosis of the kidney; increased incidence of gonadal tumors, especially gonadoblastoma; Wilms’ tumor rare (4%) |
| Hormone profile: | Elevated plasma LH, FSH levels |

LH, luteinizing hormone; FSH, follicle-stimulating hormone.

*Frasier’s syndrome and Denys-Drash syndrome may be variants of the same underlying syndrome.

as interrelated disorders of the WT1 gene and two extremes of a spectrum of clinical features rather than separate disease entities. (See section on sex differentiation.)

The genitourinary anomalies in the WAGR syndrome include renal agenesis, horseshoe kidney, urethral atresia, hypospadias, and cryptorchidism, and they are usually less severe than those observed in the Denys-Drash syndrome. (A mutation or deletion of WT1 has not been found in SR-Y-positive patients with dysgenetic gonads who do not have evidence of renal disease.)

Heterozygous mutations of the autosomal SOX9 gene cause campomelic dysplasia, often a lethal skeletal malformation, in which three fourths of affected 46,XY patients have dysgenetic male pseudohermaphroditism (Table 22-38; see also Fig. 22-17).

TABLE 22-39 — Clinical Features of SOX9 Deficiency in 46,XY Males

| Karyotype: | 46,XY |
| Inheritance: | Autosomal dominant: heterozygous loss of function mutations in the SOX9 coding region on chromosome 17q24.317q25.1 or break points with translocation 50 kb or more 5’ to the SOX9 gene |
| Genitalia: | Female ambiguous (70% of XY) normal male |
| Wolffian duct derivatives: | Absent hypoplastic present |
| Müllerian duct derivatives: | Normal hypoplastic absent |
| Gonads: | Testes dysgenetic testes ovotestes ovaries (rare) |
| Habitus: | Campomelic dysplasia, usually lethal bony dysplasia associated with male-to-female sex reversal in two thirds of affected males; testes in one third with male external genitalia |
| Prominent features of campomelic dysplasia: bowing of femora and tibiae, hypoplastic scapulae, 11 pairs of ribs, pelvic malformations, clubfeet, cleft palate, micrognathia, etc. |

disorder has an incidence of 0.05 to 1.6 per 10,000 live births. Manifestations include bowed long bones, hypoplastic scapula, a deformed pelvis, 11 pairs of ribs, a small thoracic cage, cleft palate, macrocephaly, micrognathia, hypertelorism, and a variety of cardiac and renal defects. Death from respiratory distress usually occurs in the neonatal period, but long-term survival has been reported. The external genitalia of affected 46,XY males varies from that of normal males with descended testes through ambiguous genitalia to female external genitalia, depending on the functional status of the fetal gonads. Affected 46,XX females have normal external genitalia and apparently normal ovaries. Histologic examination of the gonads from 46,XY patients with ambiguous or female external genitalia showed varying degrees of testicular dysgenesis extending to streak gonads with primordial follicles.

Tommenup and co-workers mapped the sex-reversal locus associated with campomelic dysplasia to 17q24.3q25.1 from studies of three patients with balanced de novo reciprocal translocations. The break point in these patients was distal to the growth hormone locus and proximal to the thymidine locus on 17q. Because
the murine gene. Sox9 had been localized to a region in the murine genome that is homologous to 17q and this gene is expressed in skeletal tissue, the corresponding human gene Sox9 was considered to be a candidate for campomelic dysplasia. Subsequently, missense, nonsense, frameshift, and splice junction mutations were detected in the Sox9 gene in patients with campomelic dysplasia with or without gonadal dysgenesis. However, no correlation between the mutations and the gonadal phenotype (sex reversal) has been found. In one family, the same Sox9 mutation resulted in siblings with campomelic dysplasia as a result of a germ-cell mosaicism for a Sox9 mutation in the father. However, the gonadal phenotype varied in the two 46,XY males from dysgenetic gonads to "normal" ovaries, and the affected 46,XX female had "normal" ovaries. In all patients studied, the mutation has been identified in only one SOX9 allele (heterozygous), which suggests that both the campomelic dysplasia and sex reversal are caused by haplosufficiency of the SOX9 gene. The absence of sex reversal in approximately one fourth of 46,XY individuals with campomelic dysplasia and in patients with translocations that involve break points in 17q more than 130 kb from the SOX9 gene in which no mutations have been found is unexplained.

The SOX9 gene has three exons and two introns, the first of the SOX genes to have introns, and encodes a 509-residue protein that localizes to the nucleus and contains an HMG box with 71% homology to that of the SRY protein. The HMG box binds to the same DNA motif CAA-CACAAGC as other HMG transcription factors and trans-activates transcription of a downstream target gene or genes. The trans-activation function of Sox9 appears to reside in the carboxy-terminal domain of the protein. Sox9 is expressed in the developing gonad, rete testis, and seminiferous tubule and in the mesenchyme that gives rise to skeletal tissue. During choriongenesis, Sox9 is co-expressed with COL2A1, the gene that encodes type II collagen; the SOX9 protein binds to regulatory gene sequences in the COL2A1 gene and regulates its expression. It is apparent from the study of patients with campomelic dysplasia that SOX9 is an integral part of testicular development cascade.

Steroidogenic factor 1 (SF1, or Ad4BP, adrenal 4 binding protein), a zinc finger "orphan" nuclear receptor, is a member of the steroid hormone/thyroid hormone receptor superfamily of transcription regulatory factors. SF1 encoding a gene on chromosome 9q33 and binds to a DNA motif consisting of an estrogen receptor half-site, AGGTCA, and to nucleotides S to this half-site (see Fig. 21-1A). The presence of this motif in the promoters of CYP steroid hydroxylase genes, as well as in vivo expression studies, suggests that SF1 is a key regulator of CYP steroidogenic enzymes in the adrenals and gonads. However, in the placenta the expression of CYP1A1 (P450umc) and other CYP steroidogenic genes occurs in the absence of expression of the SF1 gene, as is the case in the central nervous system where neurosteroids are produced locally. The SF1 gene is critical to the in vivo expression of AMH and the subunit of LH.

SF1 and the embryonal long terminal repeat binding protein (ELP, a protein that suppresses expression of Moloney murine leukemia virus in mouse undifferentiated embryonal carcinoma cells) are isoforms transcribed from the same gene by alternative promoter usage and splicing. SF1 and ELP are the mouse homologue of the human genes SF1 and ELP, which are co-expressed in skeletal muscle, reproductive organs, and gonads. All male and female knockout mice die of presumed adrenal insufficiency in the neonatal period. The external genitalia are female in both XX and XY mice, müllerian duct derivatives are normally developed, and the ventromedial nucleus of the hypothalamus is aplastic or hypoplastic. The expression of SF1 in the developing gonad is sexually dimorphic. At 12.5 days of embryonic development, when the bipotential gonad develops into an ovary or testis in the mouse and before expression of the CYP steroidogenic genes, SF1 expression persists in the Leydig and Sertoli cells of the testes but is extinguished in the primordial ovary.

In SF1 knockout mice the genital ridges developed normally until 10.5 days after coitus and thereafter underwent apoptosis. Therefore, SF1 appears to play a critical role in the development of the adrenals, ovaries, testes, hypothalamus, and gonadotropes and in modulation of AMH and CYP steroidogenic enzymes. As discussed previously, SF1 deficiency in three humans has been shown to cause severe adrenal insufficiency with impairment of testicular determination. Whether it affects ovarian determination is still to be determined (Table 22-40). Patients with 46,XY syndrome and putative haploinsufficiency of DMRT1 also manifest male pseudohermaphroditism, as discussed previously (Table 22-41).

Vanishing Testes Syndrome (Embryonic Testicular Regression Syndrome)

Various terms (XY gonadal dysgenesis, XY gonadal agenesis, rudimentary testis syndrome, congenital anorchia) have been used to describe the spectrum of genital anomalies resulting from cessation of testicular function during the middle phase of male sex differentiation, between 8 and 14 weeks of gestation. We first used the term vanishing testes syndrome for this form of male pseudohermaphroditism in 1957 because the genitalia in these cases suggested that the testes functioned initially and then “vanished” (for obscure reasons) at some time during the process of male sex differentiation. These patients have a 46,XY karyotype. Gonadal elements are absent, and differentiation of the genital ducts, urogenital sinus, and external genitalia is variable. At one end of the spectrum is the group of 46,XY individuals with female external and internal genitalia in whom the deficiency of embryonic testicular function presumably occurred before 8 weeks of gestation. These individuals have either no gonads (46,XY agonalism) or streak gonads. Loss of function of the fetal testes at 8 to 10 weeks of gestation would lead to ambiguous genitalia and variable development of the genital ducts, from complete absence of both müllerian and Wolffian ducts to partial development of either, a constellation referred to by some as the XY gonadal agenesis syndrome. Loss of testicular function after the critical phase of male differentiation (12 to 14 weeks) results in anorchia, a syndrome characterized by normal male differentiation both internally and externally but no gonadal tissue. The presence of normal male genitalia and absence of müllerian duct derivatives implies that fetal testicular function was normal before its loss. Sporadic and familial forms of unilaterial and bilateral anorchia, including monozoic gonads concordant and discordant for anorchia, have been described. Fetal testicular insufficiency and incomplete regression of the fetal testes after 12 to 14 weeks would be expected to produce a syndrome similar to that described by Bergada and colleagues, that is, small, rudimentary testes with microphallic and male ejaculatory ducts.

The nature of the underlying defect, which in some cases leads to absence or regression of genital ducts as well as testes and in some cases other congenital

---

**TABLE 22-40 -- Clinical Features of SF1 Deficiency in 46,XY Males:**

<table>
<thead>
<tr>
<th>Karyotype:</th>
<th>46,XY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inheritance:</td>
<td>Autosomal dominant; de novo heterozygous Glys58Glu mutation of SF1 gene on chromosome 9q33</td>
</tr>
<tr>
<td>Autosomal recessive due to homozygous Arg52Gln mutation (homozygote normal)</td>
<td></td>
</tr>
<tr>
<td>Genitalia:</td>
<td>Female</td>
</tr>
<tr>
<td>Müllerian duct derivatives:</td>
<td>Normal</td>
</tr>
<tr>
<td>Gonads:</td>
<td>Absent</td>
</tr>
<tr>
<td>Wolffian duct derivatives:</td>
<td>Absent</td>
</tr>
<tr>
<td>Habitual:</td>
<td>No pubic or auxillary hair or breast development; increased pigmentation</td>
</tr>
<tr>
<td>Hormone and metabolic profile:</td>
<td>Absent adrenals; primary adrenal insufficiency in infancy. No sex steroid secretion.</td>
</tr>
<tr>
<td>GnRH:</td>
<td>In the XY phenotypic female with a heterozygous SF1 mutation, at age 10 years:</td>
</tr>
<tr>
<td>LH</td>
<td>1.2-8.8 mIU/mL</td>
</tr>
<tr>
<td>FSH</td>
<td>17.8-38 mIU/mL</td>
</tr>
<tr>
<td>No testosterone response to HCG</td>
<td></td>
</tr>
</tbody>
</table>

*One 46,XX female reported with Arg255Leu heterozygous mutation; adrenal insufficiency; normal female genitalia; normal ovaries.*
anomalies, is not known. Several sibships with multiple affected individuals have been described. Josso and Briand reported on two siblings, one of whom was a normally differentiated male with microphallus and anorches. The other sibling had a 46,XY karyotype but was raised as a female. She had a normal clitoris, fused labioscrotal folds, a single perineal opening that led into a urogenital sinus, and a vagina. At laparotomy, absent gonads with coexistent müllerian and wolfian structures were found. This patient's phenotype was compatible with a diagnosis of XY gonadal agenesis. Despite the absent gonads, the patients had distinct phenotypic differences in the internal and external genitalia. The coexistence of so-called XY gonadal agenesis and anorches in the same sibship suggests that the disorders are related and are caused by embryonic testicular regression occurring at different stages of male development in utero; the familial cases support the operation of a rare, mutant gene in at least some patients with this syndrome. All eight boys with bilateral congenital anorches confirmed at surgical exploration were SRY positive.

The diagnosis of "true" anorches (in contrast to the testicular regression syndrome with ambiguous external genitalia) can be suspected in normally differentiated males with bilateral cryptorchidism, elevated gonadotropin levels, and low plasma AMH and inhibin levels. It is infrequently familial. We have demonstrated a diaphasic childhood pattern of gonadotropin levels in anorches males similar to that seen in females with the syndrome of gonadal dysgenesis. In particular, plasma FSH levels are elevated in infancy, decrease into the normal range in middle childhood, and rise into the agonal range after age 9 to 10 years. LHRH-induced increases in plasma FSH and LH concentrations are elevated throughout infancy and childhood. Hence, the LHRH test may be helpful diagnostically in middle childhood, when basal gonadotropin levels are normal or near normal. It has been proposed that the finding of elevated plasma FSH levels in conjunction with lack of a plasma testosterone response to HCG (1500 U/m2 intramuscularly every 48 hours × 7) establishes the diagnosis of anorches and obviates the need for laparotomy. This approach has been called into question by the finding of testes at laparotomy in two prepubescent males in whom no testosterone response to HCG was elicited. Immunoassay of plasma AMH in infancy and childhood is a useful additional test to assess for the presence of a testis; the concentration of plasma AMH in anorches is very low, as is the determination in plasma of another Sertoli cell secretion, inhibin B. Indeed, the concentration of plasma inhibin B highly correlates with the testosterone response to HCG and may well substitute for the HCG test, and it is more cost effective. Furthermore, CT or MRI, ultrasonography, and laparoscopy are useful procedures for evaluation of the patient with suspected anorches. We have deferred laparoscopic exploration of males with presumed "true" anorches (phenotypic males with nonpalpable testes, elevated gonadotropin levels, and no rise in the plasma concentration of testosterone in response to HCG) until the time of insertion of prosthetic testes. The typical findings at laparoscopy are a nubbin of testicular tissue adfixed to spermatic vessels exiting the internal inguinal ring. No recognizable testicular elements are evident in the nubbin. The true vanishing testes syndrome typified by normal male genital development has been attributed to a prenatal torsion event. An interesting case study described antenatal ultrasound findings of a male fetus with a left hydrocele and a normal right testes but at birth the right testis was nonpalpable. At surgical exploration, there was torsion of the spermatic cord on the left side and a normal testis, whereas on the right side the spermatic cord ended in a nubbin of testicular tissue. No recognizable testicular elements are evident in the nubbin. The true vanishing testes syndrome was not bilateral because the left hydrocele had a protective effect on the vascularity during the antenatal testicular torsion. The vanishing testes syndrome can be unilateral, occurring after descent but before fixation of the tunica vaginalis to the scrotal wall.

---

**TABLE 22-41 -- Clinical Features of 9p- Syndrome in 46,XY Males**

<table>
<thead>
<tr>
<th>Karyotype:</th>
<th>46,XY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inheritance:</td>
<td>Heterozygous deletion of 9q24.3 9-pter (distal short arm). The 5' end of DMRT1 is within 30 kb of the break point defining the minimal deletion causing sex reversal</td>
</tr>
<tr>
<td>Genitalia:</td>
<td>Female ambiguous male (rare)</td>
</tr>
<tr>
<td>Wolffian duct derivatives:</td>
<td>Absent hypoplastic</td>
</tr>
<tr>
<td>Müllerian duct derivatives:</td>
<td>Present hypoplastic absent (rare)</td>
</tr>
<tr>
<td>Gonads:</td>
<td>Absent dysgenetic hypoplastic testes</td>
</tr>
<tr>
<td>Habit:</td>
<td>Short stature (variable), mental retardation, microcephaly, trigonencephaly</td>
</tr>
<tr>
<td>Hormone profile:</td>
<td>Hypergonadotropic hypogonadism</td>
</tr>
</tbody>
</table>

---

**AMH/MIS**, a 148-kd glycoprotein homodimer, is secreted by the Sertoli cells of the testes beginning with differentiation of the fetal seminiferous tubules and continuing until pubertal maturation. AMH is not secreted by the fetal ovary, but postnatally (see section on sex differentiation) it is expressed in the granulosa cells of antral and preantral ovarian follicles. AMH, a member of the TGF- superfamily of growth and differentiation factors, is processed intracellularly and secreted in its mature, bioactive form. AMH binds to the AMH type II serine/threonine kinase receptor located in the mesenchyme surrounding the müllerian ducts before 8 weeks of gestation (when the müllerian ducts respond to AMH), causing epithelial-mesenchymal interaction, apoptosis of the müllerian duct epithelium, and regression of the müllerian duct. Studies in the bovine freemartin and in transgenic mice overexpressing AMH indicate that AMH can cause regression of germ cells in the ovary and reorganization of the ovary into cord-like seminiferous tubules and can inhibit CYP19 activity. Five of 21 transgenic male mice that chronically expressed human AMH exhibited mammary gland development, arrested wolffian duct differentiation, and undescended testes. These observations suggest that high levels of AMH can impair Leydig cell function and steroidogenesis in the testes. AMH levels are measurable in the normal male plasma until pubertal maturation.

AMH is encoded by a 2.75-kb gene containing five exons in the region of chromosome 19p13.3. The gene has an upstream regulatory element, the estrogen response element, half-site AGGTCA type, to which SF1 binds to regulate AMH secretion.

The AMH/MIS receptor, a serine/threonine kinase with a single transmembrane domain, is a member of the family of type II receptors for TGF--related proteins (see Fig. 22-27). The AMH type II receptor binds ligand but requires the ubiquitous type I receptor for signal transduction.

The AMH II receptor is encoded by a gene that contains 11 exons. The AMH type II receptor binds ligand but requires the ubiquitous type I receptor for signal transduction.

The AMH II receptor is encoded by a gene that contains 11 exons. This receptor is expressed in adult granulosa and Sertoli cells, which suggests a possible autocrine action of AMH in these cells.

**Persistent Müllerian Duct Syndrome (Female Ducts in Otherwise Normal Men; Hernia Uteri Inguinalis)**

AMH/MIS, a 148-kd glycoprotein homodimer, is secreted by the Sertoli cells of the testes beginning with differentiation of the fetal seminiferous tubules and continuing until pubertal maturation. AMH is not secreted by the fetal ovary, but postnatally (see section on sex differentiation) it is expressed in the granulosa cells of antral and preantral ovarian follicles. AMH, a member of the TGF- superfamily of growth and differentiation factors, is processed intracellularly and secreted in its mature, bioactive form. AMH binds to the AMH type II serine/threonine kinase receptor located in the mesenchyme surrounding the müllerian ducts before 8 weeks of gestation (when the müllerian ducts respond to AMH), causing epithelial-mesenchymal interaction, apoptosis of the müllerian duct epithelium, and regression of the müllerian duct. Studies in the bovine freemartin and in transgenic mice overexpressing AMH indicate that AMH can cause regression of germ cells in the ovary and reorganization of the ovary into cord-like seminiferous tubules and can inhibit CYP19 activity. Five of 21 transgenic male mice that chronically expressed human AMH exhibited mammary gland development, arrested wolffian duct differentiation, and undescended testes. These observations suggest that high levels of AMH can impair Leydig cell function and steroidogenesis in the testes. AMH levels are measurable in the normal male plasma until pubertal maturation.

AMH is encoded by a 2.75-kb gene containing five exons in the region of chromosome 19p13.3. The gene has an upstream regulatory element, the estrogen response element, half-site AGGTCA type, to which SF1 binds to regulate AMH secretion.

The AMH/MIS receptor, a serine/threonine kinase with a single transmembrane domain, is a member of the family of type II receptors for TGF--related proteins (see Fig. 22-27). The AMH type II receptor binds ligand but requires the ubiquitous type I receptor for signal transduction.

The AMH II receptor is encoded by a gene that contains 11 exons. Exons 1 to 3 for the signal sequence and the extracellular domain of the AMH receptor, exon 4 for the transmembrane domain, and exons 5 through 11 for the intracellular serine/threonine domain. In addition to the mesenchyme surrounding the fetal müllerian ducts (but not the epithelial cells), this receptor is expressed in adult granulosa and Sertoli cells, which suggests a possible autocrine action of AMH in these cells.

**A Distinctive disorder, persistent müllerian duct syndrome (PMDS), has been described in which 46,XY men and boys have well-developed testes, normal male ducts and external genitalia, and müllerian duct derivatives (Fig. 22-84). The diagnosis often is not made until a fallopian tube and uterus are encountered in patients undergoing inguinal hernia repair, orchioepxy, or abdominal surgery. Because of the trend for early surgical repair of an inguinal hernia or undescended testes, more cases are detected in infancy. The two anatomic forms. In the more prevalent form, there is a hernia containing a partially descended or scrotal testis and the ispsilateral tube and uterus are in the hernia. In some instances the contralateral testes and tube are present in the hernial sac as well. The presence of transverse testicular ectopia should suggest PMDS.** In the second form, the uterus, tubes, and testes are in the pelvis.
PMDS is a heterogeneous condition that is inherited in a sex-limited autosomal recessive manner. Females homozygous for null mutations in the AMH gene have normal müllerian ducts, external genitalia, and ovarian function, including fertility. Therefore AMH does not appear to play a critical role in ovarian differentiation or formation.

The retention of müllerian structures in normally differentiated males can result from failure of the testes to synthesize or secrete AMH or from a defect in the response of the duct to AMH because of an AMH receptor defect or possibly an abnormality in the timing or secretion of the hormone. Mutations in the AMH gene can lead to absence of AMH as a cause of PMDS (Fig. 22-85). AMH mutations are most common in Mediterranean and Arab countries with high rates of consanguinity, and most familial mutations are homozygous; on the other hand, mutations in the gene encoding the AMH type II receptor (AMHR-II) are more common in France and Northern Europe and are often heterozygous mutations.

In an extensive study of 69 families with PMDS, 28 mutations in the AMH gene were detected in 31 families (see Fig. 22-85). Both homozygous and compound heterozygous mutations were found in affected families, including splicing, missense, nonsense, and deletion mutations affecting the whole gene but mostly involving exons 1 and 3 of exon 5. In 27 families, deleterious mutations in the type II AMH receptor were detected, including deletion, missense, and nonsense mutations (Fig. 22-86). The most common mutation was a 27-base pair deletion in exon 10, which was found on at least one allele in 10 out of 16 families and stands in contrast to the diverse mutations in the AMH gene. In 11 families (16% of the total number), no abnormality in AMH or AMH receptor genes was detected. Patients with PMDS caused by mutations of the AMH gene have low or undetectable levels of serum AMH; in contrast, AMH concentrations are high or elevated in patients with mutations of the AMH receptor.

Treatment of PMDS is directed toward attempting to ensure fertility in males, a difficult issue because of the anatomic findings. Testicular differentiation and function are normal in these patients, but an increased prevalence of testicular degeneration has been described, which is probably secondary to torsion of the testes. Anatomic abnormalities of the epididymis and the vas deferens are common. Infertility may result from late orchiopexy or from mechanical problems associated with entrapment of the vas deferens in the müllerian derivatives, which is usually present. Early orchiopexy, proximal salpingectomy (leaving the epididymis attached to the fimbriae of the fallopian tube), dissection of the vas deferens from the lateral walls of the uterus, and a complete hysterectomy are recommended as a useful surgical approach. Despite these recommendations, men with high pelvic testes rarely have successful orchiopexy, and many of these individuals are androgen deficient.

Progestagens and synthetic estrogens, alone or in combination, have been implicated as rare causes of male pseudohermaphroditism. Courrier and Jost demonstrated an androgenic effect on the male fetus induced by a synthetic progestagen, ethisterone. Neumann and colleagues observed that relatively high doses of progesterone or of synthetic progestagens impaired urethral groove fusion in fetal male rats. Aarskog reported on 130 patients with hypospadias who were studied retrospectively. A history of maternal ingestion of oral progestagens in early pregnancy was obtained in 11 cases. In 6, the agent was administered for threatened abortion, and in 5 the progestagen in combination with estrogens was given as a pregnancy test. Hypospadias occurred anywhere from the glans to the base of the penile shaft; the location correlated with the week of gestation in which therapy was initiated. Other studies have also suggested an association between progestagens and hypospadias, although this relationship has been questioned.

Aarskog postulated that maternal progestagens may inhibit testosterone synthesis by the fetal testes and impair the reduction of testosterone to DHT at the target tissue and thereby lead to failure of urethral groove fusion and hypospadias. Some progestagens can inhibit 5-reductase activity in vitro. Inhibition of this enzymatic activity at an early fetal stage (e.g., through placental transfer of drugs given to the mother) could impair masculinization of the male external genitalia. Alternatively, progestagens may bind to androgen receptors and impair androgen action.

Kaplan described male pseudohermaphroditism in a boy whose mother received large doses of diethylstilbestrol during early pregnancy. However, no additional reports of this association have appeared. Because of the report of Herbst linking maternal diethylstilbestrol therapy during pregnancy with vaginal and cervical adenocarcinoma in daughters, abnormalities in the genital tract have been sought in males. Increased incidences of meatal stenosis, epididymal cysts, hypospadias, and abnormal semen have been observed, but hypospadias has not been reported.

Environmental Chemicals

An increase in the prevalence of disorders of the development and function of the male reproductive system, especially hypospadias and cryptorchidism, and in some European countries a fall in the sperm count and a rise in cancer of the testis, has occurred during the past 50 years. Some investigators have speculated that the increase in reproductive abnormalities observed in human males is related to an increase in the exposure in utero to exogenous estrogenic chemicals, so-called environmental estrogens, in the maternal diet, either as a natural occurrence or as a result of chemical contamination. Administration of diethylstilbestrol or the putative environmental estrogen, 4-octylphenol, to pregnant rats resulted in decreased expression of CYP17 mRNA and protein in Leydig cells of XY male offspring. Suppression of CYP17 may play a role in the putative adverse effect of environmental estrogens on fetal masculinization. The
dichlorodiphenyltrichloroethane (DDT) metabolite \( p,p' \text{DDE} \) (1,1-dichloro2,2-bis-(\( p \)-chlorophenyl) ethylene), unlike DDT itself, has little ability to bind to the estrogen receptor, \footnote{Further studies on the levels and risks of natural and environmental estrogens and antiandrogens in humans are necessary before the putative increased prevalence of certain abnormalities of the reproductive tract can be attributed to these agents as well as their putative role in the pathogenesis of the testicular dysgenesis syndrome.} but it binds to the androgen receptor and inhibits androgen action in the developing urogenital tract of rodents. \footnote{Further studies on the levels and risks of natural and environmental estrogens and antiandrogens in humans are necessary before the putative increased prevalence of certain abnormalities of the reproductive tract can be attributed to these agents as well as their putative role in the pathogenesis of the testicular dysgenesis syndrome.}
Other Sexual Abnormalities in Males

Hypospadias

Hypospadias, which may be defined as incomplete fusion of the penile urethra, is one of the common congenital anomalies. It has an estimated incidence of 4 to 8 per 1000 male births. As noted previously, the rate appears to have doubled in some countries in the 1970s and 1980s. Analysis of the family histories of patients with hypospadias revealed an increase in the occurrence of hypospadias in males in the pedigrees. This finding suggested a multifactorial mode of inheritance in some instances; the cause in most instances is unknown. Aarskog carried out a careful prospective study of 100 consecutive patients with hypospadias without other somatic anomalies, most of which were referred from a surgery clinic. One patient was a genetic female with virilizing CAH, five had sex chromosome abnormalities, one had the incomplete form of 46,XY gonadal dysgenesis, and nine were from pregnancies in which the mother had taken synthetic prostaglandin agents during the first trimester. Thus, in 15% of the patients, a pathogenetic mechanism was found or suspected. Both maternal cocaine use and environmental estrogens and androgens have also been implicated in the development of hypospadias. Even though androgen receptor defects have been suggested to play a significant role in the origin of hypospadias, androgen receptor defects are a rare cause of isolated hypospadias.

Hypospadias is a feature of many malformation syndromes, such as the Opitz syndrome. The cardinal manifestations of this syndrome are widely spaced eyes and hypospadias. This disorder is genetically heterogeneous; an X-linked form involves a locus at Xp22, and an autosomal form involves a locus at 22q11.2. Hypospadias is also a feature of the hand-foot-genital syndrome in males. Limb anomalies include short first metacarpals, short distal phalanges of the thumbs, and a short great toe. A mutation in the HOXA13 gene was detected in a pedigree with this syndrome.

The mildest and most common form of hypospadias is glandular or coronal and occurs in about 85% of cases; with surgical repair, these boys, at least in middle childhood, are not at risk for the development of "gender-atypical" behavior. Even though rare cases have been reported of 5-reductase deficiency and androgen receptor defects with isolated, simple hypospadias, extensive endocrine and cytogenetic evaluation of the otherwise normal male with glandular hypospadias and no somatic anomalies is not warranted. More severe hypospadias with or without cryptorchidism and somatic anomalies is an indication for complete evaluation, including karyotype analysis.

hCG stimulation studies, and visualization of the genitourinary tract.

Cryptorchidism

Undescended testes, the most common urogenital abnormality in malformation syndromes, is associated with more than 40 syndromes. Although normal testes may fail to descend to the scrotum because of coincidental anatomic abnormalities, in many instances cryptorchidism is caused by a defective testis. Fetal pituitary gonadotropin deficiency, either partial or complete, may play a role in some instances of cryptorchidism as well as microphallus. Cryptorchidism and its management are considered in greater detail in Chapter 16 and in several reviews.

Ambiguous Genitalia in 46,XY Males with Multiple Anomalies

The presence of ambiguous genitalia is associated with many malformation syndromes. In malformation syndromes such as the Aarskog and Opitz syndromes, the genital anomaly is of diagnostic significance. Other reports of rare causes of male pseudohermaphroditism include a patient with a putative "biologically inactive" but immunologically reactive LH and a group of familial cases in which a defect was postulated in fetal Leydig cell maturation with inadequate fetal testosterone production and impaired differentiation of germinal elements. The latter patients had ambiguous genitalia at birth but normal utilization at puberty and may represent examples of SRD5A2 deficiency, 17-HSD 3 deficiency, or partial androgen resistance.

Other Sexual Abnormalities in Females

The association of congenital absence of the vagina with normal or absent müllerian structures has been recognized for more than 100 years and is usually known as the Mayer-Rokitansky-Küster-Hauser syndrome. Congenital absence of the vagina occurs in 1 in 5000 female births. It was the second most common cause of primary amenorrhea in a series of 538 patients reviewed by Ross and van de Wiel. The principal features of the syndrome are primary amenorrhea in 46,XX females with well-developed female secondary sexual characteristics, an absent or hypoplastic vagina, and müllerian derivatives that vary from a normal uterus to bicornuate cords to absence of the uterus. Ovarian function is usually normal, and patients exhibit cyclic gonadotropin secretion with ovulation. Renal and skeletal anomalies may be present. Hearing loss, both conductive and sensorineural, occurs in 25% of patients with the Mayer-Rokitansky-Küster-Hauser syndrome. Clitoromegaly is not a feature and that distinguishes it from the adrenal and nonadrenal forms of female pseudohermaphroditism. The 46,XX karyotype and normal plasma gonadal steroid values differentiate this disorder from androgen insensitivity. Familial aggregates of women with anomalous müllerian differentiation may be explained by multifactorial inheritance.

It has been suggested that patients with skeletal and renal anomalies should be considered as a separate group, the GRES (genital-renal-ear-skeletal) syndrome. The association of uterine anomalies with malformations of the extremities is well described. The association of absence of the uterus and the upper part of the vagina, renal anomalies, and cervical somite dysplasia (Klippel-Feil syndrome) has been called the MURCS association (müllerian duct aplasia, renal agenesis/ectopia, and cervical somite dysplasia). Ultrasonography and CT and MRI scans are useful for determining the presence of a uterus and its structure.
Hematocolpos is a preventable complication if surgical reconstruction is begun before puberty is advanced. If the vagina is too small for sexual intercourse, nonsurgical or surgical correction should be undertaken at an appropriate age. Vaginal lubrication, orgasm, and coitus have been reported to be satisfactory in adults who have had successful vaginal reconstruction.
MANAGEMENT OF PATIENTS EXHIBITING AMBISEXUAL DEVELOPMENT

The advances in the management of patients with intersexuality have undergone a dramatic change over the past 50 years. A major deficiency in promoting guidelines for the modern management of intersexuality is the lack of critical outcome data or the selective nature of the available data. A background for these changes is the dramatic shift in our society's attitude toward sexual identity has come out of the closet and discarded many formerly deeply embedded Victorian attitudes. Before 1953, when one of us became committed to the clinical, scientific, and management aspects of intersex, these patients were largely managed by surgeons and urologists. Some were paternalistic authoritarians who had little tolerance for anyone questioning their empirical decision. The gonad was a cardinal criterion for determination of sex of rearing. The mantra was "It's an anatomical anomaly, fix it or cut it out."

A sea change in this approach began with Lawson Wilkins and his associates at the Harriet Lane Home of Johns Hopkins Hospital in the early 1950s. This was the beginning of group decision making by an evolving team that initially was composed of Dr. Howard Jones, a gynecologic surgeon; Drs. Joan and John Hampson, child and adult psychiatrists; John Money, a psychologist newly arrived from Boston, Wilkins and his fellows Judson Van Wyk and Melvin M. Grumbach, George Clayton, and Alfred Bongiovanni, a young faculty member in the group and the director of the pediatric endocrine laboratory. Social workers provided important family care skills. After the departure of the Hampsons for the University of Washington in the late 1950s and Lawson Wilkins' death in 1964, Money's extended studies led him to propose the gender socialization hypothesis of gender identity later referred to by his disciples as the "optimal-gender policy." The notion was that "sex of assignment and rearing were consistently and conspicuously a more reliable prognosticator" of the gender identity of an intersex patient than the chromosomal sex, gonadal sex, hormonal sex, the sex of the internal genital organs, or the degree of ambiguity of the external genitalia. Money and his associates stressed the importance of a decision about the sex of assignment in infancy.

Beginning in 1959 with the report by Phoenix and Young of "masculinization" of the female guinea pig brain by prenatal administration of testosterone, mounting evidence, at first a trickle but later virtually a torrent of experimental and behavioral studies, indicated the effect of androgens on sex dimorphic behavior. Further concern had arisen about the recommendations that infants with micropenis but normally found external genitalia and testes be assigned a female sex. Furthermore, the Intersex Society of North America raised important issues about the management of intersexuality. The article by John Colapinto in Rolling Stone and his later book, As Nature Made Him: The Boy Who Was Raised as a Girl, served to accelerate the re-examination of the clinical care of the intersex patient. This reassessment was led by psychologists, psychiatrists, and pediatric endocrinologists, as well as a number of informed pediatric surgeons and urologists. The Winter 1998 (Vol. 9, No. 4) issue of the Journal of Clinical Ethics was devoted to "Intersexuality." In a historical perspective on hermaphroditism, Dreger has emphasized how ingrained are our attitudes about "gender normality" but it also exemplifies how much our attitudes have changed.

Considerations Governing Choice of Sex for Rearing

With early, carefully weighed assignment of sex for rearing and appropriate continued management with an emphasis on continuity of care, individuals with ambiguities of the genitalia have the potential to lead well-adjusted lives and ultimately a satisfactory sex life. To obtain this favorable result, it is incumbent on the physician to make a correct diagnosis as early as possible and to provide the parents with pros and cons of sex assignment so that they may acquire sufficient knowledge to arrive at an informed decision on the sex for rearing. Lucid, simple, comprehensive discussions with the parents, taking into account their anxieties, religious views, social mores, cultural factors, and level of understanding, are critical for an appropriate gender assignment. The detection of genital ambiguity in a newborn infant can be seen as an urgent neonatal psychosocial necessity, beginning with how the parents are informed about the genital ambiguity. Once the sex for rearing is assigned, the gender role is reinforced by the use of whatever appropriate surgical, hormonal, and psychological measures are indicated.

Studies of patients reared in a sex discordant with their chromosomal sex, gonadal sex, hormonal sex, and even external genital organs have shown that no one parameter is an infallible basis on which to assign sex for rearing. A large body of evidence supports the masculinizing effect of exposure of the female as well as the male fetus to androgens; however, the magnitude of this influence cannot be predicted with certainty in the individual case. In intersex patients the degree of masculinization of the external genitalia does not correlate strongly with that of the central nervous system. The physician should consider the modern surgical advances in genital repair, including repair of hypospadias, and the use of exogenous testosterone in the treatment of androgen-sensitive microphallicus. In some cultures, the social, cultural, and economic benefits of a male gender are more compelling than phallic adequacy and are a prevailing, if not the most important, factor in the parental decision on the sex of rearing (Table 22-42).

The hormonal sex expected at maturity and the increasing possibility of fertility with advances in assisted reproduction technology are of importance. With the exception of female pseudohermaphroditids and true hermaphroditids reared as females, ambiguities of the external genitalia are caused by lesions that severely compromise but do not eliminate the possibility of fertility in view of the advances of modern reproductive techniques. A major goal in intersex patients should be the possibility of achieving cosmetic and functionally normal genitalia by surgical and endocrinologic means. In considering a decision to recommend a male sex of rearing, emphasis should be placed on the size of the phallus and glans and its potential for growth. All phenotypic males with micropenis (stretched penis length < 2.5 cm at birth) should be given a trial of testosterone enanthate in oil, 25 to 50 mg intramuscularly monthly for three doses, to ascertain the potential of the phallus for further growth before a decision on sex of rearing is made. Failure of the phallus to lengthen significantly (mean response, 2.0 ± 0.6 cm) raises the possibility of inadequate growth of the phallus in later childhood and at puberty. The parents should be made aware of the difficulties in making a dogmatic recommendation with the limited data on outcome. Principles governing the differential diagnosis and the surgical, hormonal, and psychological management of patients with genital ambiguity are treated more extensively in the following sections.

| TABLE 22-42 — Management of Ambiguous Genitalia |
| Get help! Team approach: pediatrician, endocrinologist, child mental health expert, social worker, and surgeon. |
| Arrive at a (prompt) definitive diagnosis if possible |
| Inform the parents of your diagnosis: the natural history of the disorder, prognosis, and the therapeutic options. Full disclosure! |
| Consider the parents’ level of understanding, cultural background, and religious views in order to allow them to come to a decision on the "sex" of their child and to provide truly "informed consent." |

Larsen: Williams Textbook of Endocrinology, 10th ed., Copyright © 2003 Elsevier
Abnormalities of sex differentiation should be suspected not only in infants with ambiguous genitalia (Fig. 22-87) but also in apparent females with inguinal masses, inguinal hernias, or slight clitoral enlargement. Apparent males with cryptorchidism, hypospadias, or unusually small genitalia or gonads likewise deserve close scrutiny. Sufficient investigation should be carried out in the newborn period to ascertain, if at all possible, an etiologic diagnosis to permit an informal decision by the parents on the assignment of sex. A karyotype or FISH analysis of sex chromosomes is an imperative first step in all such newborns (Fig. 22-88).
Infants with a 46,XX Karyotype

All infants with sexual ambiguity and a 46,XX karyotype should receive sufficient study in the neonatal period to differentiate the various forms of female pseudohermaphroditism (Table 22-18) from true hermaphroditism and the rare XX male.

Congenital Adrenal Hyperplasia

Female pseudohermaphrodites with virilizing CAH are reared as females in most cultures (Table 22-43). If female pseudohermaphroditism is secondary to CAH (primarily CYP21 deficiency), plasma levels of 17-hydroxyprogesterone and androstenedione and excretion of urinary 17-ketosteroids should be markedly elevated. A plasma 17-hydroxyprogesterone level higher than 90 nmol/L (3000 ng/dL) in an infant with ambiguous genitalia who is 24 hours of age or older is virtually diagnostic of 21-hydroxylase deficiency. However, premature and stressed infants may have elevated plasma 17-hydroxyprogesterone levels for 4 to 5 days. The diagnosis of CAH is sometimes difficult in the newborn period and may require multiple steroid determinations and the use of an intravenous bolus of ACTH to stimulate plasma 17-hydroxyprogesterone for the detection of 21-hydroxylase deficiency and 11-deoxycorticisol for 11-hydroxylase deficiency. Any infant with ambiguous external genitalia who fails to thrive or who develops vomiting, dehydration, and signs of hypoglycemia during the first few weeks of life should be suspected of having a severe salt-losing form of CAH. If such an infant has hyperkalemia associated with acidosis and hypocalcemia, the diagnosis is virtually assured, and vigorous therapy with hydrocortisone, salt, and mineralocorticoids should be instituted on an urgent basis to prevent collapse and sudden death. Once the diagnosis of adrenal hyperplasia is established, glucocorticoid therapy should be instituted and continued for life.

Other Forms of Female Pseudohermaphroditism

A small proportion of 46,XX female pseudohermaphrodites have aromatase (CYP19) deficiency. A history of virilization of the mother during pregnancy and elevated levels of androgens, androgen precursors, and FSH and unmeasurable levels of estrogen and estradiol are diagnostic. 46,XX infants may be presumed to have nonadrenal-induced female pseudohermaphroditism if adrenal hyperplasia, aromatase deficiency, and a glucocorticoid receptor defect have been excluded and if there is a reliable history of the mother receiving androgens or pregnancy. Patients with female pseudohermaphrodism usually have a normal uterus and fallopian tubes with ovaries in the normal location. For this reason, the diagnosis should be viewed with suspicion if there is an inguinal hernia or if gonad-like masses are palpable in the groin. Such masses are frequently testes. The presence of a uterus can often be detected in the newborn period by digital examination via the rectum; however, the use of pelvic ultrasonography is now routine because it is more informative and less stressful. If there is uncertainty, MRI of the pelvis is useful.

True Hermaphroditism

Most patients with true hermaphroditism have a 46,XX karyotype, and it may be difficult to distinguish some of them from patients with the rare nonadrenal causes of female pseudohermaphroditism. True hermaphrodites, however, often have gonads located in the labia or inguinal canals, a bipartite gonad is highly suggestive of ovotestes. In true hermaphroditism the assignment of sex should be deferred until the nature of the internal genital structures and gonads can be determined by ultrasonography, MRI, urethroscopy, or radiologic study with contrast media and, if necessary, laparoscopy. Most 46,XX true hermaphrodites are raised as females. Not infrequently, the heterologous gonadal tissue can be removed. It is important to emphasize the risk of malignant degeneration of the dysgenetic testicular tissue that is retained. In general, assignment of female sex and an attempt to preserve an ovary or ovarian tissue are appropriate except in those rare true hermaphrodites who have a 46,XY karyotype, no uterus, and adequate phallic development (Table 22-44).
Infants with a 46,XY Karyotype

Male pseudohermaphroditism is a heterogeneous group of disorders in which ambiguous genitalia result from either androgen deficiency or androgen resistance during the critical period of sex differentiation (see Fig. 22-88) (Table 22-45). The existence of both complete and partial defects in androgen biosynthesis, metabolism, and action; the occurrence of more than one gene encoding a protein with the same enzymatic activity but with different tissue specificities or developmental expression (e.g., 3-HSD, 17-HSD, 5-reductase); and the presence of different phenotypes in some patients with the same molecular genetic defect confound the clinical picture and the predictions about the natural history of these disorders.

A great effort should be made to establish an etiologic diagnosis, because this may have an important bearing on subsequent management, including family counseling. A detailed family history with construction of a pedigree is important, because many sexual abnormalities are hereditary and because this type of historical information is not always volunteered. For example, a history of aunts who have never menstruated or of an inguinal hernia or labial mass in a phenotypic female first-degree relative may suggest the diagnosis of androgen resistance. The mother should also be asked about signs of virilization during pregnancy and drugs or hormones that she may have taken during the early part of pregnancy.

Studies during the newborn period should always include an examination of the karyotype. A sufficient number of metaphase plates should be examined to reduce the possibility of overlooking mosaicism. The morphology of the sex chromosome and the autosomes should be determined and selective FISH analysis carried out.

Pelvic ultrasonography or MRI, radiographic contrast studies of the urogenital sinus, and fiberoptic endoscopic examination may aid in this initial evaluation. Laparoscopy or laparotomy is rarely indicated in the neonatal period except in infants suspected of having true hermaphroditism. Before an informal decision about sex assignment is made by the parents it is important to evaluate fully karyotypic studies, the pattern of plasma gonadal steroids before and after hCG stimulation, and other measures to identify a specific type of male pseudohermaphroditism.

Urinary steroids and plasma androgens should be measured before and after administration of corticotropin (0.15 to 0.25 mg intravenously) and hCG (1500 U/m² intramuscularly daily × 3 doses or every 48 hours × 7) to ascertain whether the patient has a block in testosterone synthesis or 5-reductase deficiency. The testosterone response to hCG may result in measurable phallic enlargement. Therefore, in addition to providing objective information about the functional capacity of the Leydig cells to secrete testosterone, the test may also provide evidence for the capacity of androgen-sensitive target tissues to respond to androgens. In the patient with male pseudohermaphroditism and no evidence of a testosterone biosynthetic error, 5-reductase-2 deficiency, or dyssgenetic male pseudohermaphroditism, clinical or biochemical evidence of testosterone responsiveness should be assessed before a sex assignment is made.

Male infants with congenital hypopituitarism or isolated hypogonadotropic hypogonadism frequently have micropenis and unilateral or bilateral cryptorchidism, and this diagnosis should be excluded by appropriate studies of pituitary function before considering sex reassignment. It is our view that all male infants with micropenis should be given a trial of testosterone parenterally before a conclusion is made that the phallus lacks the capacity for growth. Administration of a dose of 25 mg (100 mg/m²) of testosterone enanthate intramuscularly once a month for 3 months in the newborn should provide an adequate androgen stimulus to make this assessment. Rarely, this treatment may cause a slight advancement of the skeletal age, but this consideration is trivial when weighed against the momentous question of deciding the future sex of rearing. It is also important to assess penis size periodically and to repeat the course of testosterone therapy to maintain phallic size within the normal range for age. It has been suggested by inference from studies in rats that early exposure of the penis to androgens in childhood may result in a significant reduction in adult phallic size. However, data obtained by us and others suggest that early exposure to androgens does not cause a decrease in the developmentally programmed, final penile length.

In many patients, a precise etiologic diagnosis and appropriate sex assignment can be made on the basis of the criteria just stated. In the rare patient with no evidence of defective testosterone synthesis, end-organ unresponsiveness to androgen, or testicular dysgenesis, true hermaphroditism should be considered. In these patients the demonstration of both ovarian and testicular tissue at laparoscopy establishes the diagnosis. All of the findings (anatomic, karyotypic, genetic, and hormonal), along with the natural history of the specific disorder and the possibilities for surgical reconstruction and normal sexual function, need to be discussed with the parents. Their informed consent, understanding, and cooperation are critical to a successful gender assignment. Once the parental decision is made to rear the infant as a boy or as a girl, it is important to support the parents and patient in reinforcing this decision and to obtain mental health counseling. It is critical to assist the parents in addressing uncertainties and doubts that may arise during follow-up visits.

### TABLE 22-44 -- Management of True Hermaphrodite

| Most are raised as females with preservation of ovarian component. |
| When external genitalia are well masculinized and in the rare 46,XY variant, some parents select male sex of rearing. |

| TABLE 22-45 -- Management of 46,XY Male Pseudohermaphrodite |
| Administer testosterone enanthate 25 mg intramuscularly monthly × 3: "normal" response > 0.9 cm increase in phallic length |
| Raise 46,XY male pseudohermaphrodites as males except those with: |
| Complete androgen insensitivity syndrome. The dilemma presented by partial androgen insensitivity syndrome |
| Compelling reasons for sex assignment as female, including parents' informed decision |

**Larsen: Williams Textbook of Endocrinology, 10th ed. Copyright © 2003 Elsevier**
Reassignment of Sex After the Newborn Period

Children may be assigned an inappropriate sex because of errors in diagnosis or ignorance of the principles that should properly determine this choice. In such cases, the knotty decision to change the sex of rearing or to leave matters undisturbed depends largely on the age of the child and the degree to which the gender identity has been established. Money has stated that a change in the sex of rearing is feasible until the age of 18 months and is sometimes successful until 30 months, but thereafter, in our culture, serious and sometimes complex psychiatric and social consequences may be encountered. This concept has been challenged. Nevertheless, change of gender assignment in children should be undertaken after 18 months only after a review of alternatives and with the provision of close supervision and long-term counseling of the patient, parents, and siblings.

Before, at, or during adolescence, the patient may reach the decision that he or she has been reared in the wrong sex and may request assistance in changing his or her sex of assignment. If there are sufficient grounds for this belief, the request should be considered seriously and honored. Some patients may have serious psychological disturbances, and both psychiatric and legal counsel should be sought.
Reconstructive Surgery

Because the presence of ambiguous external genitalia is likely to reinforce doubt about the sexual identity of the infant or child, it is desirable to initiate reconstructive surgery as early as is medically and surgically feasible. The functional result, rather than the cosmetic, is paramount. It is highly desirable that surgery on the external genitalia be initiated before 6 months of age when practicable.

The management of clitoromegaly in female pseudohermaphrodites and in male pseudohermaphrodites reared as females has been controversial. Two different operative approaches have been used: clitoral recession, first reported by Lattimer in 1961 to replace the then widely used clitorectomy, and clitoroplasty. Clitoridectomy has long been abandoned as a mutilating procedure. Documentation of the role of the clitoris as an erotic organ in women makes it clear that clitoridectomy must be avoided. Clitoral recession has significant drawbacks mainly because of painful clitoral erections. Clitoroplasty, as recommended by Donahoe, Rink, Hutson, and Baskin, requires excision of the shaft and corpora with retention of the glans. This procedure and modifications of it are used most widely at present. Long-term data are still necessary to evaluate the efficacy of this procedure with respect to appearance and sexual function.

The extent of the initial repair of the urogenital sinus and vagina depends in large part on the skill and experience of the surgeon. These are not procedures that should be undertaken by surgeons or urologists who have not had training and experience with these techniques and are not part of a team of professionals that address clinical issues. Even when the initial repair has been done in the past by an experienced surgeon, it has not been uncommon for patients who have had vaginoplasties performed at age 18 months or earlier to require secondary operations because of stenosis of the introitus. We believe that reconstruction of a vagina in male pseudohermaphrodites reared as females and in female pseudohermaphrodites can be deferred until adolescence or until requested by the patient. A small vaginal pouch often can be enlarged by daily manipulations with a suitable mold. Even if the vagina remains too shallow for satisfactory coitus, manual dilatation makes it easier to carry out subsequent surgical correction.

<table>
<thead>
<tr>
<th>TABLE 22-46 -- Removal of the Gonads</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retain histologically normal and functional scrotal testes in 45,X/46,XY male pseudohermaphrodites.</td>
</tr>
<tr>
<td>Follow closely. Consider biopsy postpubertally to detect carcinoma in situ and perform periodic ultrasonography of testes.</td>
</tr>
<tr>
<td>Complete androgen insensitivity: testes may be retained until after puberty.</td>
</tr>
<tr>
<td>Partial androgen insensitivity or biosynthetic defects: if female sex of rearing is selected by parents, we recommend gonadal removal before puberty to prevent virilization.</td>
</tr>
</tbody>
</table>

A male with hypospadias no longer requires multiple procedures to create a phallic urethra. In recent years, new surgical techniques have reduced the number of operations and the length of hospitalization and have increased the success rate and parental expectations for "normality" in both the short and the long term. Circumcision should be avoided to preserve as much tissue as possible. Laparoscopy can be undertaken (if necessary) simultaneously with the initial operation. It is often desirable to insert prosthetic testes to give the scrotum dependency and to improve cosmetic appearance. These may be changed to adult-sized prostheses in adolescence.
Removal of the Gonads

A high incidence of gonadal tumors in patients with certain forms of gonadal dysgenesis and dysgenetic male pseudohermaphroditism especially makes it mandatory that an evaluation of this risk be given priority in deciding whether and when the gonads should be removed. Although the incidence of gonadoblastomas and germinomas (seminomas or dysgerminomas) increases near the normal time of adolescence, tumors are sometimes discovered during the first decade. Because temporizing serves no useful purpose and may expose the child to hormone secretions inappropriate to the chosen sex for rearing, it is advisable to proceed with gonadectomy concurrently with the initial repair of the external genitalia in patients who are at high risk, especially in the instance of intra-abdominal rudimentary testes. We are evaluating the use of MRI of the pelvis every 1 to 2 years to screen for gonadal neoplasms in children at risk.

Although prevalence of gonadal tumors has been reported to be as high as 9% in patients with the androgen resistance syndrome based on studies over three decades ago, some patients who developed tumors may have had atypical forms of gonadal dysgenesis; a modern survey is not available. The prevalence of gonadal malignancy before age 25 years in patients with androgen resistance appears to be relatively low. If the patient has a hernia and surgical repair is indicated, we recommend gonadectomy at that time to avoid a second operation. Otherwise, in the patient with complete androgen resistance, the undesended testes may be left in situ until after puberty. Thereafter, a frank and open discussion with the patient of the pathophysiology of androgen resistance and an assessment of the risk of gonadal malignancy needs to be undertaken to obtain informed consent for gonadectomy (Table 22-46).

There is a risk of some degree of virilization at the time of puberty in patients with the partial form of androgen resistance and in those with other forms of male pseudohermaphroditism with retained testes in whom a female sex for rearing has been assigned. In these cases, gonadectomy before puberty has been advanced and should be considered and discussed with the patient and parents. In some male pseudohermaphrodites who are raised as males, at least partial development of male secondary sexual characteristics will occur at the expected time of puberty. Provided the testes are not dysgenetic and are sufficiently descended to permit palpation, it is reasonable to leave the testes in situ. Such patients should be carefully examined at regular intervals for the presence of a tumor. MRI, ultrasonography, and examination of testicular biopsy specimens are useful in the early diagnosis of CIS and testicular neoplasm.

Hormone substitution therapy in hypogonadal patients should be prescribed in such a way that secondary sexual characteristics emerge appropriately in both timing and sequence. The goal of therapy should be to approximate normal adolescent development as closely as possible.

In females, including patients with the syndrome of gonadal dysgenesis, estrogenic hormone substitution therapy is initiated with low oral doses of estrogen (0.3 mg conjugated estrogens or 5 µg ethinyl estradiol daily) or a transdermal estradiol patch. Breast enlargement and growth of the uterus frequently occur within 3 months. Usually, cyclic therapy with estrogen and an oral progestagen is begun after 6 to 12 months of estrogen therapy or sooner if breakthrough bleeding occurs (see section on treatment of gonadal dysgenesis).

Development of male secondary sexual characteristics is usually better with repository injections of testosterone or a transdermal testosterone patch than with oral preparations. Few data are available on the use of dermal testosterone therapy in childhood or adolescence. Many oral synthetic androgens have the added disadvantage of predisposing to biliary stasis, jaundice, and hepatic tumors. Rapid virilization is usually inadvisable, and it is preferable to promote virilization gradually over many months in a manner similar to that in normal boys. The effect of gonadal steroids on skeletal maturation is dose related, whereas the effect on linear growth is less so. The relation between attained stature and skeletal maturation at the inception of therapy and the dose of androgen prescribed determines the ultimate effect of this therapy on adult height. An initial intramuscular dose of 50 mg of testosterone enanthate or other long-acting testosterone ester may be given monthly, beginning at age 12 to 13 years. Thereafter, the dose should be increased gradually over 3 to 4 years to the adult replacement dose of 200 mg every 2 weeks, usually reaching the adult level after a bone age of 17 years has been attained (see Table 22-45). In selected patients we have used the transdermal testosterone patch; initially the 2.5-mg patch is applied overnight for 8 hours. Subsequently the length of application is gradually increased to 24 hours.
Psychological Management

The newborn with ambiguous genitalia presents a clinical, social, and psychological challenge. Initially, in infants with ambiguous genitalia it is best for the physician to admit uncertainty regarding the "true sex" of the child and to urge the parents not to immediately assign a name and send out birth announcements. The filing of the birth certificate with the name should be delayed until a definite gender assignment and name has been given to the child. Clinical, cytogenetic, hormonal, and radiologic evaluation should be undertaken expeditiously. Thereafter, clearly presented, comprehensive, and informative discussions with the parents should ensue with all members of the team (i.e., endocrinologist, infant's physician, geneticist, surgeon, mental health specialist, social worker) present, if possible. This discussion should take into account the anxieties, religious views, social mores, cultural background, and level of understanding of the parents to make the best gender assignment for the infant and to obtain informed consent for the decision from the parents. A simple explanation of the normal process of sexual differentiation with appropriate illustrative material is useful because it lays the groundwork for the concept that all fetuses are bipotential initially and that sex differentiation is a complex process that may not be completed in utero. An analogy to other so-called birth defects (e.g., cleft lip, congenital heart disease) is accurate, easily understood, and less psychologically threatening. It should be stated clearly that the anatomic abnormalities can be surgically repaired by an experienced surgeon, that hormone replacement can be given if necessary, and that psychological support is available (Table 22-47). Continuing follow-up by the team members should address any questions that arise during infancy and childhood. In this age of "freedom of information," it is prudent to discuss in an age-appropriate manner and in progressive stages all aspects of the diagnosis, pathophysiology, management, and treatment of the ambiguous genitalia with the patient as soon as his or her level of increased understanding allows for this.

The implications of these observations for the management of intersex are debated in a series of published papers in the Journal of Clinical Ethics.

<table>
<thead>
<tr>
<th>TABLE 22-47 — Management of Ambiguous Genitalia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hormone therapy at puberty if necessary.</td>
</tr>
<tr>
<td>Progressive, step-by-step, age-appropriate discussion of diagnosis, pathophysiology, gender, and potential for fertility with the patient from childhood through adolescence, as well as with the parents.</td>
</tr>
<tr>
<td>Secrecy is unwarranted and counterproductive.</td>
</tr>
<tr>
<td>Involve the patient in decisions about surgery and sex hormone replacement therapy.</td>
</tr>
<tr>
<td>Provide continuing psychosocial and endocrinologic support to the patient and the family.</td>
</tr>
<tr>
<td>Long-term follow-up data on the outcome of modern management needs to be a high priority.</td>
</tr>
</tbody>
</table>

Bookmark URL: 563.html/top
References


References


Phelan JK, McCabe ERB. Mutations in NR0B1 (DAX1) and NR5A1 (SF1) responsible for adrenal hypoplasia congenita. Hum Mutat 2001; 18:472487.


Hastie ND. Life, sex, and WT1 isoformsthree amino acids can make all the difference. Cell 2001; 106:391394.


Hastie ND. Life, sex, and WT1 isoformsthree amino acids can make all the difference. Cell 2001; 106:391394.


Hastie ND. Life, sex, and WT1 isoforms: are amino acids the key to sex differentiation? J Biol Chem 2001; 276:1681716823.


Phelan JK, McCabe ERB. Mutations in NR0B1 (DAX1) and NR5A1 (SF1) responsible for adrenal hypoplasia congenita. Hum Mutat 2001; 18:472487.


Hastie ND. Life, sex, and WT1 isoformsthree amino acids can make all the difference. Cell 2001; 106:391394.

Hastie ND. Life, sex, and WT1 isoformsthree amino acids can make all the difference. Cell 2001; 106:391394.

Hastie ND. Life, sex, and WT1 isoformsthree amino acids can make all the difference. Cell 2001; 106:391394.

Hastie ND. Life, sex, and WT1 isoformsthree amino acids can make all the difference. Cell 2001; 106:391394.


Raymond CS, Murphy MW, O'Sullivan MG, Bardwell VJ, Zarkower D. Dmrt1, a gene related to worm and fly sexual regulators, is required for mammalian testis differentiation. Genes Dev 2000; 14:25872595.


References


References


534. Nielsen J, Pelsen B. Follow-up 20 years later of 34 Klinefelter males with karyotype 47,XXY and 16 hypogonadal males with karyotype 46,XY. Hum Genet 1987; 77:188192.


537. Ratcliffe SG. The sexual development of boys with the chromosome constitution 47,XXY (Klinefelter's syndrome). Clin Endocrinol Metab 1982; 11:703716.


References


References


90. Fujita H, Tanigawa Y, Yoshida Y, et al. Cytological findings of 10 cases of (X;i(X)) and one with (X;i(X)). Hum Genet 1977; 39:147155.


96. Turner C, Dennis NR, Skuse DH, Jacobs PA. Seven ring (X) chromosomes lacking the XIST locus, six with an unexpectedly mild phenotype. Hum Genet 2000; 106:93100.


985


985


References


Scheiflas F. Malignant potential of the dysgenetic gonad. Obstet Gynecol 1974; 44:298300 (Part I) and 455562 (Part II).


988


References


1037. Birnbaum MD, Rose LI. Late onset adrenocortical insufficiency was associated with menstrual dysfunction. Obstet Gynecol 1984; 63:445451.


1055. Birnbaum MD, Rose LI. Late-onset adrenocortical insufficiency was associated with menstrual dysfunction. Obstet Gynecol 1984; 63:445451.


1092. Sparks KS, Klsak I, Miller WL. Regional mapping of genes encoding human steroidogenic enzymes: P450ccc to 15q23q24; adrenodoxin to 11q22; adrenodoxin reductase to 17q24q25; and P450c17 to 10q24q25. DNA Cell Biol 1991; 10:359-365.


993


References


1245. Saaz JM, Morera AM, de Peretti E, Bertrand J. Further in vivo studies in male pseudohypothyroidism with gynecostasia due to a testicular 17-ketosteroid reductase defect (compared to a case of testicular feminization). J Clin Endocrinol Metab 1972; 34:59600.


<table>
<thead>
<tr>
<th>Reference</th>
</tr>
</thead>
</table>
115. Burrett TP, Guo W, McCabe ERB. The gene responsible for adrenal hypoplasia congenita, DAX-1, encodes a nuclear hormone receptor that defines a new class within the superfamily. Recent Prog Horm Res 1998; 51:241266.

999

144. Lala DS, Rice DA, Parker KL. Steroidogenic factor 1, a key regulator of steroidogenic enzyme expression, is the mouse homolog of fushi tarazu factor 1. Mol Endocrinol 1992; 6:12491258.
145. Sadowsky Y, Crawford PA, Woodson KG, et al. Mouse deficient in the orphan receptor steroidogenic factor 1 lack adrenal glands and gonads but express P450 side chain cleavage enzyme in the placenta and have normal embryonic serum levels of corticosteroids. Proc Natl Acad Sci USA 1990; 92:109310943.

83:28492854.

1460.
1458.
1457.
1456.
1455.
1454.
1453.
1452.
1451.
1450.
1449.
1448.
1447.
1446.
1445.
1444.
1443.
1442.
1441.
1440.
1439.


References

1496. Skakkebaek NE, Raigond-De Meys E, Main KM. Testicular dysgenesis syndrome: an increasingly common developmental disorder with environmental aspects. Hum Reprod 2001; 16:972978.


Lee PA. Should we change our approach to ambiguous genitalia? Endocrinologist 2001; 11:118-123.


Reiner W. To be male or femaleis the question. Arch Pediatr Adolesc Med 1997; 151:2242-2245.


Lee PA. Should we change our approach to ambiguous genitalia? Endocrinologist 2001; 11:118-123.


Chapter 23 - Normal and Aberrant Growth

Edward O. Reiter
Ron G. Rosenfeld

NORMAL GROWTH

ASSESSMENT OF GROWTH

Childhood is a time of growth, a process that is complex and involves the interaction of multiple, diverse factors—the "cumulative sum of millions of unsynchronized cell replications." Growth is common to all multicellular organisms and occurs by cell replication and enlargement along with the nonhomogeneous processes of cell and organ differentiation. The overall morphologic development, the rates of cellular division in different organ systems at different times, and the ultimate outcome are determined by the genetic composition of the individual interacting with external factors, including nutrition and psychosocial and economic factors.

The very nature of linear height growth, whether occurring as a continuous process or with periodic bursts of growth and arrest, has been hard to characterize definitively. During 1 year of growth monitoring, there may be marked seasonal variations of height and weight gain, with several monthly bursts of weight and then height growth. Some normal children may have a broad growth channel, with many showing diverse but characteristic tracks. Nonetheless, even though the process of growth is multifactorial and complex, children usually grow in a remarkably predictable manner. Deviation from such a normal pattern of growth can be the first manifestation of a wide variety of disease processes, including both endocrine and nonendocrine disorders and involving virtually any organ system of the body. Frequent and accurate assessment of growth is, therefore, of primary importance in the care of children.

Phases of Normal Growth

Growth occurs at differing rates during intrauterine life, early and middle childhood, and adolescence, prior to its cessation after fusion of long bone and vertebral epiphyseal growth plates. Prenatal growth averages 1.2 to 1.5 cm/week but varies dramatically (Fig. 23-1); midgestational length growth velocity of 2.5 cm/week falls to almost 0.5 cm/week immediately prior to birth. Growth velocity (Fig. 23-2 and Fig. 23-3) averages about 15 cm/year during the first 2 years of life and slows to about 6 cm/year during middle childhood. Pubertal growth begins earlier in girls than in boys but is 3 to 5 cm greater in magnitude in boys than in girls. The peak height velocity during the pubertal growth spurt is comparable to the rate of growth during the second year of life. The time of onset of the pubertal growth spurt varies in normal children, reflecting the concept of a tempo of growth or rate of maturation, as emphasized by Tanner and associates. In most normal children, the final height is not influenced by the chronologic time of the onset of the pubertal growth spurt, although the sex-related differences in adult height of approximately 13 cm are due to an earlier cessation of growth in females. Growth ceases when the skeleton achieves adult maturity.

Karlberg and associates have resolved the normal linear growth curve into three additive, partially superimposable phases. The components of this model include (1) an infancy phase, starting in midgestation and then rapidly decelerating to about 3 to 4 years of age; (2) a childhood phase, slowly decelerating during early adolescence; and (3) a sigmoid-shaped puberty phase that involves the adolescent growth spurt.

Hormonal concomitants of these phases have been suggested, but as seen in this chapter, the interplay of the growth hormone (GH)/insulin-like growth factor (IGF) axis, gonadal steroids, and thyroxine (T4) is complex, and an attempt to

![Figure 23-1 Rate of linear growth and weight gain in utero and during the first 40 weeks after birth. Length velocity is expressed in centimeters per week. The solid line depicts actual linear growth rate; the dashed line connecting the prenatal and postnatal length velocity lines depicts the theoretical curve for no uterine restriction late in gestation. The lighter dashed line depicts weight velocity. (Data from Tanner JM. Fetus into Man. Cambridge, Mass, Harvard University Press, 1978.)](https://example.com/figure23-1)

define individual predominance of one hormone at any time of life is likely an oversimplification.
Measurement

Assessment of growth requires accurate and reproducible determinations of height. Supine length is routinely measured in children younger than 2 years of age, and erect height is assessed in older children. The inherent inaccuracies involved in measuring length in infants are often obscured by the rapid skeletal growth during this period. For measurement of supine length, it is best to use a firm box with an inflexible board against which the head lies with a movable footboard on which the feet are placed perpendicular to the plane of the supine length of the infant. Optimally, the child should be relaxed, the legs should be fully extended, and the head should be positioned in the Frankfurt plane, with the line connecting the outer canthus of the eyes and the external auditory meatus perpendicular to the long axis of the trunk.

When children are old enough (and physically capable) to stand erect, it is best to employ a wall-mounted "Harpenden" stadiometer, similar to that designed by Tanner and Whitehouse for the British Harpenden Growth Study. Free-standing stadiometers are also available but require frequent recalibration. The traditional measuring device of a flexible arm mounted to a weight balance is notoriously unreliable and does not provide reliably accurate serial measurements.

As with length measurements in infants, positioning of the child in the stadiometer is critical. The child should be fully erect, with the head in the Frankfurt plane; the back of the head, thoracic spine, buttocks, and heels should touch the vertical axis of the stadiometer, and the heels should be together. Every effort should be made to correct discrepancies related to lordosis or scoliosis. Ideally, serial measurements should be made at the same time of day because standing height may undergo diurnal variation.

A trained individual, rather than an inexperienced member of the staff, should determine height. We recommend that lengths and heights be measured in triplicate, that variation is no more than 0.3 cm, and that the mean height is recorded. To determine height velocity, when several measurements are being made within a short period, the same individual should perform the determinations to eliminate interobserver variability. Even when every effort is made to obtain accurate height measurements, a minimum interval of 6 months is necessary for meaningful height velocity computation. Nine to 12 months' data are preferable so that errors of measurement are minimized and the seasonal variation in height velocity is assimilated into the data.
Growth Charts

Evaluation of a child's height must be done in the context of normal standards. Such standards can be either cross-sectional or longitudinal. Most pediatric endocrine clinics in the United States continue to use the cross-sectional data, provided by the National Center for Health Statistics (NCHS), originally introduced in 1977. Epidemiologic limitations exist in these growth charts. The original infant charts, for example, were derived from a private study of a group of subjects who were primarily white, formula-fed, middle-class infants from southwestern Ohio. Data employed for older children came from national health examination surveys conducted from 1963 to 1974.

The NCHS, now part of the Centers for Disease Control and Prevention (CDC), has recently provided a set of 16 new growth charts (8 each for boys and girls), representing revisions of 14 existing charts, and has introduced new charts for body mass index:

\[ \text{BMI} = \frac{\text{weight}}{\text{height}^2} \]

The charts show little change in average height over the last 25 years despite the perception that today's children are taller than those from three decades ago.

These charts compare individual children with the 5th, 10th, 25th, 50th, 75th, 90th, and 95th percentiles of normal children in the United States. There are, however, two major limitations of these charts when applied to the individual child.

First, they do not satisfactorily define children below the 5th or above the 95th percentiles, the very children in whom it is most critical to define the degree to which they deviate from the normal growth centiles. The NCHS data are useful in computing standard deviation scores (SDSs), which are more helpful, because a short child can be described as, for example, -4.2 or -2.5 SDS from normal. A height SDS for age is calculated as follows:

Because these are defined by cross-sectional data, however, childhood SDSs are not directly comparable with SDS during adolescence, when variation in growth rate and maturational tempo can be large.

Second, cross-sectional data are of greater value during infancy and childhood than in adolescence because differences in the timing of pubertal onset can considerably influence normal growth rates.

To address this issue, Tanner and colleagues developed longitudinal growth charts, combining longitudinal data to construct the curve shapes with centile widths obtained from a large cross-sectional survey, thus accounting for variability in the timing of puberty. Such charts are of particular value in assessing growth during adolescence and puberty and for plotting sequential growth data on any given child.

The data from cross-sectional and longitudinal growth studies have been employed to develop height velocity standards, enhancing the value of linear growth velocity measurements in an individual. Carefully documented height velocity data are invaluable in assessing the child with abnormalities of growth. Although there is considerable variability in the normal height velocity in children of different ages, between age 2 years and the onset of puberty, children
normally grow with remarkable fidelity relative to the normal growth curves. The physician should note any "crossing" of height percentiles during this age period, and abnormal height velocities always warrant further evaluation.

![Figure 23-7 Head circumference for-age and weight-for-length percentiles for boys (birth to 36 months). Developed by the National Center for Health Statistics in collaboration with the National Center for Chronic Disease Prevention and Health Promotion (2000).](http://www.cdc.gov/growthcharts)

Syndrome-specific growth curves have been developed for a number of clinical conditions associated with growth failure (e.g., Turner's syndrome, achondroplasia, Down's syndrome). Such growth profiles are invaluable for tracking the growth of children with these clinical conditions. Deviation of growth from the appropriate disease-related growth curve suggests the possibility of a second underlying cause.
Body Proportions

Many abnormal growth states, including both short stature and excessive stature, are characterized by disproportionate growth. The following determinations should be made as part of the evaluation of short stature:

1. Occipitofrontal head circumference.
2. Lower body segment: distance from top of pubic symphysis to the floor.
3. Upper body segment: sitting height (height of stool should be subtracted from standing height).

Published standards exist for these body proportion measurements, which must be evaluated relative to the patient’s age. The upper segment/lower segment ratio, for example, ranges from 1.7 in the neonate to slightly below 1.0 in the adult.
Skeletal Maturation

The growth potential in the tubular bones can be assessed by evaluation of the progression of ossification within the epiphyses. The ossification centers of the skeleton appear and progress in a predictable sequence in normal children, and skeletal maturation can be compared with normal age-related standards. This forms the basis of bone age or skeletal age, the only readily available quantitative determination of net somatic maturation and thus a mirror of the tempo of growth and maturation. It is not clear which factors determine this normal maturational pattern, but it is certain that genetic factors and multiple hormones, including T<sub>4</sub>, GH, and gonadal steroids, are involved.

Recent studies in patients with mutations of the gene for the estrogen receptor or for the aromatase enzyme have shown that estrogen is primarily responsible for epiphyseal fusion, although it seems unlikely that estrogen is solely responsible for all aspects of skeletal maturation.

After the neonatal period, a radiograph of the left hand and wrist is commonly used for comparison with the published standards of Greulich and Pyle. An alternative method for assessing bone age from radiographs of the left hand involves a scoring system for developmentally identified stages of each of 20 individual bones, a technique that has been adapted for computed assessment. The left hand is used because radiographs of the entire skeleton would be tedious and expensive and would involve additional radiation exposure. However, the hand does not contribute to height, and accurate evaluation of growth potential sometimes calls for radiographs of the legs and spine.

A number of important caveats concerning bone age must be considered. Experience in determination of bone age is essential to minimize intraobserver variance, and clinical studies involving bone age generally benefit from having a single reader perform all interpretations. The normal rate of skeletal maturation differs between boys and girls and among different ethnic groups. The standards of Greulich and Pyle are separable by sex but were developed in American white children between 1931 and 1942. Finally, both the Greulich and Pyle and the Tanner and Whitehouse standards involved normal children and may not be applicable to children with skeletal dysplasias, endocrine abnormalities, or other forms of growth retardation.
Prediction of Adult Height

The extent of skeletal maturation observed in an individual can be employed to predict the ultimate height potential. Such predictions are based on the observation that the more delayed the bone age (relative to chronologic age), the longer the time before epiphyseal fusion prevents further growth.

The most commonly used method for height prediction, based on Greulich and Pyle's *Radiographic Atlas of Skeletal Development,* [37]

---

was developed by Bayley and Pinneau [42] and relies on bone age, height, and a semiquantitative allowance for chronologic age (Table 23-1). The system of Tanner and colleagues [38, 43] employs height, bone age, chronologic age, height and bone age increments during puberty in the previous year, and menarchal status. Roche and associates [44] employ the combination of height, bone age, chronologic age, midparental height, and weight. Further, attempts have been made to calculate final height predictions without requiring the use of skeletal age [45, 46] by using multiple regression analyses with available data such as height, weight, birth measurements, and midparental stature. All of these systems are, by nature, empirical and are not absolute predictors. The more advanced the bone age, the greater the accuracy of the adult height prediction because a more advanced bone age places a patient closer to final height.

All methods of predicting adult height are based on data from normal children, and none has been documented to be accurate in children with growth abnormalities. For this kind of precision, it would be necessary to develop disease-specific (e.g., achondroplasia, Turner's syndrome) atlases of skeletal maturation.
Parental Target Height

Because genetic factors are important determinants of growth and height potential, it is useful to assess a patient's stature relative to that of siblings and parents. Tanner and associates developed a growth chart modifying the heights of children, aged 2 to 9 years, by the midparental height. Further, the child's predicted adult height (see earlier) may be related to a parental target height, or the mean parental height with the addition or subtraction of 6.5 cm for boys and girls, respectively. The two standard deviation (2 SD) range for this calculated parental target height is about ± 10 cm, so that calculated target heights, like predicted adult heights, are approximations.

Recent statistical reassessment shows a tendency for regression to the mean of the children's height as related to the midparental target height. Failure to realize this may lead to inappropriately using short parental height as an explanation for marked short stature in a child. Nevertheless, when a child's growth pattern clearly deviates from that of parents or siblings, the possibility of underlying pathology should be considered. Although it is certainly important to measure the heights of parents and siblings, rather than accept their statural claims, one must recall as well that it is not always possible to know the heights of the true biologic parents.
ENDOCRINE REGULATION OF GROWTH

The Pituitary Gland

The concept of the pituitary as a "master gland," controlling the endocrine activities of the body, has been replaced by recognition of the importance of the brain and, particularly, the hypothalamus in regulating hormonal production and secretion. Nevertheless, the pituitary gland is central to understanding the regulation of growth.

Embryologically, the pituitary gland is formed from two distinct sources: Rathke's pouch, a diverticulum of the primitive oral cavity (stomodeal ectoderm), gives rise to the adenohypophysis. The neurohypophysis (posterior pituitary) originates in the neural ectoderm of the floor of the forebrain, which also develops into the third ventricle. The adenohypophysis normally constitutes 80% of the weight of the pituitary and consists of anterior, intermediate, and infundibular lobes. In humans, the anterior lobe is the largest component and houses the most hormone-producing cells.

Rathke's pouch, the origin of the adenohypophysis, can be identified in the 3-mm embryo during the 3rd week of pregnancy. GH-producing cells can be found in the adenohypophysis by 9 weeks of gestation. Vascular connections between the anterior lobe of the pituitary and the hypothalamus develop about this time, although hormone production can occur in the pituitary in the absence of connections with the hypothalamus. Somatotrophs can frequently be demonstrated in the pituitary in anencephalic newborns. Nevertheless, the initiation of development of the anterior pituitary is probably dependent on responsiveness of the oral ectoderm to inducing factors from the ventral diencephalon and hypothalamus.

A complex orchestration of temporarily sequenced and geographically restricted expression of multiple extracellular signaling peptides and intracellular transcription factors regulates this developmental process. The developing pituitary gland and hypothalamus are in close anatomic juxtaposition, and their embryonic development is likely to be codependent. Some of the diencephalic factors that have been identified to be critical in formative and patterning of Rathke's pouch, which, in the mouse, is initiated on embryonic day 8 (e8), are bone morphogenic proteins 4 and 2 (BMP-4/2), Wnt5a, and fibroblast growth factor 8 (FGF-8). The dorsal neuroepithelial signal, BMP-4, is needed for "organ commitment" of the pituitary, whereas a BMP-2 (ventral) and FGF-8 (dorsal) gradient determines pituitary cell phenotypes [i.e., somatotrophs and the Pit-1-dependent lines, somatotrophs, lactotrophs, and thyrotrophs (ventral) and melanotrophs and corticotrophs (dorsal)]. It seems that reciprocal interaction of at least two transcription factors, Pit-1 and GATA-2, are important in implementing the cell-determination signals of BMP-2 and FGF-8.

Explant studies in the mouse have demonstrated that if Rathke's pouch is removed from the oral ectoderm on e10.5 and incubated in appropriate culture medium, differentiation of each of the pituitary cell types continues, indicating that by that point, organogenesis of the anterior pituitary is no longer dependent on hypothalamic signals, although such signals may remain critically involved in pituitary hormone production.

A number of pituitary-specific transcription factors are involved in the determination of pituitary cell lineages and cell-specific expression of anterior pituitary hormones. To date, defects in several homeodomain transcription factors shown to be involved in human anterior pituitary development and differentiation have now been associated with various combinations of pituitary hormone deficiencies (see Fig. 23-14 and Table 23-5). Because additional gene defects have been implicated in abnormal murine pituitary development, it seems likely that the number of human genetic defects will expand.

In the adult, the mean pituitary size is 13 × 9 × 6 mm. The mean weight is 600 mg (range, 400 to 900 mg), is slightly greater in women than in men, and increases during pregnancy. In the newborn, pituitary weight averages about 100 mg. Normally, the pituitary resides in the sella turcica, immediately above and partially surrounded by the sphenoid bone. The volume of the sella turcica is a good index of pituitary size and may be reduced in the child with pituitary hypoplasia. The adult mean pituitary size is 13 × 9 × 6 mm.
Growth Hormone

Chemistry

Human GH is produced as a single chain, 191amino acid, 22-kd protein (Fig. 23-16). It is not glycosylated, but it does contain two intramolecular disulfide bonds. GH shares sequence homology with prolactin, choriionic somatomammonio-tropin (CS) (placental lactogen), and a 22-kd GH variant (GH-V) secreted only by the placenta that differs from pituitary GH by 13 amino acids. The genes for these proteins have probably evolved from a common ancestral gene, even though the genes are located on different chromosomes (chromosome 6 for prolactin, chromosome 17 for GH). The genes for GH, prolactin, and placental lactogen share a common structural organization, with four introns separating five exons. In fact, the GH subfamily contains five members, whose genes are located on a 78-kb section of chromosome 17; the 5’ to 3’ order of the genes are GH, a CS pseudogene, CS-A, GH-V, and CS-B.

Normally, about 75% of GH produced by the pituitary is of the mature, 22-kd form. Alternative splicing of the second codon results in deletion of amino acids 32 to 46, yielding a 20-kd form, which normally accounts for 5% to 10% of pituitary GH. The remainder of pituitary GH includes desamidated and N-acetylated forms and various GH oligomers.

Secretion

The pulsatile pattern characteristic of GH secretion largely reflects the interplay of two hypothalamic regulatory peptides, growth hormonereleasing hormone (GHRH) and somatostatin (somatotropin resease-inhibiting factor [SRIF]), with presumed modulation by putative other GH-releasing factors. GHRH activity is species-specific, presumably reflecting the specificity of binding to a G protein-receptorvated receptor on the pituitary somatotrophs. Regulation of GH production by GHRH is mediated largely at the level of transcription and is enhanced by increases in intracellular cyclic adenosine monophosphate (cAMP) levels.

The GHRH receptor is a member of the G protein-coupled receptor family B-III, also called the secretin family, and has partial sequence identity with receptors for vasactive intestinal polypeptide, secretin, calcitonin, and parathyroid hormone.

In the dwarf transgenic mouse model with diminished GHRH production, pituitary somatotroph proliferation is markedly decreased. Mutations of the GHRH gene itself have not yet been reported, but anatomic and functional abnormalities of the connection between the hypothalamus and the anterior pituitary, which prevent interaction of GHRH with its receptor on the somatotroph, are the most important causes of clinical growth hormone deficiency (GHD).

The Gsh-1 homeobox gene, which is expressed in the developing central nervous system (CNS) but not in the pituitary, plays an important role, nonetheless, in mouse pituitary development. Mice with mutations in this gene do not produce GHRH and, presumably, do not produce gonadotropin-releasing hormone (GnRH) because anterior pituitary hypoplasia and deficiencies of GH, prolactin, and luteinizing hormone (LH) are present. The effects of this gene on hypothalamic releasing factors are analogous to those of the Pit-1 or PROPl genes (see later) at the pituitary level.

Instead of regulating GH synthesis, somatostatin appears to affect the timing and amplitude of pulsatile GH secretion. The pulsatile secretion of GH in vivo is believed to result from a simultaneous reduction in hypothalamic somatostatin release and increase in GHRH release. Conversely, a trough of GH secretion occurs when somatostatin is released in the face of diminished GHRH activity.

The regulation of the reciprocal secretion of GHRH and somatostatin is imperfectly understood. Multiple neurotransmitters and neuropeptides are involved in the regulation of the release of these hypothalamic factors.

Synthetic hexapeptides capable of stimulating GH secretion are termed GH secretagogues (GHSs). These peptides stimulate GH release and enhance the GH response to GHRH, although they work at receptors distinct from those for GHRH, at hypothalamic and pituitary sites. Finding 40% to 60% homology to the G protein-coupled GHS receptor in the pufferfish indicates that structure and function of this receptor have been highly conserved for about 400 million years, certainly suggesting a fundamental role for the natural ligand of this receptor.

Kojima and co-workers identified a putative endogenous ligand, a 28amino acid with the serine 3 residue N-octanoylated, referred to as ghrelin. It is found primarily in the stomach (and throughout the gastrointestinal tract) but also in...
TABLE 23-1 -- Prediction of Adult Stature: Fraction of Adult Height Attained at Each Bone Age

<table>
<thead>
<tr>
<th>Bone Age (years/months)</th>
<th>Girls Retarded</th>
<th>Girls Average</th>
<th>Girls Advanced</th>
<th>Boys Retarded</th>
<th>Boys Average</th>
<th>Boys Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>0.733</td>
<td>0.720</td>
<td>0.680</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>63</td>
<td>0.742</td>
<td>0.729</td>
<td>0.690</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>66</td>
<td>0.751</td>
<td>0.738</td>
<td>0.700</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>69</td>
<td>0.763</td>
<td>0.751</td>
<td>0.709</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>0.770</td>
<td>0.757</td>
<td>0.712</td>
<td>0.718</td>
<td>0.695</td>
<td>0.670</td>
</tr>
<tr>
<td>73</td>
<td>0.779</td>
<td>0.765</td>
<td>0.722</td>
<td>0.728</td>
<td>0.702</td>
<td>0.676</td>
</tr>
<tr>
<td>76</td>
<td>0.788</td>
<td>0.772</td>
<td>0.732</td>
<td>0.738</td>
<td>0.709</td>
<td>0.683</td>
</tr>
<tr>
<td>79</td>
<td>0.797</td>
<td>0.782</td>
<td>0.742</td>
<td>0.747</td>
<td>0.716</td>
<td>0.689</td>
</tr>
<tr>
<td>80</td>
<td>0.804</td>
<td>0.790</td>
<td>0.750</td>
<td>0.756</td>
<td>0.723</td>
<td>0.696</td>
</tr>
<tr>
<td>83</td>
<td>0.813</td>
<td>0.801</td>
<td>0.760</td>
<td>0.765</td>
<td>0.731</td>
<td>0.703</td>
</tr>
<tr>
<td>86</td>
<td>0.823</td>
<td>0.810</td>
<td>0.771</td>
<td>0.773</td>
<td>0.739</td>
<td>0.709</td>
</tr>
<tr>
<td>89</td>
<td>0.836</td>
<td>0.821</td>
<td>0.784</td>
<td>0.779</td>
<td>0.746</td>
<td>0.715</td>
</tr>
<tr>
<td>90</td>
<td>0.841</td>
<td>0.827</td>
<td>0.790</td>
<td>0.786</td>
<td>0.752</td>
<td>0.720</td>
</tr>
<tr>
<td>93</td>
<td>0.851</td>
<td>0.836</td>
<td>0.800</td>
<td>0.794</td>
<td>0.761</td>
<td>0.728</td>
</tr>
<tr>
<td>96</td>
<td>0.858</td>
<td>0.844</td>
<td>0.809</td>
<td>0.800</td>
<td>0.769</td>
<td>0.734</td>
</tr>
<tr>
<td>99</td>
<td>0.866</td>
<td>0.853</td>
<td>0.819</td>
<td>0.807</td>
<td>0.777</td>
<td>0.741</td>
</tr>
<tr>
<td>100</td>
<td>0.874</td>
<td>0.862</td>
<td>0.828</td>
<td>0.812</td>
<td>0.784</td>
<td>0.747</td>
</tr>
<tr>
<td>103</td>
<td>0.884</td>
<td>0.874</td>
<td>0.841</td>
<td>0.816</td>
<td>0.791</td>
<td>0.753</td>
</tr>
<tr>
<td>106</td>
<td>0.896</td>
<td>0.884</td>
<td>0.856</td>
<td>0.819</td>
<td>0.795</td>
<td>0.758</td>
</tr>
<tr>
<td>109</td>
<td>0.907</td>
<td>0.896</td>
<td>0.870</td>
<td>0.821</td>
<td>0.800</td>
<td>0.763</td>
</tr>
<tr>
<td>110</td>
<td>0.918</td>
<td>0.906</td>
<td>0.883</td>
<td>0.823</td>
<td>0.804</td>
<td>0.767</td>
</tr>
<tr>
<td>113</td>
<td>0.922</td>
<td>0.910</td>
<td>0.887</td>
<td>0.827</td>
<td>0.812</td>
<td>0.776</td>
</tr>
<tr>
<td>116</td>
<td>0.926</td>
<td>0.914</td>
<td>0.891</td>
<td>0.832</td>
<td>0.818</td>
<td>0.786</td>
</tr>
<tr>
<td>119</td>
<td>0.929</td>
<td>0.918</td>
<td>0.897</td>
<td>0.839</td>
<td>0.827</td>
<td>0.800</td>
</tr>
<tr>
<td>120</td>
<td>0.932</td>
<td>0.922</td>
<td>0.901</td>
<td>0.845</td>
<td>0.834</td>
<td>0.809</td>
</tr>
<tr>
<td>123</td>
<td>0.942</td>
<td>0.932</td>
<td>0.913</td>
<td>0.852</td>
<td>0.843</td>
<td>0.818</td>
</tr>
<tr>
<td>126</td>
<td>0.949</td>
<td>0.941</td>
<td>0.924</td>
<td>0.860</td>
<td>0.853</td>
<td>0.828</td>
</tr>
<tr>
<td>129</td>
<td>0.957</td>
<td>0.950</td>
<td>0.935</td>
<td>0.869</td>
<td>0.863</td>
<td>0.839</td>
</tr>
<tr>
<td>130</td>
<td>0.964</td>
<td>0.958</td>
<td>0.945</td>
<td>0.880</td>
<td>0.876</td>
<td>0.850</td>
</tr>
<tr>
<td>133</td>
<td>0.971</td>
<td>0.967</td>
<td>0.955</td>
<td>0.890</td>
<td>0.890</td>
<td>0.863</td>
</tr>
<tr>
<td>136</td>
<td>0.977</td>
<td>0.974</td>
<td>0.963</td>
<td>0.902</td>
<td>0.875</td>
<td></td>
</tr>
<tr>
<td>139</td>
<td>0.981</td>
<td>0.978</td>
<td>0.968</td>
<td>0.914</td>
<td>0.890</td>
<td></td>
</tr>
<tr>
<td>140</td>
<td>0.983</td>
<td>0.980</td>
<td>0.972</td>
<td>0.927</td>
<td>0.905</td>
<td></td>
</tr>
<tr>
<td>143</td>
<td>0.986</td>
<td>0.983</td>
<td>0.977</td>
<td>0.938</td>
<td>0.918</td>
<td></td>
</tr>
<tr>
<td>146</td>
<td>0.989</td>
<td>0.986</td>
<td>0.980</td>
<td>0.948</td>
<td>0.930</td>
<td></td>
</tr>
<tr>
<td>149</td>
<td>0.992</td>
<td>0.988</td>
<td>0.993</td>
<td>0.958</td>
<td>0.943</td>
<td></td>
</tr>
<tr>
<td>150</td>
<td>0.994</td>
<td>0.990</td>
<td>0.996</td>
<td>0.968</td>
<td>0.958</td>
<td></td>
</tr>
<tr>
<td>153</td>
<td>0.995</td>
<td>0.991</td>
<td>0.998</td>
<td>0.973</td>
<td>0.967</td>
<td></td>
</tr>
<tr>
<td>156</td>
<td>0.996</td>
<td>0.993</td>
<td>0.990</td>
<td>0.976</td>
<td>0.971</td>
<td></td>
</tr>
<tr>
<td>159</td>
<td>0.997</td>
<td>0.994</td>
<td>0.992</td>
<td>0.980</td>
<td>0.976</td>
<td></td>
</tr>
<tr>
<td>160</td>
<td>0.998</td>
<td>0.996</td>
<td>0.993</td>
<td>0.982</td>
<td>0.980</td>
<td></td>
</tr>
<tr>
<td>163</td>
<td>0.999</td>
<td>0.996</td>
<td>0.994</td>
<td>0.985</td>
<td>0.983</td>
<td></td>
</tr>
<tr>
<td>166</td>
<td>0.999</td>
<td>0.997</td>
<td>0.995</td>
<td>0.987</td>
<td>0.985</td>
<td></td>
</tr>
<tr>
<td>169</td>
<td>0.9995</td>
<td>0.998</td>
<td>0.997</td>
<td>0.989</td>
<td>0.988</td>
<td></td>
</tr>
<tr>
<td>170</td>
<td>1.00</td>
<td>0.999</td>
<td>0.998</td>
<td>0.991</td>
<td>0.990</td>
<td></td>
</tr>
<tr>
<td>173</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.993</td>
</tr>
<tr>
<td>176</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.9995</td>
<td>0.995</td>
</tr>
<tr>
<td>179</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.995</td>
<td></td>
</tr>
<tr>
<td>180</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td>0.996</td>
</tr>
<tr>
<td>183</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.998</td>
<td></td>
</tr>
<tr>
<td>186</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

*Average: Bone age within 1 year of chronologic age.


The developmental abnormality holoprosencephaly, which can be associated with GHD, may be caused by gene defects affecting PACAP and PACAP receptor expression.

The hypothalamus, heart, lung, and adipose tissue. Administration of ghrelin stimulates food intake and obesity and raises plasma GH concentrations and, to a lesser extent, adrenocorticotropic hormone (ACTH). These data suggest that ghrelin is an important stimulus for nutrient allocation for growth and metabolism and a central component of the GH regulatory system. Pituitary adenylate cyclase-activating peptide (PACAP), a hypothalamic peptide possibly involved in the regulation of GH secretion, is a member of the PACAP/glucagon superfamily. The developmental abnormality holoprosencephaly, which can be associated with GHD, may be caused by gene defects affecting PACAP and PACAP receptor expression.
The synthesis and secretion of GH are also regulated by the IGF peptides. Inhibition of GH secretion by IGF-I and IGF-II have been identified in varied pituitary cell systems. Inhibition of GH secretion by IGF-I and IGF-II in rat anterior pituitary cells has been demonstrated in a perfusion system, and spontaneous GH secretion is diminished in humans treated with synthetic IGF-I.

The episodic release of GH from pituitary somatotrophs results in intermittent increases in serum levels of GH that are separated by periods of low or undetectable levels, during which time GH secretion is minimal. The pulsatile nature of GH secretion has been demonstrated by frequent serum sampling, coupled with the use of immunofluorometric or chemoluminescent assays of GH. Under normal circumstances, serum GH levels are less than 0.04 µg/L between secretory bursts. Consequently, it is impractical to assess GH secretion by random serum sampling. Extensive sampling studies in people at different ages, in healthy people, and in many patients with abnormal conditions have defined GH pulses, basal secretion, and diurnal variability.

Computer programs have been developed to indicate whether changes in GH levels in various life periods and under diverse clinical circumstances occur because of a change in a secretory mass or pulse frequency, an altered clearance, or a combination of these processes. A model-free measure, is applied to quantify the degree of orderliness of GH release patterns. The impact of the specific nature of pulsatile GH secretion on its biologic actions is under study. For example, it appears that better statural growth is associated with large swings of GH output but of relatively uniform magnitude in an irregular sequence (high approximate entropy).

By 9 to 12 weeks of gestation, GH-secreting cells have been identified, and by 7 to 9 weeks, immunoreactive pituitary GH is present. By 5 weeks of gestation, fetal pituitary cells secrete GH in vitro, before the hypothalamic-portal vascular system is differentiated. By at least 6 weeks of gestation, Pit-1 messenger RNA (mRNA) and Pit-1 protein are expressed; its abundant presence early in gestation suggests an important role in cytodifferentiation and cell proliferation.

By the end of the first trimester, GH can be identified in fetal serum, with peak levels of around 150 µg/L in midgestation. Throughout the latter part of pregnancy, serum levels fall and are lower in term than in premature infants, perhaps reflecting feedback by the higher serum levels of IGF peptides characteristic of the later stages of gestation.

Through childhood and early puberty, mean levels of GH decrease from values of 25 to 35 µg/L in the neonatal period to approximately 5 to 7 µg/L. During adolescence, 24-hour GH secretion peaks, undoubtedly contributing to the high serum levels of IGF-I characteristic of puberty. During mid to late puberty, the increase in GH production is due both to enhanced pulse amplitude and increased mass of GH per secretory burst rather than a change in pulse frequency.

In the face of stable levels of the GH-binding protein (GHBP), the enhanced pubertal GH production is associated with higher levels of free GH. After adolescence, GH and IGF production begins to decline and continues to fall throughout adult life. Normal young adult men generally experience 6 to 10 GH secretory bursts per 24 hours, a value similar to that in younger children and in adolescents. In contrast, 24-hour GH production rates for normal men range from 0.25 to 0.52 mg/m² surface area about 20% to 30% of pubertal levels; this is largely due to decreased GH pulse amplitude with age. Indeed, puberty may be considered, with some justification, a period of "acromegaly," whereas aging, with its decrease in GH secretion, has been termed the somatopause.

Physiologic states that affect GH secretory function include obesity, fasting, and exercise. Stress and gonadal steroids. Maximal GH secretion occurs during the night, especially at the onset of the first slow-wave sleep (stages III and IV). Rapid-eyemovement sleep, however, is associated with low GH secretion. A circadian rhythm of somatostatin secretion, on which is superimposed episodic bursts of GHRH release, may help explain the nocturnal augmentation of GH production. When testosterone was administered to boys with delayed puberty, spontaneous GH release was enhanced, but such a change was not duplicated by administration of nonaromizable androgens, emphasizing the possible unique importance of estrogen on GH secretion.

The effects of testosterone on serum IGF-I levels may, in part, be independent of GH because individuals with mutations of the GH receptor (GHR) still experience a rise in serum IGF-I during puberty.

With a combination of deconvolution analysis, approximate...
entropy, and cosine regression analysis, Velthuis and associates \cite{94,103} carefully evaluated intensive GH sampling data, derived from measurements in sensitive GH assays, in prepubertal and pubertal boys and girls. In addition to the amplified secretory burst mass originating from jointly increased GH pulse amplitude and duration, they found that sex steroids selectively affected faceted of GH neurosecretory control; estrogen increases basal GH secretion rate and irregularity of GH release patterns, whereas testosterone stimulates greater GH secretory burst mass and IGF-1 concentrations.

Obesity is characterized by markedly decreased GH production, reflected by nearly a marked decrease in number of GH secretory bursts and of half-life. Obesity in childhood and adolescence, similarly, is characterized by decreased GH production but normal IGF and increased GHB levels and often increased linear growth. The hyponatremia associated with obesity causes lowered IGF-binding protein (IGFBP)-1 and, perhaps, higher free IGF-I levels. Endogenous GH secretion and levels achieved during provocative tests in these obese subjects \cite{104} approximate the diagnostic range of GHD. Fasting increases both the number and amplitude of GH secretory bursts, presumably reflecting decreased somatostatin secretion and enhanced GHRH release while lowering GHBP concentrations. Rapid changes in levels of IGFBPs in response to altered nutrition and changes of insulin levels may modify the effect of IGF-I on its negative feedback and effector sites. \cite{105,106}

Body mass also influences GH production in normal prepubertal and pubertal children and adults. \cite{107,108}

**Growth Hormone Receptor/Growth Hormone Binding Protein**

Leung and colleagues \cite{124} cloned both the rabbit and human complementary DNAs (cDNAs) for the GHR. Each contains an open reading frame of 620 amino acids and encodes a mature receptor of 620 amino acids and a predicted molecular weight of 70 kd before glycosylation. There are three domains: (1) an extracellular, hormone-binding domain, (2) a single membrane-spanning domain, and (3) a cytoplasmic domain.

In humans, the most important circulating GH-binding protein appears to be derived from proteolytic cleavage of the extracellular domain of the receptor. \cite{126} In the mouse \cite{127} and rat, \cite{128} however, there are multiple transcripts for the GHR; the larger, 3.4 to 4.8 kb, transcript codes for the intact receptor and the 1.2 to 1.9 kb transcript codes for the soluble GHBP.

The coding and 3' untranslated regions of the human GHR are encoded by nine exons, numbered 2 to 10. \cite{129} The gene for the human GHR is located on chromosome 5p13.1-p12, where it spans more than 87 kb. \cite{130} The GHR shows sequence homology with the prolactin receptor and with receptors for interleukin (IL)-2, IL-3, IL-4, IL-6, and IL-7, as well as receptors for erythropoietin, granulocyte-macrophage colony-stimulating factor (GM-CSF), and interferon. \cite{131} The GHR is a member of the class 1 hematopoietic cytokine family. \cite{132} Examination of the crystal structure of the GH/GHR complex revealed that the complex consists of one molecule of GH bound to two GHR molecules, indicating a GH-induced receptor dimerization, which is necessary for GH action. \cite{133}

**Growth Hormone Receptor/Growth Hormone Binding Protein**

After binding to its receptor, GH stimulates phosphorylation of a protein with an apparent molecular weight ratio of 120 kd. \cite{134} Although it was originally suspected that the GHR might be capable of autophosphorylation, it is now apparent that the major tyrosine-phosphorylated protein is associated with the receptor rather than being the receptor itself. JAK2 has been recently identified as the critical GH-R-associated tyrosine kinase. \cite{135} The presumed sequence of steps in GH action is shown in Figure 23-20 (Figure Not Available):

1. Binding of GH to the membrane-associated GHR.
2. Sequential dimerization of the GHR through binding to each of two specific sites on GH.
3. Interaction of the GHR with JAK2.
4. Tyrosine phosphorylation of both JAK2 and the GHR.
5. Changes in cytoplasmic and nuclear protein phosphorylation and dephosphorylation.

GH-dependent and JAK2-dependent phosphorylation have been demonstrated for many cytoplasmic signaling molecules that, after forming homo- and heterodimers, translocate into the nucleus, bind DNA, and activate transcription. \cite{136,137,138} How all of these seemingly redundant pathways intersect to mediate the various anabolic and metabolic actions of GH remains to be elucidated.

The major GHBP in human plasma binds GH with high specificity and affinity but with relatively low capacity, because about 45% of circulating GH is bound. \cite{139,140,141} The GHBP is, in essence, the extracellular domain of the GHR and has an apparent molecular weight ratio of approximately 55 kd. An additional GHBP, not
that is, low levels are associated with states of growth hormone insensitivity (GHI). In the rapidly growing child, however, levels of GHBP are quite low. Initial assays for GHBP involved incubation of serum with \(^{125}\text{I}\)-GH and separation of bound from free radioligand. Carlson and co-workers have developed a ligand-mediated immunofunctional assay for measurement of GHBP.

Levels of GHBP are low in early life, rise through childhood, and plateau during the pubertal years and adulthood. Once puberty is reached, levels are usually constant for a given individual. Impaired nutrition, diabetes mellitus, hypothyroidism, chronic liver disease, and a spectrum of inherited abnormalities of the GHR are associated with low levels of GHBP, whereas obesity, refeeding, early pregnancy, and estrogen treatment can cause elevated levels of GHBP.

A direct correlation exists between GHBP levels and body mass index. Serum GHBP levels correlate inversely with 24-hour GH production; this reciprocal relationship between GH production and GHBP in normal subjects and in subjects with idiopathic short stature (ISS) may result from adjustments of GH secretion to accommodate GH levels that may be genetically determined or modulated by environmental factors such as nutritional status. Assays of serum levels of GHBP are useful in identifying subjects with GH insensitivity due to genetic abnormalities of the GHR. Patients with GHI due to nonreceptor abnormalities, defects of the intracellular domain of the GHR, or inability of the receptor to dimerize may, however, have normal serum levels of GHBP.

Inhibition of GH signaling by several members of the GH-inducible suppressors of cytokine signaling (SOCS) family has been reported. The importance of SOCS proteins in controlling growth is demonstrated by the finding of gigantism in SOCS-2 knockout mice. Endotoxin and proinflammatory cytokines, such as IL-1 and tumor necrosis factor (TNF-), which can also induce SOCS proteins, produce GHI. SOCS-3 induced by IL-1 and TNF- or by endotoxin in vivo may play a role in the GHI induced by sepsis. Critically ill patients with septic shock treated with GH had increased mortality, possibly related to induction of GHI in specific tissues as a consequence of endotoxemia and cytokinemia.

**Growth Hormone Actions**

According to the somatomedin hypothesis, the anabolic actions of GH are mediated through the IGF peptides. Although this theory is largely true, GH is also capable of inducing effects that are independent of IGF activity. Indeed, the actions of GH and IGF are, on occasion, contradictory, as evident in the diabetogenic effects of GH and IGF-1 and tumor necrosis factor (TNF-), which can also induce SOCS proteins, produce GHI. SOCS-3 induced by IL-1 and TNF- or by endotoxin in vivo may play a role in the GHI induced by sepsis. Critically ill patients with septic shock treated with GH had increased mortality, possibly related to induction of GHI in specific tissues as a consequence of endotoxemia and cytokinemia.

2. Bone: stimulation of osteoclast differentiation and activity, stimulation of osteoblast activity, and increase of bone mass by endochondral bone formation.
3. Adipose tissue: acute insulin-like effects, followed by increased lipolysis, inhibition of lipoprotein lipase, stimulation of hormone-sensitive lipase, decreased glucose transport, and decreased lipogenesis.
4. Muscle: increased amino acid transport, increased nitrogen retention, increased lean tissue, and increased energy expenditure.

The concept of IGF-independent actions of GH is supported by in vivo studies, in which GH-I cannot duplicate all of the effects of GH, such as nitrogen retention and insulin resistance. The administration of GH for 1 to 3 weeks to calorically restricted normal or obese men results in significant nitrogen retention, although this effect does not persist with prolonged therapy. The effects of GH in normal human aging and in catabolic states are subjects of active investigation.
Insulin-Like Growth Factors

Historical Background

The IGFs (or somatomedins) are a family of peptides that are, in part, GH-dependent and that mediate many of the anabolic and mitogenic actions of GH. Originally identified in 1957 by their ability to stimulate $^35$S sulfate incorporation into rat cartilage and termed sulfation factor, concurrent investigations indicated that only one component of the insulin-like activity of normal serum could be blocked by the addition of anti-insulin antibodies. The remaining activity, termed nonsuppressible insulin-like activity (NSILA), was subsequently demonstrated to contain two soluble, low-molecular-weight (7-kd) forms, named NSILA-1 and NSILA-II. A third line of investigation arose from studies by Dulak and Temin on the mitogenic nature of bovine serum; the mitogenic factor was termed multiplication-stimulating activity (MSA) and shares metabolic and mitogenic activities with both sulfation factor and NSILA.

In 1972, the restrictive labels of sulfation factor and NSILA were replaced by the term somatomedin. The following criteria for a somatomedin were established:

1. The concentration in serum must be GH-dependent.
2. The factor must possess insulin-like activity in extraskeletal tissues.
3. The factor must promote the incorporation of sulfate into cartilage.
4. The factor must stimulate DNA synthesis and cell multiplication.

Purification yielded two somatomedin peptides: a basic peptide (somatomedin-C) and a neutral peptide (somatomedin-A). In 1978, Rinderknecht and Humble described two active somatomedins from human plasma, and after demonstrating a striking structural resemblance to proinsulin, renamed them insulin-like growth factors (IGFs).

IGF Structure and Molecular Biology

IGF-I, a basic peptide of 70 amino acids, correlates with somatomedin-C, and IGF-II is a slightly acidic peptide of 67 amino acids. The two peptides share 45 of 73 possible amino acid positions and have approximately 50% amino acid homology to insulin. Like insulin, both IGFs have A and B chains connected by disulfide bonds. The connecting C-peptide region is 12 amino acids long for IGF-I and 8 amino acids for IGF-II, bearing no homology with the C-peptide region of proinsulin. IGF-I and IGF-II also differ from proinsulin in possessing carboxy-terminal extensions, or D-peptides, of 8 and 6 amino acids, respectively. This structural similarity explains the ability of both IGFs to bind to the insulin receptor and of insulin to bind to the type 1 IGF receptor (see later). On the other hand, structural differences probably also explain the failure of insulin to bind with high affinity to the IGFBPs (see later).

IGF Variants

There are several variants of the two IGF peptides. Rinderknecht and Humble reported that up to one fourth of the IGF-II isolated from human plasma lacked the N-terminal alanine. Jansen and colleagues demonstrated that an IGF-II cDNA isolated from a human liver library predicted an IGF-II variant in which Ser was replaced by Arg-Leu-Pro-Gly, and Zumstein and associates identified this variant peptide subsequently in human plasma. Zumstein and associates isolated a 10-kd IGF-II variant from human plasma that contains a 21-residue carboxy extension, representing a portion of the E domain of pro-IGF-II (see later). In one peptide fragment isolated, Ser was replaced by Cys-Gly-Asp. A 25-kd IGF-II variant was isolated by Gowan and colleagues, presumably representing a carboxy-terminal extension.

The significance of "big" IGF-II forms is still uncertain. In general, these variants appear capable of binding to IGF and insulin receptors and to IGFBPs and can participate in formation of the 150-kd IGF/IGFBP-3/acid-labile subunit (ALS) ternary complex. Big IGF-II can be produced by mesenchymal tumors and can cause nonsiled-cell tumor hypoglycemia (NICTH).

Daughaday and co-workers described a patient with a leiomyosarcoma and recurrent hypoglycemia, in whom 70% of serum IGF-II was in higher-molecular-weight forms. Removal of the tumor eliminated big IGF-II from the serum and corrected the hypoglycemia. The presence of big IGF-II in NICTH has been confirmed in multiple laboratories, but it is unclear why hypoglycemia occurs in the face of normal total serum IGF-II levels.

Zapf has proposed that NICTH occurs when secretion of big IGF-II results in suppression of GH, insulin, and 7-kd IGF-II, leading to decreased production of IGF-I, IGFBP-3, and the ALS and increased production of IGFBP-2. This leads to a shift in the distribution of IGF-II from the 150-kd ternary complex to the 40- to 50-kd molecular-weight complex, composed of IGFBP-3, IGFBP-2, and a number of other low-molecular-weight IGFBPs. It is presumed that this results in increased bioavailability of IGF-II to target tissues, enhanced glucose consumption, and decreased hepatic glucose production.

Big forms of IGF-I have not been as thoroughly documented as with IGF-II. Powell and associates, however, have reported that IGF-I forms with an apparent molecular weight ratio as high as 19 kD may be found in uredine serum. Large molecular forms of IGF-I have also been identified in conditioned media of human fibroblast cell lines.

Two IGF-I precursor molecules have been identified. The first 134 amino acids of each are identical, comprising the signal peptide (48 amino acids), the mature IGF-I molecule (70 amino acids), and the first 16 amino acids of the E domain of the precursor. IGF-I A has additional 19 amino acids, and IGF-I B has additional 61 amino acids (total 195 residues). Alternative splicing of the IGF-I gene presumably generates the two mRNAs. The primary IGF-I translation product in human, rat, and mouse contains 180 amino acids, including a 24-residue signal peptide, the 67-amino acid mature IGF-II sequence, and a carboxy-terminal E peptide of 89 amino acids.

The IGF-I Gene

The IGF genes (Fig. 23-21) are expressed differently in the embryo, fetus, child, and adult. Single large genes encode both IGF-I and IGF-II. The human IGF-I gene is located on the long arm of chromosome 12, and it contains at least six exons. Exons 1 and 2 encode alternative signal peptides, probably each containing several transcription start sites. Exons 3 and 4 encode the remaining signal peptide, the remainder of the mature IGF-I molecule, and part of the trailer peptide (E peptide). Exons 5 and 6 encode, alternatively, used segments of the trailer peptide (resulting in the IGF-IA and IGF-IB forms) and 3' untranslated sequences with multiple different polyadenylation sites. The wide diversity of IGF-I mRNAs thus reflects the following:

1. Multiple leader exons and transcription start sites.
3. Multiple polyadenylation sites in exon 6.

The IGF-II Gene

The human IGF-II gene (Fig. 23-22) is located on the short arm of chromosome 11, adjacent to the insulin gene and spans 35 kb of genomic DNA, containing 9 exons. Exons 1 to 6 encode 5' untranslated RNA; exon 7 encodes the signal peptide and most of the mature protein; and exon 8 encodes the
carboxy-terminal portion of the protein and part of the trailer peptide, whose coding is completed in exon 9.

Thus, multiple mRNA species exist for both IGF-I and IGF-II, allowing for tissue-specific expression of specific transcripts and for developmental and hormonal regulation. The mechanisms involved in the regulation of IGF gene expression include the existence of multiple promoters, heterogeneous transcription initiation within each of these promoters, alternative splicing of various exons, differential RNA polyadenylation, and variable mRNA stability. Translation of IGF-I genes may also be under complex control.

**Regulation of IGF Gene Expression**

GH appears to be the primary regulator of IGF-I gene transcription, which begins as early as 30 minutes after intraperitoneal injection of GH into hypophysectomized rats. Transcriptional activation by GH affects both IGF-I promoters equivalently, resulting in a 20-fold rise in IGF-I mRNA. This coordinated, rapid induction of all IGF-I mRNA species coincides with induction of Spi 2.1 gene by GH, although the relationship between these two processes is still not clear. Furthermore, there may be tissue-to-tissue variability in GH-induced expression of IGF-I mRNA. Other factors that influence IGF-I gene expression include estrogen, which stimulates IGF-I mRNA expression in the uterus but inhibits GH-stimulated IGF-I transcription in the liver. The pubertal rise in serum IGF-I levels reflects the effect of gonadal steroids on IGF-I transcription, some of which results from the pubertal rise in GH secretion and some of which is due to a direct effect of gonadal steroids on IGF synthesis or secretion, because a pubertal rise in serum IGF levels is also observed in patients with GH deficiency.

The factors involved in the regulation of IGF-II gene expression are less clear. In humans and rats, IGF-II gene expression is high in fetal life and has been detected as early as the blastocyst stage in mice. Serum levels of IGF-II are high in midgestation in pregnant rabbits. Fetal tissues generally have high IGF-II mRNA levels that decline postnatally, although brain IGF-II mRNA remains high in the adult rat. IGF-II mRNA is expressed constitutively in a number of mesenchymal and embryonic tumors, including Wilms' tumor, rhabdomyosarcoma, neuroblastosarcoma, pheochromocytoma, hepatoblastoma, leiomyoma and leiomyosarcoma, liposarcoma, and colon carcinoma. Production of big IGF-II by these tumors may cause NICTHE (see earlier).

A tumor suppressor gene associated with Wilms' tumor (WT1) has been mapped to 11p13, close to the IGF-II locus (11p15.5), consistent with the possibility of a direct effect of WT1 on IGF-II gene transcription and suggesting an autocrine role for IGF-II in some tumors. This may be relevant in the embryonal tumors of Beckwith-Wiedemann syndrome (BWS), in which there may be loss of heterozygosity in the 11p15 maternally derived chromosome and paternal isodisomy, consistent with parental imprinting and a twofold increase in gene dosage of the active IGF-II allele.

**IGF Imprinting**

Gene regulation for the IGF system may also be subject to genomic imprinting, a process that influences the expression of specific genes. Namely, certain autosomal genes are expressed only from one of the two theoretically available alleles, in a manner that is specific for the parent of origin. The result is a heritable difference in gene expression, depending on whether a specific allele is inherited from the mother or the father. Allele-specific imprinting is exemplified by abnormalities of chromosome 15q11-13, where deletions of the paternal chromosome result in Prader-Willi syndrome, and deletions of the maternal locus are associated with Angelman syndrome, two phenotypically distinct conditions. The molecular mechanisms responsible for genomic imprinting involve variable DNA methylation. The first evidence for imprinting in the IGF axis emerged from studies of targeted gene disruption of Igf2 in the mouse which caused fetal growth retardation only when the disrupted allele was inherited from the father (i.e., maternally imprinted). The human IGF-II gene is similarly imprinted. In tissues where only maternal chromosomes are present, such as ovarian teratomas, no IGF-II expression is observed, whereas gene expression is observed in tissues where only paternal chromosomes are present (complete hydatidiform mole). Loss of imprinting (or relaxation of imprinting) of the IGF-II gene has been observed in rhabdomyosarcomas, lung cancers, Wilms' tumors, and choriocarcinoma. In such situations, IGF-II may act as an autocrine or paracrine growth factor for neoplastic tissue. Furthermore, in Wilms' tumor, loss of imprinting of the IGF-II gene is associated with reduced expression of the putative tumor suppressor gene H19. The H19 gene appears to be imprinted in a reciprocal manner to Igf2/Igf2, and the two genes may be coordinately regulated, because the genes are located near each other on the same chromosome.

The genes for the type II IGF receptor, which is the same as the cation-independent mannose-6-phosphate (M6P) receptor, are also imprinted, although in a different manner. Thus, the mouse Igf2 receptor gene and the human IGF type 2 receptor gene are both expressed by the maternal allele (i.e., paternally imprinted). If IGF-II functions as a fetal growth factor, there is potential for both maternal and paternal regulation of fetal size.

**Targeted Disruption of IGF Genes**

The role of the IGF axis in fetal growth has been firmly established by a series of studies involving IGF and IGF receptor null mutations. Unlike GH and GHR knockouts, which are near normal size at birth, mice with knockouts of the gene for either IGF-I or IGF-II have birth weights approximately 60% of normal. Mouse mutants lacking both IGF-I and the GHR are only 17% of normal. These observations indicate that both IGF-I and IGF-II are important embryonic
and fetal growth factors but that GH itself does have some independent role as well. Although fetal size was proportionately reduced in both situations and although morphogenesis was grossly normal, a higher neonatal death rate was observed following disruption of the gene for IGF-I. Growth delay began on day e11 for IGF-II knockouts and on day e13.5 for IGF-I knockouts. Those mice with IGF-I gene disruptions who survived the immediate neonatal period continued to have growth failure postnatally, with weights 30% of normal by 2 months of age. Indeed, postnatal growth was poorer than that observed in mice with GHR, GHRH receptor mutations, or Pit-1 mutations, indicating that both GH-dependent and GH-independent factors are necessary for normal growth.

A similar prenatal and postnatal growth phenotype has been observed in the one reported case of an IGF-I gene deletion. When the genes for both IGF-I and IGF-II were disrupted, weight at birth was only 30% of normal, and all animals died shortly after birth, apparently from respiratory insufficiency secondary to muscular hypotonia. Specific ablation of hepatic IGF-I production through the Cre/loxP recombination system confirmed that the liver is the principal source of circulating IGF-I but demonstrated that an 80% lowering of serum IGF-I levels had no apparent effect on postnatal growth. Presumably, either local (paracrine) chondrocyte production of IGF-I or other tissues (possibly adipose) maintain adequate endocrine sources of IGF-I to account for growth preservation.

Supportive data for the predominant role in growth of locally produced IGF-I are the modest decrement of postnatal growth seen in ALS null mice. These murine models are complex as IGFBP-3 levels are reduced despite increased GH and, in contrast with the human, rise after treatment with exogenous IGF-I. Further, the free IGF-I levels are normal in these animals but do not prevent an increment of GH production.

In another study, knockout of the gene for the type 1 IGF receptor resulted in birth weights 45% of normal and 100% neonatal lethality. Concurrent knockout of genes for IGF-I and the type 1 IGF receptor resulted in no further reduction in birth size (45% of normal), consistent with the concept that all IGF-I actions in fetal life are mediated through this receptor. On the other hand, simultaneous knockout of the genes for IGF-II and the type 1 IGF receptor resulted in further reduction of birth size to 30% of normal (as with simultaneous knockouts of IGF-I and IGF-II), suggesting that some of the fetal anabolic actions of IGF-II are mediated by a secondary mechanism (perhaps placental growth). This pathway does not appear to involve the type 2 IGF receptor, because knockout of this paternally imprinted gene results in an increased birth weight but death in late gestation or at birth. Because this receptor normally degrades IGF-II, increased growth reflects excess IGF-II acting through the IGF-I receptor; there is, however, variable accumulation of IGF-II in such mouse tissues. Knockout of the type 2 IGF receptor plus IGF-II causes a birth weight 60% of normal (as is the case with knockout of IGF-II alone) but allows fetal survival.

Several conclusions can be drawn from these studies:

1. IGF-I is important for both fetal and postnatal growth.
2. IGF-II is more important than GH for postnatal growth.
3. IGF-II is a major fetal growth factor.
4. The type 1 IGF receptor mediates anabolic actions of both IGF-I and IGF-II.
5. The type 2 IGF receptor is bifunctional, serving to both target lysosomal enzymes and to enhance IGF-II turnover.
6. IGF-I production is involved in normal fertility.
7. Placental growth is impaired only with IGF-II knockouts.

Whether these studies in mice are fully applicable to humans is yet unknown.

**IGF Peptide Assays**

Bioassays methods for IGF activity included stimulation of [35S] sulfate incorporation, using various modifications of the original method described by Salmon and Daughaday. A wide variety of other bioassays used stimulation of the synthesis of DNA, RNA, or protein or of glucose uptake. Such assays are cumbersome, subject to interference by IGFBPs, and incapable of distinguishing between IGF-I and IGF-II. When somatomedin-C (and later, IGF-I and IGF-II) were purified, it became possible to develop radioreceptor assays (RRAs) and competitive proteinbinding assays; development of specific antibodies permitted the development of accurate and specific measurement of IGF-I and IGF-II.

The issue of IGFBPs must be addressed in any IGF assay. For example, the discrepant results found in uremic sera assayed for IGF by bioassay, RRA, and immunassay are due to the interference of IGFBPs in RRA; such interference is a particular problem in conditions with a relatively high IGFBP/IGF peptide ratio and at the extremes of the assay (i.e., GHD or acromegaly). The most effective way to deal with IGFBPs is to separate them from IGF peptides by chromatography under acidic conditions; however, this is a labor-intensive procedure and has been occasionally replaced by an acid extraction procedure. Although this latter method may be reasonably effective for most serum samples, it is problematic in conditions of high IGFBP/IGF peptide ratios, such as conditioned media from cell lines and sera from newborns and from subjects with GHD or uremia.

Alternative methods include the use of antibodies generated against synthetic peptides, such as the C-peptide region of IGF-I or IGF-II, which does not bind to IGFBPs. In general, such antibodies have high specificity but relatively low affinity. An alternative approach, developed by Blum and colleagues, involves use of an antibody with high specificity for IGF-II, which permits the addition of excess unlabelled IGF-I, to saturate endogenous IGFBPs.

Bang and co-workers have bypassed the interference of IGFBPs by employing a truncated IGF-I radioligand, which has decreased affinity for IGFBPs. Currently, the most practical and effective way to perform accurate IGF assays with minimal interference by IGFBPs is to use the sandwich assay method. These assays, which can be performed in either enzyme-linked immunosorbent assay (ELISA) or immunoradiometric assay, do not employ a radiolabeled IGF molecule, which can bind to IGFBPs, as in conventional radioimmunoassays, and lead to erroneous readings if IGFBP levels are elevated. The absolute IGF values obtained in many of the assays may be falsely high because of low purity and inconsistent amino acid analyses of local standards, but this can be avoided by the use of the World Health Organization International Reference Reagent 87/518 calibration standard.

**Serum Levels of IGF Peptides**

In human fetal serum, IGF-I levels are relatively low and are positively correlated with gestational age. Some but not all groups have reported a correlation between fetal cord serum IGF-I levels with birth weight. IGF-I levels in human newborn serum are generally 30% to 50% of adult levels. Serum levels rise during childhood and attain adult levels at the onset of sexual maturation. During puberty, IGF-I levels rise to two to three times the adult range. Thus, levels during adolescence correlate better with Tanner stage (or bone age) than with chronologic age. Girls with gonadal dysgenesis show no adolescent increase in serum IGF-I, clearly establishing the role of the pubertal rise in IGF-I with the production of gonadal steroids. The pubertal rise in gonadal steroids may stimulate IGF-I production directly, by first leading to a rise in GH secretion, but patients with GH due to GHR mutations show a pubertal rise in serum IGF-I despite a decline in GH levels, thereby suggesting a direct effect of gonadal steroids on IGF-I.

After adolescence, or at least after 20 to 30 years of age, serum IGF-I levels demonstrate a gradual and progressive age-associated decline. a decline that is possibly responsible for the negative nitrogen balance, decrease in muscle mass, and osteoporosis of aging. This provocative hypothesis is unproven but has generated interest in the potential use of GH and IGF-I therapy in normal aging.

Human newborn levels of IGF-II are generally 50% of adult levels. By 1 year of age, however, adult levels are attained, with little, if any subsequent decline, even up to the seventh or eighth decade. This pattern of IGF-II levels in humans is different than that in the rat or mouse, in which serum IGF-II levels are also high in the fetus but rapidly decline postnatally to undetectable levels in the adult.

**Measurement of IGF Levels in Growth Disorders**

The GH dependency of the IGFs was established in the initial report from Salmon and Daughaday and further clarified with the development of sensitive and
specific immunoassays that distinguish between IGF-I and IGF-II. IGF-I levels are more GH-dependent than are IGF-II levels and are more likely to reflect subtle differences in GH secretory patterns. However, serum IGF-I levels, as stated earlier, are influenced by age, degree of sexual maturation, and nutritional status. As a result, construction of age-defined normative values may be misleading. IGF-I levels in normal children younger than 5 years of age are low, and there is overlap between the normal range and values in GH-deficient children. Assessment of serum IGF-II levels is less age-dependent, especially after 1 year of age, but IGF-II is less GH-dependent than is IGF-I.

Moore and associates performed GH stimulation tests in 78 children with heights below the 5th percentile and serum IGF-I levels lower than 0.5 U/mL. Although 19 of these children were subsequently discovered to have GH deficiency on the basis of standard provocative tests, there was an overlap of serum IGF-I levels between GH-deficient children and children with other forms of short stature and normal provocative GH levels. It was only in children with bone ages greater than 12 years that serum IGF-I levels permitted discrimination between GHD and normal short children. Similarly, Reiter and Lovinger found that 4 of 16 children with low provocative GH levels had normal serum IGF-I levels, whereas 7 of 25 children with normal provocative GH levels had low serum IGF-I levels.

Rosenfeld and colleagues evaluated the efficacy of IGF-I and IGF-II measurements in 68 GH-deficient patients, 197 children with normal stature, and 44 normal children with short stature (Fig. 23-24 and Fig. 23-25). Eighteen percent of the GH-deficient children had serum IGF-I levels within the normal range for age, and 32% of normal short children had low IGF-I levels. Low IGF-II levels were found in 52% of GH-deficient children and in 35% of normal short children. However, the use of combined IGF-I/IGF-II assays provided better discrimination. Only 4% of GH-deficient children had normal plasma levels of both IGF-I and IGF-II. Furthermore, only 0.5% of normal children and 11% of normal short children had low serum levels of both IGF-I and IGF-II.

The observation that many "normal short" children have low serum levels of IGF-I, IGF-II, or both calls into question the criteria by which the diagnosis of GHD is made. Given that provocative GH testing is both arbitrary and nonphysiologic and the inherent variability in GH assays, it is not surprising that the correlation between IGF-I levels and provocative GH levels is imperfect. These points are further supported by recent observations with immunoassays for IGFBP-3 (see later).

**IGF Receptors**

A study was published in which the binding of IGFs to the insulin receptor provided an explanation for their insulin-like activity. Shortly thereafter, Megyesi and co-workers identified distinct receptors for insulin and IGF in rat hepatic membranes. At least two classes of IGF receptors exist; insulin, at high levels, competes for occupancy of one form of IGF receptor but has essentially no affinity for the second form of receptor.

Structural characterization of these receptors documented the differences in the two forms of receptor (Fig. 23-26). The type 1 IGF receptor is closely related to the insulin receptor; both are heterotetramers comprised of two membrane-spanning subunits and two intracellular subunits. The subunits contain the binding sites for IGF-I and are linked by disulfide bonds. The subunits contain a transmembrane domain and an adenosine triphosphate (ATP)-binding site and a tyrosine kinase domain that constitute the presumed signal transduction mechanism for the receptor. One mole of the full heterotetrameric receptor appears to bind one mole of ligand.

Although the type 1 IGF receptor has been commonly termed the IGF-I receptor, the receptor binds both IGF-I and IGF-II with high affinity, and both IGF peptides appear capable of activating tyrosine kinase by binding to this receptor. In studies involving transfection and overexpression of the type 1 IGF receptor cDNA, the Kd for IGF-I is typically in the range of 0.2 to 1 nM, affinity for IGF-II is usually slightly less but varies from study to study. The affinity of the type 1 receptor for insulin is generally 100-fold less, thereby explaining the relatively weak mitogenic effect of insulin.

Ullrich and associates deduced the structure of the human type 1 IGF receptor from cDNA; the mature peptide constitutes 1337 amino acids with a predicted molecular mass of 151,869. The translated heterodimer is subsequently cleaved at an Arg-Lys-Arg-Arg sequence at positions 707 to 710, and the released and subunits are linked by disulfide bonds to form the mature (1,2)-receptor in which two alpha chains are joined by secondary disulfide bonds. The subunits are extracellular and contain a cysteine-rich domain, which is critical for IGF binding. As is the case with the insulin receptor, the subunit has a short extracellular domain, a hydrophobic transmembrane domain, and the intracellular tyrosine kinase domain and ATP-binding site. Like the insulin receptor, the type 1 IGF receptor undergoes ligand-induced autophosphorylation, principally on tyrosines 1131, 1135, and 1136. Both receptors are believed to have evolved from a common ancestor protein but are encoded by genes on separate chromosomes (chromosome 15 for the type 1 IGF receptor and chromosome 19 for the insulin receptor).

The type 1 IGF receptor gene spans greater than 100 kb of genomic DNA, with 21 exons; the genomic organization resembles that of the type 1 IGF receptor gene. By Northern blot hybridization, human mRNA reveals two bands of 11 and 7 kb; rat tissues contain only the 11-kb band. Type 1 IGF receptor mRNA is most abundant in embryonic tissues and appears to decrease encoded by exons 4 to 10. The peptide cleavage site involved in generation of the and subunits is encoded by exon 11, and the tyrosine kinase domain of the subunit is encoded by exons 16 to 20. It is in the latter region that the type 1 IGF receptor and insulin receptor share the greatest sequence homology, ranging from 80% to 95%; interspecies homology in this region of the receptors is also high. Exon 21 encodes 3 untranslated sequences.

Type 1 IGF receptor mRNA has been identified in virtually every tissue except liver. By Northern blot hybridization, human mRNA reveals two bands of 11 and 7 kb; rat tissues contain only the 11-kb band. Type 1 IGF receptor mRNA is most abundant in embryonic tissues and appears to decrease encoded by exons 4 to 10. The peptide cleavage site involved in generation of the and subunits is encoded by exon 11, and the tyrosine kinase domain of the subunit is encoded by exons 16 to 20. It is in the latter region that the type 1 IGF receptor and insulin receptor share the greatest sequence homology, ranging from 80% to 95%; interspecies homology in this region of the receptors is also high. Exon 21 encodes 3' untranslated sequences.

Type 1 IGF receptor mRNA has been identified in virtually every tissue except liver. By Northern blot hybridization, human mRNA reveals two bands of 11 and 7 kb; rat tissues contain only the 11-kb band. Type 1 IGF receptor mRNA is most abundant in embryonic tissues and appears to decrease encoded by exons 4 to 10. The peptide cleavage site involved in generation of the and subunits is encoded by exon 11, and the tyrosine kinase domain of the subunit is encoded by exons 16 to 20. It is in the latter region that the type 1 IGF receptor and insulin receptor share the greatest sequence homology, ranging from 80% to 95%; interspecies homology in this region of the receptors is also high. Exon 21 encodes 3' untranslated sequences.
with age. The type 1 IGF receptor is present at the embryonic eight-cell stage (the type 2 IGF receptor is first demonstrable at the 2-cell stage) and becomes widely expressed postimplantation, consistent with the observation that this receptor is essential for normal fetal growth.

As with other growth factor receptor tyrosine kinases, binding of ligand (IGF-I or IGF-II) induces receptor autophosphorylation of critical tyrosine residues in the type 1 receptor.\textsuperscript{326} Mutations of the ATP-binding site or of critical tyrosine residues in the subunit result in loss of IGF-stimulated thymidine incorporation and glucose uptake.

Autophosphorylation appears to occur by transphosphorylation of sites on the opposite subunit.\textsuperscript{327} The activated type 1 IGF receptor is capable of phosphorylating other tyrosine-containing substrates, such as insulin receptor substrate 1 (IRS-1), a 185-kd protein that is the predominant substrate of the insulin receptor kinase and IRS-2 (Fig. 23-28) (Figure Not Available).\textsuperscript{328} IRS-1 contains specific phosphorysosyRNA motifs that can associate with proteins containing SH2 (src homology 2) domains, such as PI3-kinase (phosphatidylinositol-3 kinase),\textsuperscript{329} Grb2 (growth factor receptor-bound protein 2),\textsuperscript{330} Syp (a phosphotyrosine phosphatase),\textsuperscript{331} and Nck (an oncogenic protein).\textsuperscript{332} The substrates, which are phosphorylated by the IGF receptor, include the members of the IRS family, particularly IRS-1 and IRS-2, as both of the knockout mice models for these genes result in poor growth (as well as insulin resistance).\textsuperscript{333} Other IRS molecules may have a negative feedback role in regulating IGF action.\textsuperscript{334}

Activation of the type 1 IGF receptor also leads to tyrosine phosphorylation of Shc (src homology domain-containing protein),\textsuperscript{335} which then associates with Grb2 and activates Ras, leading to a cascade of protein kinases, including Raf, MAP kinase kinases, MAP kinases, and S6 kinase.\textsuperscript{336} Thus, phosphorylation of IRS-1 by either the type 1 IGF or insulin receptor activates multiple signaling cascades that ultimately influence nuclear transcription and gene expression. It is, presumably, at this level that the IGF peptides exert their mitogenic and anabolic actions. Given that insulin and IGF peptides activate similar, if not identical, signaling pathways through their own specific receptors, it is unclear how the cell distinguishes between these overlapping ligands. Whether this merely reflects the relative levels of receptors or whether divergent downstream pathways exist for insulin and IGF action remain questions for future investigation.\textsuperscript{337}

Although targeted disruption of the gene for the type 1 IGF receptor causes fetal growth retardation, a clear role for this receptor in the cell cycle has not been established. Fibroblast cell lines derived from mouse embryos homozygous for the knockout gene still undergo cell cycle-dependent division, although at a slower rate.\textsuperscript{338} On the other hand, the transformed phenotype of some cells may be critically dependent on expression of the type 1 IGF receptor. The SV40TAg is capable of inducing a transformed phenotype in a cell only in the presence of intact type 1 IGF receptors.\textsuperscript{339} NIH 3T3 cells and Rat-1 fibroblasts that are made to overexpress the type 1 IGF receptor develop IGF-dependent neoplastic transformation with colony formation in soft agar and tumor formation in nude rats.\textsuperscript{340} Prager and colleagues\textsuperscript{341} have shown that truncation of the type 1 IGF receptor at the amino terminal increases transforming potential, suggesting that the subunit normally restricts this function and that the binding of IGF to the receptor represses the effects of mitogenic stimulation.

Variants of both the a and subunits are present in placenta,\textsuperscript{342} muscle,\textsuperscript{343} and brain.\textsuperscript{344} These variants may explain seemingly anomalous competitive binding studies.\textsuperscript{345} The molecular mechanisms for the formation of such receptor variants have not been identified; nor is it clear if they differentially bind IGF-I, IGF-II or insulin. The formation of IGF-insulin receptor hybrids that contain an IGF hemireceptor disulfide-linked to an insulin hemireceptor (see Fig. 23-26)\textsuperscript{346} appears to be ligand-dependent,\textsuperscript{347} and studies with monoclonal antibodies specific for the insulin or type 1 IGF receptor suggest that such receptors develop in cells with abundant native receptors, such as muscle and placenta.\textsuperscript{348} Such hybrids have near-normal affinity for IGF-I but decreased affinity for insulin. The physiologic significance of such hybrid receptors is unknown.

The type 2 IGF receptor bears no structural homology with either the insulin or type 1 IGF receptors. It has an apparent molecular mass of 220,000 under physiologic significance of such hybrid receptors is unknown. Fibroblast cell lines derived from mouse embryos homozygous for the knockout gene still undergo cell cycle-dependent division, although at a slower rate. On the other hand, the transformed phenotype of some cells may be critically dependent on expression of the type 1 IGF receptor. The SV40TAg is capable of inducing a transformed phenotype in a cell only in the presence of intact type 1 IGF receptors. NIH 3T3 cells and Rat-1 fibroblasts that are made to overexpress the type 1 IGF receptor develop IGF-dependent neoplastic transformation with colony formation in soft agar and tumor formation in nude rats. Prager and colleagues have shown that truncation of the type 1 IGF receptor at the amino terminal increases transforming potential, suggesting that the subunit normally restricts this function and that the binding of IGF to the receptor represses the effects of mitogenic stimulation.

Variants of both the a and subunits are present in placenta, muscle and brain. These variants may explain seemingly anomalous competitive binding studies. The molecular mechanisms for the formation of such receptor variants have not been identified; nor is it clear if they differentially bind IGF-I, IGF-II or insulin. The formation of IGF-insulin receptor hybrids that contain an IGF hemireceptor disulfide-linked to an insulin hemireceptor appears to be ligand-dependent, and studies with monoclonal antibodies specific for the insulin or type 1 IGF receptor suggest that such receptors develop in cells with abundant native receptors, such as muscle and placenta. Such hybrids have near-normal affinity for IGF-I but decreased affinity for insulin. The physiologic significance of such hybrid receptors is unknown.

The type 2 IGF receptor bears no structural homology with either the insulin or type 1 IGF receptors. It has an apparent molecular mass of 220,000 under physiologic significance of such hybrid receptors is unknown. Fibroblast cell lines derived from mouse embryos homozygous for the knockout gene still undergo cell cycle-dependent division, although at a slower rate. On the other hand, the transformed phenotype of some cells may be critically dependent on expression of the type 1 IGF receptor. The SV40TAg is capable of inducing a transformed phenotype in a cell only in the presence of intact type 1 IGF receptors. NIH 3T3 cells and Rat-1 fibroblasts that are made to overexpress the type 1 IGF receptor develop IGF-dependent neoplastic transformation with colony formation in soft agar and tumor formation in nude rats. Prager and colleagues have shown that truncation of the type 1 IGF receptor at the amino terminal increases transforming potential, suggesting that the subunit normally restricts this function and that the binding of IGF to the receptor represses the effects of mitogenic stimulation.

Variants of both the a and subunits are present in placenta, muscle and brain. These variants may explain seemingly anomalous competitive binding studies. The molecular mechanisms for the formation of such receptor variants have not been identified; nor is it clear if they differentially bind IGF-I, IGF-II or insulin. The formation of IGF-insulin receptor hybrids that contain an IGF hemireceptor disulfide-linked to an insulin hemireceptor appears to be ligand-dependent, and studies with monoclonal antibodies specific for the insulin or type 1 IGF receptor suggest that such receptors develop in cells with abundant native receptors, such as muscle and placenta. Such hybrids have near-normal affinity for IGF-I but decreased affinity for insulin. The physiologic significance of such hybrid receptors is unknown.

The type 2 IGF receptor bears no structural homology with either the insulin or type 1 IGF receptors. It has an apparent molecular mass of 220,000 under physiologic significance of such hybrid receptors is unknown. Fibroblast cell lines derived from mouse embryos homozygous for the knockout gene still undergo cell cycle-dependent division, although at a slower rate. On the other hand, the transformed phenotype of some cells may be critically dependent on expression of the type 1 IGF receptor. The SV40TAg is capable of inducing a transformed phenotype in a cell only in the presence of intact type 1 IGF receptors. NIH 3T3 cells and Rat-1 fibroblasts that are made to overexpress the type 1 IGF receptor develop IGF-dependent neoplastic transformation with colony formation in soft agar and tumor formation in nude rats. Prager and colleagues have shown that truncation of the type 1 IGF receptor at the amino terminal increases transforming potential, suggesting that the subunit normally restricts this function and that the binding of IGF to the receptor represses the effects of mitogenic stimulation.

Variants of both the a and subunits are present in placenta, muscle and brain. These variants may explain seemingly anomalous competitive binding studies. The molecular mechanisms for the formation of such receptor variants have not been identified; nor is it clear if they differentially bind IGF-I, IGF-II or insulin. The formation of IGF-insulin receptor hybrids that contain an IGF hemireceptor disulfide-linked to an insulin hemireceptor appears to be ligand-dependent, and studies with monoclonal antibodies specific for the insulin or type 1 IGF receptor suggest that such receptors develop in cells with abundant native receptors, such as muscle and placenta. Such hybrids have near-normal affinity for IGF-I but decreased affinity for insulin. The physiologic significance of such hybrid receptors is unknown.

The type 2 IGF receptor bears no structural homology with either the insulin or type 1 IGF receptors. It has an apparent molecular mass of 220,000 under physiologic significance of such hybrid receptors is unknown. Fibroblast cell lines derived from mouse embryos homozygous for the knockout gene still undergo cell cycle-dependent division, although at a slower rate. On the other hand, the transformed phenotype of some cells may be critically dependent on expression of the type 1 IGF receptor. The SV40TAg is capable of inducing a transformed phenotype in a cell only in the presence of intact type 1 IGF receptors. NIH 3T3 cells and Rat-1 fibroblasts that are made to overexpress the type 1 IGF receptor develop IGF-dependent neoplastic transformation with colony formation in soft agar and tumor formation in nude rats. Prager and colleagues have shown that truncation of the type 1 IGF receptor at the amino terminal increases transforming potential, suggesting that the subunit normally restricts this function and that the binding of IGF to the receptor represses the effects of mitogenic stimulation.

Variants of both the a and subunits are present in placenta, muscle and brain. These variants may explain seemingly anomalous competitive binding studies. The molecular mechanisms for the formation of such receptor variants have not been identified; nor is it clear if they differentially bind IGF-I, IGF-II or insulin. The formation of IGF-insulin receptor hybrids that contain an IGF hemireceptor disulfide-linked to an insulin hemireceptor appears to be ligand-dependent, and studies with monoclonal antibodies specific for the insulin or type 1 IGF receptor suggest that such receptors develop in cells with abundant native receptors, such as muscle and placenta. Such hybrids have near-normal affinity for IGF-I but decreased affinity for insulin. The physiologic significance of such hybrid receptors is unknown.
production of inositol triphosphate and diacylglycerol in proximal tubules and canine kidney membranes. Tally and co-workers reported that IGF-II stimulates the growth of a K562 human erythroleukemia cell line, an action not duplicated by either IGF-I or insulin. Minnilli and colleagues reported that IGF-II appears capable of acting as an autocrine growth factor and cell motility factor for human rhabdomyosarcoma cells, actions apparently mediated through the type 2 receptor, and IGF-II may activate a calcium-permeable cation channel via the type 2 IGF receptor, perhaps through coupling to a permissus toxicinsensitive, guanine nucleotide-binding protein (G protein). IGF-II is found in cells transfected with the human type 2 IGF receptor cDNA, IGF-II decreased CAMP accumulation promoted by cholera toxin or forskolin. Mutations or truncation of the small cytoplasmic domain of the receptor prevented these IGF-II actions. The type 2 IGF receptor also binds other molecules such as Msp-containing enzymes (e.g., cathepsin and urokinase), which may be important in the removal of these enzymes from the cellular environment, thus modulating tissue remodeling. In addition, the type 2 IGF receptor binds retinoid x and may mediate some of the growth inhibitory effects of retinoids.

As discussed earlier, knockout of the type 2 IGF receptor results in excessive growth. This receptor, therefore, may act as a growth inhibitory component of the IGF system responding to and mediating multiple antimitogenic systems.

**IGF-Binding Proteins** (Fig. 23-29) (Figure Not Available)

In contrast with insulin, the IGFs circulate in plasma complexed to a family of binding proteins that extend the serum half-life of the IGF peptides, transport the IGFs to target cells, and modulate the interaction of the IGFs with surface membrane receptors. The identification and characterization of IGF-binding proteins (IGFBPs) in body fluids and in conditioned media from cultured cells have been facilitated by the development of a number of biochemical and assay techniques, including gel chromatography, RRAAs, affinity cross-linking, Western ligand blotting, immunoblotting, and specific radioimmunoassays. However, study of the molecular biology of the IGFBPs has provided the most information concerning their structural inter-relationship.

**IGF-Binding Protein Structure** (Fig. 23-29)

To date, the cDNAs for six distinct human and rat IGFBPs have been cloned and sequenced. Their structural characteristics are summarized in Figure 23-30. The amino acid sequences of the six cloned mammalian IGFBPs are highly conserved. Within a species, the IGFBPs share an overall amino acid sequence homology in the order of 50%, and between species there is more than 80% sequence homology for individual IGFBPs. Perhaps the most impressive similarity in structure is the conservation of the number and placement of the cysteine residues. The total number of cysteines varies from 16 to 20 (18 are conserved in human IGFBP-3) to 5; IGFBP-6 conserves 16 of the 18, and IGFBP-4 has 2 additional cysteines in the middle region of the protein), and each of the IGFBPs has cysteine-rich regions at the amino-terminus and carboxyl-terminus. Conservation of the spatial order of the cysteines presumably indicates that the secondary structure of the IGFBPs, which is dependent on disulfide bonding, must also be conserved.

Disulfide bonding is essential for formation of the IGF

**IGF-Binding Proteins** (Fig. 23-28) (Figure Not Available)

Schematic representation of the insulin-like growth factor I receptor (IGF-IR). On binding IGF-I, the IGF receptor undergoes autophosphorylation at multiple tyrosine residues. The intrinsic kinase activity of the receptor also phosphorylates insulin receptor substrate 1 (IRS-1) at multiple tyrosine residues. Various SH domain-containing proteins, including PI 3-kinase, Syp, Fyn, and Nick, associate with specific phosphotyrosine-containing motifs within IRS-1. These docking proteins recruit diverse other intracellular substrates, which then activate a cascade of protein kinases, including Raf-1 and one or more related kinases. These protein kinases, in turn, activate various other elements, including nuclear transcription factors. Alterations in expression of various IGF-I-responsive genes results in longer-term effects of IGF-I, including growth and differentiation. This model of signal transduction cascades also shows a potential mechanism for the inhibition of apoptosis. (From Le Roth D, Bondy C, Yaker S, et al: The somatomedin hypothesis. Endor Rev 1991; 20:573. © The Endocrine Society.)

**IGF-Binding Protein Structure** (Fig. 23-30)

Figure 23-30 Amino acid sequences of insulin-like growth factor-binding proteins (IGFBPs) 1 to 6, deduced from nucleotide sequences. Sequences in the amino-terminal and carboxyl-terminal residues are aligned to show maximal similarities. Dashes indicate gaps. Residues that are identical in five or six of the six IGFBPs are shaded. (From Rechler MM: Insulin-like growth factor binding proteins. Vitam Horm 1993; 47:114.)

properties of the IGFBPs, such as cell association, IGF enhancement, and IGF-independent actions (see later) may be dependent on specific sequences in these midregions. An RGD (arginine-glycine-aspartic acid) sequence near the carboxyl-terminus of IGFBP-1 and IGFBP-2 is the minimum sequence required for the binding of many extracellular matrix proteins to membrane receptors of the integrin protein family, and IGFBPs may associate with the cell surface through such amino acid sequences. However, IGFBP-3, which lacks an RGD sequence, also binds to cell membranes, possibly to specific receptors (see later).

Under most conditions, the IGFBPs appear to inhibit IGF action, presumably by competing with IGF receptors for binding IGF peptides. For example, IGF analogues with decreased affinity for IGF receptors have increased biologic potency. In studies involving transfection of the human IGF-I-responsive gene into fibroblasts, increased expression of IGFBP-3 inhibited cell growth, even in the absence of added IGF, suggesting a direct inhibitory role of the binding protein. Under some conditions, however, the IGFBPs appear to enhance IGF action, perhaps by facilitating the delivery of IGF to target receptors.

The discovery of several groups of cysteine-rich proteins that contain domains similar to the amino-terminus of the IGFBPs has led to the proposal of an IGFBP superfamily, which includes the family of six high-affinity IGFBPs, as well as a number of IGFBP-related proteins (IGFBP-Ps). Three of the IGFBP-Ps (Mac25/IGFBP-P1; connective tissue growth factor, CTGF/IGFBP-P2; Nov/IGFBP-P3) have been shown to bind IGFs, although with considerably lower affinity than the IGFBPs. Like the IGFBPs, the IGFBP-Ps are modular proteins and the highly preserved amino-terminal domain appears to represent the consequence of exon shuffling of an ancestral gene. The role, if any, of the IGFBP-Ps in normal IGF physiology is unclear, but they seem likely that they influence cell growth by IGF-independent and IGF-dependent mechanisms.

Analysis of IGFBPs is further complicated by the presence of IGFBP-proteases, which degrade IGF, initially reported in the serum of pregnant women, proteases for IGFBP-2, IGFBP-3, IGFBP-4, and IGFBP-5 are present in serum, seminal plasma, cerebrospinal fluid and urine. It is likely that multiple IGFBP-proteases exist, including calcium-dependent serine proteases, kallikreins, cathepsins, and matrix metalloproteases. Proteolysis of IGFBPs complicates their assay and must be taken into consideration when measuring the various IGFBPs in biologic fluids. The physiologic significance of limited proteolysis of IGFBPs remains to be determined, although protease activity usually decreases the affinity of the IGFBP for IGF peptides (Fig. 23-31) and may enhance the mitogenic and anabolic effects of IGF peptides in this way. In prostate epithelial cells and prostate-specific antigen acts as a potent IGF-3 protease (Fig. 23-32) and Fig. 23-32).

The identification and characterization of IGFBPs in body fluids and in conditioned media from cultured cells have been facilitated by the development of a number of biochemical and assay techniques, including gel chromatography, RRAAs, affinity cross-linking, Western ligand blotting, immunoblotting, and specific radioimmunoassays. However, study of the molecular biology of the IGFBPs has provided the most information concerning their structural inter-relationship.

**IGF-Binding Protein Structure** (Fig. 23-30)

IGF-binding proteins (IGFBPs) 1 to 6, deduced from nucleotide sequences. Sequences in the amino-terminal and carboxyl-terminal residues are aligned to show maximal similarities. Dashes indicate gaps. Residues that are identical in five or six of the six IGFBPs are shaded. (From Rechler MM: Insulin-like growth factor binding proteins. Vitam Horm 1993; 47:114.)

properties of the IGFBPs, such as cell association, IGF enhancement, and IGF-independent actions (see later) may be dependent on specific sequences in these midregions. An RGD (arginine-glycine-aspartic acid) sequence near the carboxyl-terminus of IGFBP-1 and IGFBP-2 is the minimum sequence required for the binding of many extracellular matrix proteins to membrane receptors of the integrin protein family, and IGFBPs may associate with the cell surface through such amino acid sequences. However, IGFBP-3, which lacks an RGD sequence, also binds to cell membranes, possibly to specific receptors (see later).

Under most conditions, the IGFBPs appear to inhibit IGF action, presumably by competing with IGF receptors for binding IGF peptides. For example, IGF analogues with decreased affinity for IGF receptors have increased biologic potency. In studies involving transfection of the human IGF-I-responsive gene into fibroblasts, increased expression of IGFBP-3 inhibited cell growth, even in the absence of added IGF, suggesting a direct inhibitory role of the binding protein. Under some conditions, however, the IGFBPs appear to enhance IGF action, perhaps by facilitating the delivery of IGF to target receptors.

The discovery of several groups of cysteine-rich proteins that contain domains similar to the amino-terminus of the IGFBPs has led to the proposal of an IGFBP superfamily, which includes the family of six high-affinity IGFBPs, as well as a number of IGFBP-related proteins (IGFBP-Ps). Three of the IGFBP-Ps (Mac25/IGFBP-P1; connective tissue growth factor, CTGF/IGFBP-P2; Nov/IGFBP-P3) have been shown to bind IGFs, although with considerably lower affinity than the IGFBPs. Like the IGFBPs, the IGFBP-Ps are modular proteins and the highly preserved amino-terminal domain appears to represent the consequence of exon shuffling of an ancestral gene. The role, if any, of the IGFBP-Ps in normal IGF physiology is unclear, but it seems likely that they influence cell growth by IGF-independent and IGF-dependent mechanisms.

Analysis of IGFBPs is further complicated by the presence of IGFBP-proteases, which degrade IGF, initially reported in the serum of pregnant women, proteases for IGFBP-2, IGFBP-3, IGFBP-4, and IGFBP-5 are present in serum, seminal plasma, cerebrospinal fluid and urine. It is likely that multiple IGFBP-proteases exist, including calcium-dependent serine proteases, kallikreins, cathepsins, and matrix metalloproteases. Proteolysis of IGFBPs complicates their assay and must be taken into consideration when measuring the various IGFBPs in biologic fluids. The physiologic significance of limited proteolysis of IGFBPs remains to be determined, although protease activity usually decreases the affinity of the IGFBP for IGF peptides (Fig. 23-31) and may enhance the mitogenic and anabolic effects of IGF peptides in this way. In prostate epithelial cells and prostate-specific antigen acts as a potent IGF-3 protease (Fig. 23-32) and Fig. 23-32).
IGF-Binding Proteins as Carrier Proteins

Given the high affinity of the IGFs for IGF-I and IGF-II (Kₐ = 10⁻¹⁰ M⁻¹), virtually all IGF-I and IGF-II in serum is complexed to IGF-BPs. In normal adult serum, 75% to 80% of the IGF-I and IGF-II peptides are carried in a ternary complex consisting of one molecule of IGF plus one molecule of IGFBP-3 plus one molecule of an 8-kd protein termed the acid-labile subunit (ALS). Binding of ALS to form the full ternary complex occurs after the binding of IGF by IGFBP-3 or IGF-BP-5, although ALS may bind to IGFBP-3 even in the absence of IGF. The 150-kd ternary complex is too large to leave the vascular compartment and extends the half-life of IGF peptides from approximately 10 minutes to 1 to 2 hours for IGF in the IGFBP-3 ternary complex to 12 to 15 hours for IGF in the ternary complex. The presence of cell membrane proteins or receptors that specifically bind IGFBP-3 provides a potential mechanism for IGF-independent growth inhibitory actions of IGFBP-3.

IGF peptides in the 40- to 50-kd molecular-weight peak may not be restricted to the vascular compartment. IGFBP-1, IGFBP-2, and IGFBP-4, at least, can probably cross endothelial barriers. In the fetus and neonate, whose IGFBP-3 levels are relatively low, and in GHD and GH, binding of IGFs by IGFBP-1, IGFBP-2, IGFBP-4, and IGFBP-5 may predominate over binding to IGFBP-3. Similarly, in tumorinduced hypoglycemia associated with increased serum levels of IGF-II, ternary complex formation may be decreased, and most IGF peptides are found in the low-molecular-weight peak.

IGF-Binding Proteins as Modulators of IGF Action

In general, the binding affinity of IGF-BPs for IGF peptides is higher than that of IGF receptors, implying that IGF-BPs can modulate IGF binding to its receptors, thereby modulating IGF biologic actions. Coincubation of cells with IGF-I and a molar excess of IGFBP-3 results in an inhibition of IGF-stimulated thymidine incorporation in human fibroblasts, lipogenesis in rat epididymal adipocytes, and glucose consumption in mouse fibroblasts. Termination of inhibition apparently requires dissociation of IGFs from the IGF-IGFBP complex by mass action, proteolysis, or other mechanisms. IGF-BP proteases have been identified in a variety of body fluids and cell culture media and are postulated to play a role in altering IGF availability by lowering the affinities of IGF-BPs for their ligand, thereby increasing the availability of IGFs to cell membrane receptors (see earlier).

Under certain conditions, the IGF-BPs potentiate IGF action. In human and bovine fibroblasts, DNA synthesis and -aminoisobutyric acid transport are potentiated when cells are preincubated with IGFBP-3, whereas IGFBP-3 is inhibitory if added at the same time as IGF-I. These observations have suggested that cell association of IGFBP-3 during preincubation is essential for its IGF-potentiating effect, perhaps allowing IGFBP-3 to serve as a reservoir for IGFs and bringing the ligand into closer proximity to the type 1 IGF receptors. This cell surface association of IGFBP-3 may involve interaction with heparin and heparin sulfate proteoglycans on the cell membrane or specific IGFBP-3 receptors.

IGF-Independent Actions of IGF-Binding Proteins

The IGF-BPs are bioactive molecules that, in addition to binding IGF, have a variety of IGF-independent functions. These include growth inhibition in some cell types, direct induction of apoptosis, and modulation of the effects of other non-IGF growth factors. These effects of IGF-BPs are mediated by binding to their own receptors. The IGF-BP signaling pathways are currently being unraveled and involve interaction of IGF-BPs with nuclear retinoid receptors as well as with other molecules on the cell surface and in the cytoplasm.

IGFBP-3, itself, appears to have intrinsic inhibitory effects on cells, independent of its interaction with IGF. Vailland and co-workers found that the stimulation of DNA synthesis by basic FGF is inhibited by simultaneous treatment with IGFBP-3, even in the presence of levels of insulin, suggesting that sequentialization of IGF peptides from type 1 IGF receptors is not the only means whereby IGFBP-3 inhibits cell growth. IGFBP-3 is also more effective than immunoneutralization of IGF-I in inhibiting serum-stimulated DNA synthesis, and IGFBP-3 inhibits FSH-stimulated DNA synthesis in cultured ovarian granulosa cells, with or without added IGF. Under the same conditions, IGFBP-2 is less inhibitory, despite its higher affinity for IGF peptides. Expression of a transfected human IGFBP-3 cDNA in mouse fibroblasts inhibits both IGF-stimulated and insulin-stimulated cell proliferation. Similar studies of fibroblasts derived from mouse embryos homozygous for a targeted disruption of the type 1 IGF receptor again demonstrated inhibition with overexpression of IGFBP-3. These studies strongly support an IGF-independent action for IGFBP-3.

IGFBP-3 binds with high affinity to the surface of various cell types, including human breast cancer cells and rat chondrocytes, and inhibits monolayer growth of these cells in an IGF-independent manner. Furthermore, transcausal regulation of IGFBP-3 expression may be the mechanism for the inhibition of breast cancer cell growth by both transforming growth factor 2 and retinoic acid (Reduction of IGFBP-3 production may be a common pathway for multiple hormones and growth factors involved in the modulation of cell growth. For example, estrogen inhibits expression and secretion of IGFBP-3, whereas antestrogens stimulate production of IGFBP-3 in estrogen receptorpositive human breast cancer cells. Similarly, the mitogenic action of epidermal growth factor (EGF) in human cervical epithelial cells is associated with inhibition of IGFBP-3 expression, and the inhibitory effect of retinoic acid is accompanied by increased IGFBP-3 expression. Regulation of IGFBP-3 gene expression plays a role in signaling by p53, a potent tumor suppressor protein.

The presence of cell membrane proteins or receptors that specifically bind IGFBP-3 provides a potential mechanism for IGF-independent growth inhibitory actions of IGFBP-3.
IGFBP-3 is the predominant IGFBP in adult serum, where it carries approximately 75% of the total IGF, primarily as part of the 150-kd ternary complex. Serum levels capable of ternary complex formation are increased in insulin resistance.

Analyses and radioimmunoassays for IGFBP-3 reflect the altered affinity of IGFBP-3 fragments for IGF ligands, although some proteolytic fragments of IGFBP-3 are also produced.

The mature IGFBP-3 protein has a molecular weight of approximately 29 kd; however, because it is nonglycosylated, its affinity for IGF-I and IGF-II is decreased.

IGFBP-1 affinity for IGF-I and thereby inhibit IGF action. The ability of IGFBP-1 to inhibit or potentiate IGF action may depend on post-translational modifications of IGFBP-1, such as phosphorylation, which appears to enhance its ability to inhibit IGF-I.

Although most in vitro studies are consistent with an inhibitory effect of IGFBP-1 on IGF actions, presumably reflecting interference with IGF ligand-receptor interactions, IGFBP-1 potentiates IGF effects in certain cell systems, probably as the result of the binding of IGFBP-1 to cell membranes through its Arg-Gly-Asp (RGD) sequence. RGD is an integrin receptor recognition sequence that presumably allows IGFBP-1 to associate with the integrin (fibronectin) receptor. The ability of IGFBP-1 to inhibit or potentiate IGF action may depend on post-translational modifications of IGFBP-1, such as phosphorylation, which appears to enhance its ability for IGF-I and thereby inhibit IGF action.

Figure 23-33

The acute modulation of serum IGFBP-1 levels may regulate the free IGF-I pool and thereby influence the ability of IGF-I to bind to the type I IGF receptor (lanes 11 and 12 in Figure 23-33A). Otherwise, knockout of the IGFBP-2 gene or overexpression of IGFBP-1 in transgenic mice appears to have little effect on phenotype, possibly reflecting "redundancy" in the IGF system, in which one IGFBP can compensate for loss of another.

The existence of a low-molecular-weight IGFBP in cerebrospinal fluid was inferred from studies demonstrating a 34-kd IGFBP that did not react with antibodies to IGFBP-1 (or IGFBP-3). This IGFBP appeared to be consistent with a previous observation of CSF IGFBPs with preferential affinity for IGF-II.

Figure 23-34

Affinity cross-linking of [125I]IGF-I (A) and [125I]IGF-II (B) to membranes from HepG2 breast cancer cells. In the absence of unlabeled insulin-like growth factor (IGF) peptide (lane 1), IGF-I and IGF-II were not detected.

The mature IGFBP-3 protein has a molecular weight of approximately 28 kd; however, because it is nonglycosylated, it normally migrates as a doublet of 40 to 46 kd. Glycosylation does not alter its affinity for IGF-I or IGF-II. IGFBP-3 also undergoes serine phosphorylation of IGFBP-3, although its physiologic significance is uncertain. Perhaps the most significant post-translational modification of IGFBP-3 is proteolysis (see earlier). Discrepancies between immunoblot analyses and radioimmunoassays for IGFBP-3 reflect the altered affinity of IGFBP-3 fragments for IGF ligands, although some proteolytic fragments of IGFBP-3 are capable of inhibiting IGF action. In pregnancy serum levels of IGFBP-3 remain elevated, as do levels of IGFBP-3 in the conditioned media of prostatic epithelial cells. Interestingly, IGFBP-2 gene expression is markedly reduced in prostatic stromal cells from patients with benign prostatic hyperplasia, suggesting that IGFBP-2 may inhibit stromal growth.

Serum levels of IGFBP-2 are frequently elevated in patients with prostate carcinoma.

The mature IGFBP-3 protein is GH-dependent, due to either a direct GH effect or regulation by IGF. GH administration to hypophysectomized rats increases serum levels of IGFBP-3 in a manner similar to that seen with IGF-I administration. The protein was actually identified and purified from several different tissues, including amniotic fluid and Hep G2 conditioned media. Placental membranes (placental protein 12), and endometrium (pregnancy-associated 47k, -globulin). Its gene is 5.2 kb long, located on the short arm of chromosome 7, and composed of four exons. The mature protein is 30 kd and is nonglycosylated. mRNAs for IGFBP-3 are expressed in decidua (although not in placental trophoblasts), liver, and kidney.

IGFBP-1 may be involved in reproductive functions, including endometrial cycling, ovocyte maturation and fetal growth. It is the major IGFBP in fetal serum in early gestation, reaching levels as high as 3000 µg/L by the second trimester. Levels of IGFBP-1 in newborn serum are inversely correlated with birth weight, consistent with an inhibitory role on fetal IGF action.

IGFBP-1 also appears to have an important metabolic role, because its gene expression is enhanced in catabolic states, and serum levels undergo diurnal variation. Insulin suppresses and glucocorticoids enhance IGFBP-1 mRNA levels. Its gene is 5.2 kb long, located on the short arm of chromosome 7, and composed of four exons. The mature protein is 30 kd and is nonglycosylated. mRNAs for IGFBP-3 are expressed in decidua (although not in placental trophoblasts), liver, and kidney.

The IGFBP-2 is similar to IGFBP-1 in its lack of N-glycosylation and in the presence of an RGD sequence, perhaps allowing cell association and potentiation of IGF action. Nevertheless, knockout of the IGFBP-2 gene or overexpression of IGFBP-1 in transgenic mice appears to have little effect on phenotype, possibly reflecting "redundancy" in the IGF system, in which one IGFBP can compensate for loss of another.

The existence of a low-molecular-weight IGFBP in cerebrospinal fluid was inferred from studies demonstrating a 34-kd IGFBP that did not react with antibodies to IGFBP-1 (or IGFBP-3). This IGFBP appeared to be consistent with a previous observation of CSF IGFBPs with preferential affinity for IGF-II.

IGFBP-2 is expressed in secretory endometrium and endometrial tumors and is the major IGFBP in seminal fluid and in the conditioned media of prostatic epithelial cells. Interestingly, IGFBP-2 gene expression is markedly reduced in prostatic stromal cells from patients with benign prostatic hyperplasia, suggesting that IGFBP-2 may inhibit stromal growth.

Serum levels of IGFBP-2 are frequently elevated in patients with prostate carcinoma.

The IGFBP-3 gene is located on chromosome 7 in proximity to the gene for IGF-I. It contains four exons homologous to those of IGFBP-1 and IGFBP-2 and a fifth exon, consisting of 3' untranslated sequences. In all human tissues studied to date, a single 2.6-kb mRNA has been observed, whereas an additional 1.7-kb mRNA species suggests alternative splicing in baboons. mRNAs levels are high in liver, but IGFBP-3 is secreted by hepatic endothelia (portal venous and sinusoidal) and Kupffer cells, whereas ALS is synthesized in hepatocytes.

IGFBP-3 is GH-dependent, due to either a direct GH effect or regulation by IGF. GH administration to hypophysectomized rats increases serum levels of IGFBP-3. Serum levels of IGFBP-2 are frequently elevated in patients with prostate carcinoma. On the other hand, IGF-I treatment of patients with GHI does not alter serum IGFBP-3 levels, whereas GH treatment of GH-deficient patients does increase serum levels. Whether these observations mean that GH has a direct effect on IGFBP-3 or reflect GH regulation of ALS and ternary complex formation is unclear.

Figure 23-33

The effect of insulin-like growth factor binding protein 3 (IGFBP-3) proteolysis by prostate-specific antigen (PSA) on the ability of IGF-I to inhibit IGF-I (A) and IGF-II (B) action. (From Cohen P, Puehl DM, Graves HC, Rosenfeld RG. Biological effects of prostate-specific antigen as an insulin-like growth factor binding protein-3 protease. J Endocrinol 1994; 142:407415.)

Figure 23-34

Affinity cross-linking of [125I]IGF-I (A) and [125I]IGF-II (B) to membranes from HepG2 breast cancer cells. In the absence of unlabeled insulin-like growth factor (IGF) peptide (lane 1), IGF-I was markedly displaced by unlabeled IGF-I or IGF-II (lanes 2 to 5) but not by unlabeled IGF-II/insulin hybrid molecule (lanes 6 and 7) or by an IGF analogue with decreased affinity for IGFBP-binding proteins (QALY, lanes 9 and 10 in A). However, addition of [Leu27] IGF-II, which has decreased affinity for the type I IGF receptor (lanes 11 and 12 in A and lanes 7 and 9 in B), resulted in "unmasking" of the 130-kd subunit of the type I IGF receptor (A) and the 250-kd type II IGF receptor. (From Oh Y, Muller HL, Lamson G, Rosenfeld RG. Insulin-like growth factor [IGF] independent action of IGFBP-binding protein 3 in HepG2 human breast cancer cells: cell surface binding and growth inhibition. J Biol Chem 1993; 268:1496414971.)

The mature IGFBP-3 protein has a molecular weight of approximately 28 kd; however, because it is nonglycosylated, it normally migrates as a doublet of 40 to 46 kd. Glycosylation does not alter its affinity for IGF-I or IGF-II. IGFBP-3 also undergoes serine phosphorylation of IGFBP-3, although its physiologic significance is uncertain. Perhaps the most significant post-translational modification of IGFBP-3 is proteolysis (see earlier). Discrepancies between immunoblot analyses and radioimmunoassays for IGFBP-3 reflect the altered affinity of IGFBP-3 fragments for IGF ligands, although some proteolytic fragments of IGFBP-3 are capable of inhibiting IGF action. In pregnancy serum levels of IGFBP-3 remain elevated, as do levels of IGFBP-3 in the conditioned media of prostatic epithelial cells. Interestingly, IGFBP-2 gene expression is markedly reduced in prostatic stromal cells from patients with benign prostatic hyperplasia, suggesting that IGFBP-2 may inhibit stromal growth.

Serum levels of IGFBP-2 are frequently elevated in patients with prostate carcinoma. On the other hand, IGF-I treatment of patients with GHI does not alter serum IGFBP-3 levels, whereas GH treatment of GH-deficient patients does increase serum levels. Whether these observations mean that GH has a direct effect on IGFBP-3 or reflect GH regulation of ALS and ternary complex formation is unclear.

IGFBP-3 is the predominant IGFBP in adult serum, where it carries approximately 75% of the total IGF, primarily as part of the 150-kd ternary complex. Serum levels
are reduced in patients with GHD or GHI, conditions in which assays for serum IGFBP-3 have important diagnostic value (see later).

IGFBP-3 associates with cell membranes. Affinity cross-linking studies employing [125I]IGF-I and a human breast cancer cell line have demonstrated no binding to the type 1 IGF receptor but rather to membrane-associated 45-kd IGFBP-3 (see Fig. 23-34). When IGF analogues with selective affinity for IGFBPs were added, a typical 125I-subunit of the type 1 IGF receptor was uncovered, demonstrating that membrane-associated IGFBP-3, with its high affinity for IGF peptides, normally "mask" the IGF receptors. Oh and associates [24] demonstrated that the binding of IGFBP-3 to cell membrane proteins was specific, cation-dependent, and of high affinity. Whether these proteins constitute genuine IGFBP-3 receptors remains to be demonstrated, although they may mediate IGF-independent actions of IGFBP-3. Alternatively, IGFBP-3 may associate with heparin-containing proteoglycans both in the extracellular matrix and in the cell membrane, because both IGFBP-3 and IGFBP-5 contain heparin-binding consensus sequences in their COOH termini. However, treatment of cell monolayers with heparinase or chondroitinase has only minor effect on IGFBP-3 binding.

Like other IGFBPs, IGFBP-3 inhibits IGF action, especially when the binding protein is present in excess. Presumably, inhibition of IGF action by IGFBP-3 reflects a sequestering of IGF peptides away from the type 1 receptor. Proteolysis of IGFBP-3, resulting in a decrease in affinity for IGF ligands, decreases the inhibitory effects of the binding protein.

**IGF-Binding Protein 4**

The **IGFBP-4** gene, located on chromosome 17, contains four exons. A single 2.6-kb mRNA has been identified with high expression in liver. The protein is the smallest of the IGFBPs with 237 amino acids in humans, including 20 cysteines and one N-linked glycosylation site. In immunoblots of most biologic fluids, IGFBP-4 is a 24/28-kd doublet; deglycosylation eliminates the 28-kd band. IGFBP-4 appears to interact with connective tissues, but there is no evidence of membrane association, consistent with a primary role for IGFBP-4 as a soluble, extracellular IGFBP.

IGFBP-4 was initially isolated on the basis of its ability to inhibit IGF-stimulated cell proliferation in bone. There is no evidence for any IGF-potentiating effects. The inhibitory effects of IGFBP-4 are reduced by proteolysis of the protein, much as has been observed with IGFBP-3 degradation. IGFBP-4 proteases are produced by a wide variety of cells, including neuroblasts, smooth muscle, fibroblasts, osteoblasts, and prostatic epithelium. Activation of IGFBP-4 proteolysis occurs in the presence of IGF-I or IGF-II, presumably reflecting a conformational change in IGFBP-4 resulting from IGF occupancy. The clinical use of IGFBP-4 measurements is minor.

**IGF-Binding Protein 5**

Complementary DNAs for IGFBP-5 have been isolated and sequenced from rat ovary and human placenta and from a human osteosarcoma. The gene is located on chromosome 5 and contains four exons. A single 6.0-kb mRNA is expressed in a wide variety of tissues, particularly in kidney. Mature IGFBP-5 is produced as a 252-amino acid protein with no N-linked glycosylation sites but with one O-linked glycosylation site.

The addition of excess IGFBP-5 to human osteosarcoma cells inhibits IGF-stimulated DNA and glycosyn synthesis. However, when IGFBP-5 adheres to fibroblast extracellular matrix, it potentiates the growth-stimulatory effects of IGF on DNA synthesis. The affinity of IGFBP-5 for IGF-I is reduced approximately sevenfold when the binding protein is associated with extracellular matrix, providing a potential mechanism for release of IGFs to cell surface receptors. Association of IGFBP-5 with extracellular matrix also appears to protect it from proteolysis. Addition of IGFBP-5 to conditioned medium from fibroblasts results in proteolysis to a 21-kd fragment that does not potentiate IGF action, whereas the deposition of IGFBP-5 in extracellular matrix of fibroblasts makes it relatively resistant to degradation. Andreass and Bimbarm [55] purified a 23-kd IGFBP-5 fragment from U-2 osteosarcoma cells that has reduced affinity for IGFs but enhances IGF-stimulated mitogenesis. The 23-kd IGFBP-5 fragment stimulates mitogenesis in an IGF-independent manner, presumably by binding to a specific "receptor" on the cell membrane.

Unlike proteolysis of IGFBP-4, which is enhanced by addition of IGFs, degradation of IGFBP-5 is inhibited by the binding of IGF peptides. Proteolysis of IGFBP-5 results in the formation of 15- to 23-kd fragments demonstrated on immunoblots. Degradation of IGFBP-5 may have particular importance in the regulation of granulosa cell activity. In healthy ovarian follicles, neither IGFBP-4 nor IGFBP-5 is expressed, whereas both binding proteins are expressed in atretic follicles, thereby providing a mechanism for intrafollicular regulation

**IGF-Binding Protein 6**

The human IGFBP-6 gene is located on chromosome 12 and contains four exons. IGFBP-6 transcripts include a major 1.3-kb mRNA and a minor 2.2-kb transcript. The mature peptide contains 216 amino acids and has a molecular mass of approximately 23 kd, although it may migrate at a higher molecular weight on SDS gels, presumably reflecting O-glycosylation. Although IGFBP-6 binds both IGF-I and IGF-II, it has a significantly greater affinity for IGF-I. IGFBP-6 is found in relatively high levels in cerebrospinal fluid, as is also the case for IGFBP-2, which also binds IGF-II with selectively high affinity. IGFBP-6 may also have a role in regulating ovarian activity, perhaps by functioning as an antigonadotropin.
Gonadal Steroids

Although androgens and estrogens do not contribute substantially to normal growth before puberty, the adolescent rise
Thyroid Hormone

Thyroid hormone is a major contributor to postnatal growth, although, like GH, it is of relatively little importance to growth of the fetus. Hypothyroidism postnatally can cause profound growth failure and virtual arrest of skeletal maturation. In addition to a direct effect on epiphyseal cartilage, thyroid hormones appear to have a permissive effect on GH.

### TABLE 23-2 -- Classification of Growth Retardation

<table>
<thead>
<tr>
<th>I. Primary Growth Abnormalities</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Osteochondrodysplasias</td>
</tr>
<tr>
<td>B. Chromosomal abnormalities</td>
</tr>
<tr>
<td>C. Intrauterine growth retardation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>II. Secondary Growth Disorders</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Malnutrition</td>
</tr>
<tr>
<td>B. Chronic disease</td>
</tr>
<tr>
<td>C. Endocrine disorders</td>
</tr>
<tr>
<td>1. Hypothyroidism</td>
</tr>
<tr>
<td>2. Cushing's syndrome</td>
</tr>
<tr>
<td>3. Pseudohypoparathyroidism</td>
</tr>
<tr>
<td>4. Rickets</td>
</tr>
<tr>
<td>5. IGF deficiency</td>
</tr>
<tr>
<td>a. GHD due to hypothalamic dysfunction</td>
</tr>
<tr>
<td>b. GHD due to pituitary GH deficiency</td>
</tr>
<tr>
<td>c. GH resistance</td>
</tr>
<tr>
<td>(1) Primary GH insensitivity</td>
</tr>
<tr>
<td>(2) Secondary GH insensitivity</td>
</tr>
<tr>
<td>6. Primary defects of IGF synthesis</td>
</tr>
<tr>
<td>7. Primary defects of IGF transport and clearance</td>
</tr>
<tr>
<td>8. IGF insensitivity</td>
</tr>
<tr>
<td>(1) Defects of the type 1 IGF receptor</td>
</tr>
<tr>
<td>(2) Postreceptor defects</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>III. Idiopathic Short Stature</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Genetic short stature</td>
</tr>
<tr>
<td>B. Constitutional delay of growth and maturation</td>
</tr>
<tr>
<td>C. Heterozygous defects of the GH receptor</td>
</tr>
</tbody>
</table>

GH, growth hormone; GHD, growth hormone deficiency; IGF, insulin-like growth factor.

Patients with hypothyroidism have decreased spontaneous GH secretion and blunted responses to GH provocative tests. Treatment with thyroid hormone results in rapid "catch-up" (accelerated) growth, which is typically accompanied by marked skeletal maturation, potentially causing overly rapid epiphyseal fusion and compromise of adult height.
GROWTH RETARDATION

A classification of growth retardation is shown in Table 23-2. Growth disorders are subdivided into these categories:

1. Primary growth abnormalities, in which the defect(s) appears to be intrinsic to the growth plate.
2. Secondary growth disorders, or growth failure resulting from chronic disease or endocrine disorders (the newly introduced category of "IGF deficiency" can result from GHRH deficiency, GHD, or GH or IGF insensitivity).
3. ISS, including variants of normal (constitutional delay of growth and maturation [CDGM] and genetic short stature), heterozygous mutations of the GH receptor gene (GHR gene, a variant of GHI) and as yet unclarified mutations throughout the growth-related genome.
PRIMARY GROWTH ABNORMALITIES

Osteochondrodysplasias

The osteochondrodysplasias encompass a heterogeneous group of disorders characterized by intrinsic abnormalities of cartilage or bone, or both. These conditions share the following features:

1. Genetic transmission.
2. Abnormalities in the size and shape of bones of the limbs, spine, and/or skull.
3. Radiologic abnormalities of the bones (generally).

More than 100 osteochondrodysplastic conditions have been identified to date on the basis of physical features and radiologic characteristics, and biochemical, molecular, and genetic studies of these conditions will undoubtedly lead to the recognition of additional types. An international classification for the osteochondrodysplasias, developed in 1970 and revised in 1978 and 1992, is summarized in Table 23-3. Of note, the category of dysostosis has been dropped from the classification, which focuses on developmental disorders of bone and cartilage.

Diagnosis of osteochondrodysplasias can be difficult. Although the underlying molecular and biochemical defects have been identified in many of these conditions, clinical and radiologic evaluation remain central to the diagnosis. Frequently, the clinical features are characteristic, and the diagnosis can be made at birth or even prenatally by ultrasonography. The family history is critical, although many cases are due to fresh mutations, as is generally the case in the classical autosomal dominant achondroplasia and hypochondroplasia. Measurement of body proportions should include arm span, sitting height, upper and lower body segments, and head circumference. Clinical and radiologic evaluation should be used to determine whether involvement is of the long bones, skull, and vertebrae and whether abnormalities are primarily at the epiphyses, metaphyses, or diaphyses.

Two of the more common osteochondrodysplasias are achondroplasia and hypochondroplasia.

TABLE 23-3 – Classification of Osteochondrodysplasias

<table>
<thead>
<tr>
<th>I. Defects of the Tubular (and Flat) Bones and/or Axial Skeleton</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Achondroplasia group</td>
</tr>
<tr>
<td>B. Achondrogenesis</td>
</tr>
<tr>
<td>C. Spondyloepiphyseal dysplasia (perinatally lethal)</td>
</tr>
<tr>
<td>D. Metatropic dysplasia group</td>
</tr>
<tr>
<td>E. Short rib dysplasia group (with/without polydactyly)</td>
</tr>
<tr>
<td>F. Atelosteogenesis/diastrophic dysplasia group</td>
</tr>
<tr>
<td>G. Kniest-Stickler dysplasia group</td>
</tr>
<tr>
<td>H. Spondyloepiphyseal dysplasia congenita group</td>
</tr>
<tr>
<td>I. Other spondylo epiphyseal dysplasias</td>
</tr>
<tr>
<td>J. Dysostosis multiplex group</td>
</tr>
<tr>
<td>K. Spondyloepiphyseal dysplasias</td>
</tr>
<tr>
<td>L. Epiphyseal dysplasias</td>
</tr>
<tr>
<td>M. Chondrodysplasia punctata (stippled epiphyses) group</td>
</tr>
<tr>
<td>N. Metaphyseal dysplasias</td>
</tr>
<tr>
<td>O. Brachyrrachia (short spine dysplasia)</td>
</tr>
<tr>
<td>P. Mesomelic dysplasias</td>
</tr>
<tr>
<td>Q. Acro/acro-mesomelic dysplasias</td>
</tr>
<tr>
<td>R. Dysplasias with significant (but not exclusive) membranous bone involvement</td>
</tr>
<tr>
<td>S. Bent bone dysplasia group</td>
</tr>
<tr>
<td>T. Multiple dislocations with dysplasias</td>
</tr>
<tr>
<td>U. Osteodysplastic primordial dwarfism group</td>
</tr>
<tr>
<td>V. Dysplasias with increased bone density</td>
</tr>
<tr>
<td>W. Dysplasias with defective mineralization</td>
</tr>
<tr>
<td>X. Dysplasias with increased bone density</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>II. Disorganized Development of Cartilaginous and Fibrous Components of the Skeleton</th>
</tr>
</thead>
<tbody>
<tr>
<td>Achondroplasia</td>
</tr>
</tbody>
</table>

Achondroplasia is the most common of the osteochondrodysplasias, with a frequency of about 1:26,000. Although transmitted as an autosomal dominant disorder, 80% to 90% of cases appear to be due to new mutations. Achondroplasia is due to a mutation in a transmembranous domain of the gene for FGF receptor 3 (FGF-R3) located on the short arm of chromosome 4p16.3.

Most cases identified to date are caused by activating mutations at a "hot spot" at nucleotide 1138 (codon380, gly380arg) of the FGF-R3 gene and because these mutations create new recognition sites for restriction enzymes, they can be easily diagnosed. The mutation rate reported at this site indicates that it may be the most mutable gene in the human genome. The homogeneity of mutation in achondroplasia probably explains the minimal heterogeneity in its phenotype. Infants homozygous for this condition have severe disease, typically dying in infancy from respiratory insufficiency due to the small thorax. Diminished growth velocity is present from infancy, although short stature may not be evident until after 2 years of age. Mean adult heights in males and females are 130 and 120 cm, respectively. Growth curves for achondroplasia have been developed and are of value in following patients.

With increasing age, the diagnosis of achondroplasia becomes easier because these patients have characteristic abnormalities of the skeleton, including...
megaloecephaly, low nasal bridge, lumbar lordosis, short trident hand, and rhizomelia (shortness of the proximal legs and arms) with skin redundancy. Radiologic findings include small, cuboid-shaped vertebral bodies with short pedicles and progressive narrowing of the lumbar interpedicular distance. The iliac wings are small, with narrow sciatic notches. The small foramen magnum may lead to hydrocephalus, and spinal cord and/or root compression may result from kyphosis, stenosis of the spinal canal, or disc lesions. GH secretion in these children is comparable to that in normal children.

In a mouse with an equivalent FGF-R3 mutation showing many features of human achondroplasia, there is ligand-independent dimerization and phosphorylation of FGF-R3 with activation of STAT proteins and up-regulation of cell cycle inhibition. Additionally, such mutant mice also exhibit down-regulation of expression of the Indian hedgehog and PTHrP receptor genes, which are also involved in bone formation. As a result of the overexpression of this receptor activity, there is abnormal chondrogenesis and osteogenesis during endochondral ossification.

Hypochondroplasia

Hypochondroplasia is an autosomal dominant disorder, previously described as a "mild form" of achondroplasia, that frequently results from a mutation (Asn540Lys) in the FGF-R3 gene. The two disorders do not occur in the same family. About 70% of affected individuals are homozygous for a mutation in the FGF-R3 gene, but locus heterogeneity exists as other unidentified mutations cause a similar phenotype. Mullis and co-workers, using restriction enzyme analysis, suggested that the IGF-I gene may be a candidate gene for hypochondroplasia, but other molecular abnormalities are likely to be found.

The facial features of achondroplasia are absent, and both the short stature and rhizomelia are less pronounced. Adult heights typically are in the 120- to 150-cm range. In contrast with achondroplasia, poor growth may not be evident until after 2 years of age, but stature then deviates progressively from normal. Occasionally, the disproportionate short stature is not apparent until adulthood. Outward bowing of the legs may be accompanied by genu varum. Lumbar interpedicular distances diminish between L1 and L5, and, as with achondroplasia, there may be flaring of the pelvis and narrow sciatic notches. The diagnosis is exceedingly difficult to make in young children. Mild variants of the syndrome may not be clinically distinguishable from normal, and radiologic studies should be performed if a question arises.
Mild intrauterine growth retardation (IUGR), with mean birth weights and lengths of 2800 g and 48.3 cm, respectively. Clinical trials of GH therapy in such cases are currently in progress.

Deletion of the long arm of chromosome 18 has an estimated prevalence of 1 in 40,000 live births. In a review of 50 cases, 64% of children (mean age 5.8 ± 4.5 years) had heights greater than 2 SD below the mean, with only 6% greater than 0 SDS. Fifteen percent had serum IGF-I concentrations, and 9% had IGFBP-3 concentrations below -2 SD. Seventy-two percent of children had reduced GH responses to provocative testing, although such testing was not always rigorous. Clinical trials of GH therapy in such cases are currently in progress.

**Chromosomal Abnormalities**

Abnormalities of autosomes or sex chromosomes may cause growth retardation, frequently associated with somatic abnormalities and mental retardation, as in deletion of chromosome 5 or trisomy 16 or 13. Such abnormalities, however, may be subtle, and the diagnosis of Turner's syndrome must be considered in any girl with unexplained short stature. In many cases, the precise cause of growth failure is not clear because the genetic defects do not affect known components of the GH-IGF system. The chromosomal lesion may directly influence normal tissue growth and development or, indirectly, modulate local responsivity to IGF.

**Down's Syndrome**

Trisomy 21, or Down's syndrome, is probably the most common chromosomal disorder associated with growth retardation, affecting approximately 1 in 600 live births. On average, newborns with Down's syndrome have birth weights 500 g below normal and are 2 to 3 cm shorter. Growth failure continues postnatally and is typically associated with delayed skeletal maturation and a delayed and incomplete pubertal growth spurt. Adult heights range from 135 to 170 cm in men and 127 to 158 cm in women. The cause of growth failure in Down's syndrome and in other autosomal defects is unknown.

Attempts to find underlying hormonal explanations for growth retardation have been unsuccessful, even though hypothryoidism due to Hashimoto's thyroiditis is more common than normal in Down's syndrome and should be sought. Marginal levels of GH secretion and low serum levels of IGF-I have been reported in Down's syndrome, and exogenous GH may be efficacious in the short term. It is more likely, however, that the growth failure reflects a generalized biochemical abnormality of the epiphyseal growth plate.

**Gonadal Dysgenesis**

In girls with gonadal dysgenesis (Turner's syndrome), short stature is the single most common feature, occurring more frequently than delayed puberty, cubitus valgus, or webbing of the neck. In large series of such individuals, short stature occurs in 95% to 100% of girls with a 45,X karyotype. Several distinct phases of growth have been identified in girls with Turner's syndrome:

1. Mild intrauterine growth retardation (IUGR), with mean birth weights and lengths of 2800 g and 48.3 cm, respectively.
2. Slow growth during infancy falling to -3 SDS by 3 years of age.
3. Delayed onset of the "childhood phase" of growth and progressive decline in height velocity from 3 years of age until approximately 14 years of age, resulting in further deviation from normal height percentiles.
4. A prolonged adolescent growth phase, characterized by a partial return toward normal height, followed by delayed epiphyseal fusion.

Mean adult heights in the United States and Europe range from 142.0 to 146.8 cm (lower in Asia). There are important genetic and ethnic influences on growth in these girls. Parental height correlates well with final patient height, and a cross-cultural study in 15 countries demonstrated a strong correlation between adult height and midparental height. In girls with gonadal dysgenesis and Turner's syndrome, adult heights range from 135 to 167 cm in women. Parental height correlates well with final patient height, and a cross-cultural study in 15 countries demonstrated a strong correlation between adult height and midparental height.

**(r = 0.91)** between final height in Turner's syndrome and in the normal population with an approximate 20-cm deficit.

The cause of growth failure in Turner's syndrome remains unclear. Girls have a skeletal dysplasia and are haploinsufficient for the SHOX gene (short stature homeobox-containing gene) located in the pseudoautosomal region of the short arm of the X chromosome. Mutations and deletions of this gene are associated with poor height growth and several syndromes of skeletal dysplasia including Madelung's deformity, which is also seen in Turner's syndrome. The incidence of IUGR is much greater in girls lacking two copies of the SHOX gene (46% versus 7%).

Most patients have normal GH and IGF levels during childhood; reports of low GH or IGF levels in adolescents with Turner's syndrome are likely due to low serum levels of gonadal steroids. Growth impairment is evident prior to the period when activity of the GH-IGF axis is decreased. Nevertheless, GH therapy is capable of both accelerating short-term growth and increasing adult height. This diagnosis must be considered in all girls with unexplained growth failure and especially in girls who are short for family but are growing between the 5th and 10th percentiles in the first decade of life. Nonetheless, a recent study found that the mean age at diagnosis lagged 5.3 years behind the age at which patients with Turner's syndrome fell below the 5th percentile, where its frequency is approximately 1/1000. Such data affirm the need for vigorous assessment of all girls who are either absolutely short or relatively small for family heights.

**18q Deletions**

Deletion of the long arm of chromosome 18 has an estimated prevalence of 1 in 40,000 live births. In a review of 50 cases, 64% of children (mean age 5.8 ± 4.5 years) had heights greater than 2 SD below the mean, with only 6% greater than 0 SDS. Fifteen percent had serum IGF-I concentrations, and 9% had IGFBP-3 concentrations below -2 SD. Seventy-two percent of children had reduced GH responses to provocative testing, although such testing was not always rigorous. Clinical trials of GH therapy in such cases are currently in progress.
Intrauterine Growth Retardation

Infants with IUGR comprise a heterogeneous group with birth weight and/or length below the 3rd or 10th percentile for gestational age, depending on the study. They may also be referred to as small-for-gestational-age (SGA) infants, in contrast with those who are appropriate for gestational age (AGA). The importance of this distinction, in addition to a number of issues influencing neonatal morbidity, is in the prediction of later growth: Most AGA low-birth-weight infants experience catch-up growth during the first 2 years of life, in contrast with the slower, attenuated growth of SGA infants who may have persistent height deficits throughout childhood and adolescence. First-trimester growth failure has been closely associated with low birth weight and low-birth-weight percentile. The earlier in gestation that fetal growth is impaired, the less likely that complete recapture of lost growth will occur.

IUGR can arise from abnormalities in the fetus, the placenta, or the mother (Table 23-4). Factors affecting fetal growth include nutrition provided by the maternal-placental system, alterations of fetal IGF production, and as yet unclarified genes. Although it is understandable why uterine constraint or twin pregnancies might result in limited fetal growth, the reason for abnormal fetal growth in most cases of IUGR is unclear.

### TABLE 23-4 -- Causes of Intrauterine Growth Retardation

<table>
<thead>
<tr>
<th>I. Intrinsic Fetal Abnormalities</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Chromosomal disorders</td>
</tr>
<tr>
<td>B. Syndromes associated with primary growth failure</td>
</tr>
<tr>
<td>1. Russell-Silver syndrome</td>
</tr>
<tr>
<td>2. Seckel’s syndrome</td>
</tr>
<tr>
<td>3. Noonan’s syndrome</td>
</tr>
<tr>
<td>4. Progeria</td>
</tr>
<tr>
<td>5. Cockayne’s syndrome</td>
</tr>
<tr>
<td>6. Bloom’s syndrome</td>
</tr>
<tr>
<td>7. Prader-Willi syndrome</td>
</tr>
<tr>
<td>8. Rubenstein-Taybi syndrome</td>
</tr>
<tr>
<td>C. Congenital infections</td>
</tr>
<tr>
<td>D. Congenital anomalies</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>II. Placental Abnormalities</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Abnormal implantation of the placenta</td>
</tr>
<tr>
<td>B. Placental vascular insufficiency; infarction</td>
</tr>
<tr>
<td>C. Vascular malformations</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>III. Maternal Disorders</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Malnutrition</td>
</tr>
<tr>
<td>B. Constraints on uterine growth</td>
</tr>
<tr>
<td>C. Vascular disorders</td>
</tr>
<tr>
<td>1. Hypertension</td>
</tr>
<tr>
<td>2. Toxemia</td>
</tr>
<tr>
<td>3. Severe diabetes mellitus</td>
</tr>
<tr>
<td>D. Uterine malformations</td>
</tr>
<tr>
<td>E. Drug ingestion</td>
</tr>
<tr>
<td>1. Tobacco</td>
</tr>
<tr>
<td>2. Alcohol</td>
</tr>
<tr>
<td>3. Narcotics</td>
</tr>
</tbody>
</table>

The implications of IUGR may extend beyond fetal life. Although most SGA infants exhibit catch-up growth by 2 years of age, a large subgroup remains small. In a retrospective study of 47 individuals who had IUGR, 23 men had an adult height of 162 cm and 24 women had an adult height of 148 cm. Larger studies demonstrated that SGA children had a fivefold to sevenfold greater chance of short stature than AGA children did. Ten percent to 15% of SGA infants will have short stature, and this group makes up as much as 20% of all short children. In a study of a more severely affected neonatal intensive care unit SGA population, 27% had not yet achieved catch-up by 6 years of age. Final adult height is -0.8 to -0.9 SDS, which is a mean deficit of 3.6 to 4 cm when adjusted for family stature. The endocrinologic mechanisms of the poor growth are varied but may include abnormalities of GH production and secretory patterns and insensitivity to GH and IGF-I action.

The childhood and adolescent endocrine disorders associated with the SGA children include premature adrenarche, insulin resistance, functional ovarian hyperandrogenism, and an attenuated pubertal growth spurt. Furthermore, SGA infants have an increased risk of hypertension, maturity-onset diabetes mellitus, and cardiovascular disease later in life. Not all of these problems appear to occur in those IUGR babies who do not have catch-up growth, although insulin resistance has been described. Whether IUGR is causally related to these disorders or is a symptom of an underlying inborn metabolic disorder is not yet known.

In contrast with the role of the endocrine system in postnatal growth, intrauterine growth is less dependent on fetal pituitary hormones. Athyreotic and agonadal infants are of normal length and weight at birth. Pituitary GH is synthesized and secreted by the latter half of the first trimester, with midgestational levels peaking at 150 µg/L and then falling to around 30 µg/L at term. Because the anencephalic fetus is normal in size, the pituitary was thought to be unnecessary for fetal growth. However, documentation of birth size of rats and humans with congenital GHD and of human newborns with mutations of the GH or GHR genes indicate that GH from the fetal pituitary makes a small contribution to birth size. Infants with neonatal GHD are around -0.5 to

---

**Intrinsic Fetal Factors**
Other syndromes associated with moderate to profound growth failure include Bloom's syndrome, de Lange's syndrome, leprechaunism (mutations of the insulin syndrome caused by hypothalamic dysfunction. Hypertension, and congestive heart failure. Skeletal hypoplasia results in severe growth retardation, which usually becomes evident by 6 to 18 months of age. Subcutaneous fat, accompanied by alopecia, hypoplasia of the nails, joint limitation, early onset of atherosclerosis, typically followed by angina, myocardial infarction, The senile appearance characteristic of progeria be delayed or incomplete. Mental retardation of variable degrees is present in approximately 25% to 50% of patients. Although birth weight is generally within the normal range, mean growth in length and weight is below the 3rd percentile through much of childhood with a falling height velocity not dissimilar to that seen in Turner's syndrome except for the late and attenuated pubertal increments. GH secretory abnormalities do not account for the short stature, although endogenous GH production may be reduced somewhat. Exogenous IGF-I administration increases neonatal growth rate and protein and fat accretion in pigs with IUGR. Hepatic levels of GHR mRNA and of GHR are low, perhaps explaining the modest impact of GH on IGF production and linear growth. In neonates with IUGR, GH levels are elevated, and exogenous GH treatment has little or no effect on growth, body composition, or energy expenditure, further supporting a state of relative insensitivity to GH at this developmental stage. With defects of the GHR, neonatal IGF levels are low, suggesting a role for GH in regulating IGF production. Similarly, IGFBP-1's are identifiable in serum and other biological fluids in the fetus and newborn. However, serum levels of IGFBP-3 and ALS, the major serum carriers of IGF peptides in the adult, are low in the fetus and newborn. Thus, the components of the IGF system are apparently regulated directly by glucose levels or indirectly by fetal insulin secretion, with less impact of GH levels.

The role of insulin production in fetal growth is demonstrated by somatic overgrowth of the hyperinsulinemic infants of mothers with diabetes and of infants with the syndrome of persistent neonatal hyperinsulinemic hypoglycemia (nosedobilastosis). In contrast, infants with pancreatic agenesis or with abnormalities of the insulin receptor in the "leprechaun" syndrome are SGA. Further, the inverse relationship of insulin and IGFBP-1 levels and the finding that fetal IGFBP-1 levels are elevated in IUGR support an important role for insulin in fetal growth regulation. In addition to these well-characterized endocrine profiles, cord blood cortisol levels and neonatal insulin levels are inversely related to IGF-I and directly to IGFBP-3 concentrations. In infants with IUGR, a close correlation (r = -0.54) was observed between cord blood cortisol levels and length growth during the first 3 months of life. This is a period during which substantial catch-up growth occurs in some, but clearly not all, IUGR infants.

Infants with IUGR exhibiting poor postnatal growth, particularly when the abnormalities are intrinsic to the fetus, have frequently been categorized as having "primordial growth failure." Several syndromes are briefly noted in the following sections.

**Russell-Silver Syndrome**

Russell-Silver syndrome (RSS) is a condition that was independently described by Russell and by Silver and associates. Although this syndrome is probably due to a heterogeneous group of disorders, the common findings include IUGR, postnatal growth failure, congenital hemihypertrophy, and small, triangular facies. Nonspecific findings include clinodactyly, precocious puberty, delayed closure of the fontanelles, and delayed bone age. Adults are short, with final heights about 4 SD below the mean. Endogenous GH secretion in prepubertal children with RSS is similar to that in other short IUGR children and less than in AGA short children.

Because no genetic or biochemical basis for this disorder has been identified, RSS is often used incorrectly as a designation for IUGR of unknown etiology. Maternal uniparental disomy of chromosome 7 exists in 7% to 10% of cases. Candidate genes in chromosome region 7p11-13, such as those for IGFBP-1, IGFBP-3, and IGF-R, all show biallelic expression, but a growth suppression gene, GB10, which binds to the insulin and IGF-I receptors and replicates asynchronously, remains a candidate gene for overexpression. Paternally expressed imprinted genes at 7q32, PEG/MEST, and g2-COP are other candidates.

**Seckel's Syndrome**

Originally described by Mann and Russell in 1959. The condition most commonly termed Seckel's syndrome is also known as Seckel's bird-headed dwarfism. The syndrome is an autosomal recessive disorder characterized by IUGR and severe postnatal growth failure, combined with microcephaly, prominent nose, and micrognathia. Final height is typically 90 to 110 cm, with moderate to severe mental retardation. The nature of the underlying defect is unknown, although the gene defect may be at 3p22.1-q24.

**Noonan's Syndrome**

Although Noonan's syndrome shares certain phenotypic features with Turner's syndrome, the two disorders are clearly distinct. In Noonan's syndrome, the sex chromosomes are normal and transmission is apparently autosomal dominant; neither the gene locus nor product has been identified, although linkage with chromosome 12 has been demonstrated. Both males and females may be affected, which may explain the misleading terms Turner-like syndrome and male Turner's syndrome. Affected individuals typically have webbing of the neck, a low posterior hairline, ptosis, cubitus valgus, and malformed ears. Cardiac abnormalities are primarily right-sided (pulmonary valve) rather than the left-sided lesions (aorta, aortic valve) characteristic of Turner's syndrome.

Although birth weight is generally within the normal range, mean growth in length and weight is below the 3rd percentile through much of childhood with a falling height velocity not dissimilar to that seen in Turner's syndrome except for the late and attenuated pubertal increments. GH secretory abnormalities do not account for the short stature, although endogenous GH production may be reduced somewhat. Microphallus and cryptorchidism are common, and puberty may be delayed or incomplete. Mental retardation of variable degrees is present in approximately 25% to 50% of patients.

**Progeria**

The senile appearance characteristic of progeria (Hutchinson-Gilford syndrome) is generally apparent by 2 years of age. There is a progressive loss of subcutaneous fat, accompanied by alopecia, hypoplasia of the nails, joint limitation, early onset of atherosclerosis, typically followed by angina, myocardial infarction, hypertension, and congestive heart failure. Skeletal hypoplasia results in severe growth retardation, which usually becomes evident by 6 to 18 months of age.

**Cockayne's Syndrome**

Cockayne's syndrome, like progeria, is characterized by a premature senile appearance. Retinal degeneration, photosensitivity of the skin, and impaired hearing may also be present. Growth failure typically appears at 2 to 4 years of age. Transmission is as an autosomal recessive disorder.

**Prader-Willi Syndrome**

Growth failure in Prader-Willi syndrome may be evident at birth and is more impressive postnatally. It is considered at length in the discussion of IGF deficiency syndrome caused by hypothalamic dysfunction. Other syndromes associated with moderate to profound growth failure include Bloom's syndrome, de Lange's syndrome, leprechaunism (mutations of the insulin receptor gene), Ellis-van Creveld syndrome, Aarskog's syndrome, Rubinstein-Taybi syndrome, multiple endocrine (Perner's syndrome), Dubowitz's syndrome, and Johanson-Blizzard syndrome. It is interesting that the gene for the ghrelin receptor is close to the mapped location of de Lange's syndrome.
Maternal and Placental Factors

Maternal factors and placental insufficiency can impair fetal growth. Although such affected infants have better growth potential compared with infants with "primordial growth failure," postnatal growth is not always normal. Maternal nutrition is an important contributor to fetal growth and to the child's growth during the first year of life. Fetal growth retardation may also result from alcohol consumption during pregnancy and from use of cocaine, marijuana, and tobacco. The mechanisms of such drug-induced fetal growth retardation are unclear but probably include uterine vasoconstriction and vascular insufficiency, placental abruption, and premature rupture of membranes. Although maternal tobacco use is, statistically, a major contributor to reduced fetal size, it is unlikely, by itself, to result in severe IUGR.

The maternal hormonal milieu is affected by placental steroids and peptides, especially placental GH and human placental lactogen (hPL), also called human chorionic somatomammatropin (hCS), which influence the production of maternal IGF-I. Maternal IGF affects placental function and may facilitate transport of nutrients to the fetus and maternal IGF-I levels correlate with fetal growth. Hasegawa and colleagues have found increased levels of free (non-IGFBP bound) IGF-I levels during normal human pregnancy, possibly due to accelerated proteolysis of IGFBP-3.

The placenta has multiple functions, including the transport of nutrients, oxygen, and waste and production of hormones, and it consumes oxygen and glucose brought to it by the uterine circulation. Placental GH affects maternal IGF production that in turn affects placental function. The finding of normal levels of placental GH and IGF-I in a woman with Pit-1 deficiency supports the importance of placental GH action. Additionally, ghrelin message and peptide are present in human placentae during the first trimester.

hPL is a major regulator of glucose, amino acid, and lipid metabolism in the mother, aiding in the mobilization of nutrients for transport into the fetus. Damage to the placenta by vascular disease, infection, or intrinsic abnormalities of the syncytiotrophoblasts can impair these important functions. Sometimes, but not always, examination of the placenta can reveal diagnostic information as to the pathogenesis of IUGR. An X-linked homeobox gene, Esx1, detected only in extraembryonic tissues and human testis, is a chromosomally imprinted regulator of placental morphogenesis. Heterozygous and homozygous mutant mice were born 20% smaller than normal and had large edematous placentae. Vasculature was abnormal at the maternal-fetal interface, presumably causing the growth retardation.
SECONDARY GROWTH DISORDERS

Malnutrition

Given the worldwide presence of undernutrition, it is not surprising that inadequate caloric and protein intake is the most common cause of growth failure. Marasmus refers to cases with an overall deficiency of calories, including protein malnutrition. Subcutaneous fat is minimal, and protein wasting is marked. Kwashiorkor refers to inadequate protein intake, although it may also be characterized by some caloric undernutrition. In both conditions, multiple deficiencies of vitamins and minerals are apparent. Frequently, the two conditions overlap. Decreased weight growth generally precedes the failure of linear growth by a very short time in the neonatal period and by several years at older ages. Stunting of growth in early life has life-long consequences, resulting in diminished height growth.

Both acute and chronic malnutrition affects the GH-IGF system. The impaired growth is usually associated with elevated basal and stimulated serum GH levels, but in generalized malnutrition (marasmus) GH levels may be normal or low. In both conditions, serum IGF-I levels are reduced. Malnutrition may consequently be considered a form of GHI, with serum IGF-I levels reduced despite normal or elevated GH levels. GHBP levels, as a reflection of GHR content, are decreased. GH may be an adaptive response, whereby protein is spared by the lipolytic and anti-insulin actions of GH. Reduced serum IGF-I levels would serve to shift calories from anabolic to survival requirements. These adaptive mechanisms are accompanied by changes in serum IGFBPs to further limit IGF action during periods of malnutrition.

Inadequate calorie and protein intake complicates many chronic diseases that are characterized by growth failure. Anorexia is a common feature of renal failure and inflammatory bowel disease and occurs with cyanotic heart disease, congestive heart failure, CNS disease, and other illnesses. Some of these conditions, furthermore, may be characterized by deficiencies of specific dietary components, such as zinc, iron, and vitamins, necessary for normal growth and development.

Undernutrition may also be voluntary, as with dieting and food fads. Caloric restriction is especially common in girls during adolescence, when it may be associated with anxiety concerning obesity, and in gymnasts and ballet dancers. Anorexia nervosa and bulimia are extremes of "voluntary" caloric deprivation and are commonly associated with impaired growth, prior to epiphyseal fusion, that may result in diminished final adult height. Adolescent bone mineral accretion is impaired and significant osteopenia may persist into adulthood. Later in adolescence, malnutrition may cause delayed puberty and menarche and a variety of metabolic alterations. In anorexia nervosa, hormonal profiles are similar to those in protein-energy malnutrition, with high basal levels of GH but low levels of IGF-1, IGFBP-3, and GHBP. GHBP and IGFBP-3 levels correlate with body mass index, as in normal children. The hormones of the GH-IGF axis return to normal levels with refeeding.
Chronic Diseases

Malabsorption and Gastrointestinal Diseases

Intestinal disorders that impair absorption of calories or protein cause growth failure, for many of the reasons cited earlier. Growth retardation may predate other manifestations of malabsorption and chronic inflammatory bowel disease. Accordingly, celiac disease (gluten-induced enteropathy) and regional enteritis (Crohn's disease) should be considered in the differential diagnosis of unexplained growth failure. Serum levels of IGF-I may be reduced, reflecting the malnutrition, and it is crucial to discriminate between these conditions and GHD or related disorders causing IGF-I deficiency. Documentation of malabsorption requires demonstration of fecal wasting of calories, especially fecal fat, along with other measures of gut dysfunction such as the D-xylene or breath hydrogen studies.

In celiac disease (Fig. 23-42), impaired linear growth may be the first manifestation of disease. In European studies, celiac disease is the cause of unexplained growth impairment in 5% to 20% of unselected patients. The onset and progression of puberty may be delayed, and menarche may be late. Accordingly, a screening test for celiac disease is needed to obviate the standard diagnostic tests involving multiple intestinal biopsies in assessments of asymptomatic patients.

Both immunoglobulin G (IgG) and IgA antiendomysial antibodies have relatively high sensitivity and specificity but have been largely supplanted by IgA tissue transglutaminase antibodies. IgA deficiency is the most common immunodeficiency in assessments of antibody data. Nonetheless, the diagnosis of celiac disease ultimately requires demonstration of the characteristic mucosal flattening in small bowel biopsy. Changes in the serologic profiles mirror the clinical status obviating the need for subsequent biopsies. Gluten withdrawal is a highly effective treatment for celiac disease and results in rapid catch-up growth and decreased clinical symptoms during the first 6 to 12 months of treatment. Low IGF-I and IGFBP-3 levels return to normal during this period. Most children who receive appropriate dietary management ultimately achieve a normal final height.

Growth failure in Crohn's disease is probably due to a combination of malnutrition from malabsorption and anorexia, chronic inflammation, inadequacy of trace minerals in the diet, and use of glucocorticoids. IGF-I levels are low, especially with impaired growth. One third to two thirds or more of children with Crohn's disease have impaired growth at diagnosis, and occasional patients have significant growth failure as the first evidence of Crohn's disease. Osteopenia is common. An elevated erythrocyte sedimentation rate, anemia, and low serum albumin are useful clues, but diagnosis ultimately requires colonoscopy and biopsy along with gastrointestinal imaging studies.

Long-term treatment includes enteral and parenteral nutrition, anti-inflammatory agents, alternate-day steroid therapy, and judicious operative intervention. Newer therapeutic alternatives may include GH. Permanent impairment of linear growth and deficits of final height may occur in 30% of patients.

Chronic Liver Disease

Impaired linear growth and short stature with chronic liver disease in childhood are caused by decreased food intake, fat and fat-soluble vitamin malabsorption, trace element deficiencies, and abnormalities of the GH-IGF system. Decreased levels of IGF-I, IGFBP-3, and increased GH secretion define the acquired GH syndrome.

A close correlation between GH-dependent peptides and liver function indicates the dominant regulatory role of damaged hepatocytes in end-stage liver failure. Low levels of full-length and truncated GHRs in the cirrhotic liver and consequent diminished release of GHB P substantiate the GHI. Despite provision of adequate calories, insensitivity to the action of GH persists.

Liver transplantation prolongs life expectancy, but linear growth is variably improved in the early post-transplantation years. Exogenous glucocorticoid administration presumably plays a major role in the continued growth retardation; GH and IGFBP-3 production are normal, but the amount of "free IGF" may be decreased because IGFBP-3 levels are relatively high. Post-transplantation growth is inversely correlated with age and directly correlated with degree of growth improvement at transplantation. Exogenous GH treatment, for a period of 18 months, enhances growth rates and increases median height SDS by 0.7 unit.

Cardiovascular Disease

Congenital heart disease with cyanosis or chronic congestive failure can cause growth failure. As many as 27% of children with varied cardiac lesions were below the 3rd percentile for height and weight in one survey, and 70% were lower than the 50th percentile in another. Because cardiac defects are usually congenital, many infants have dysmorphic features and IUGR.

Inadequate cardiac intake is the most common cause of growth impairment in children with congenital heart disease frequently associated with anorexia and vomiting. Chronic congestive heart failure is associated with malabsorption that includes protein-losing enteropathy, intestinal lymphangiectasia, and steatorrhea. Greater cardiac and respiratory work and the relatively higher ratio of metabolically active, energy-utilizing brain and heart to the growth-related body mass (cardiac cachexia) causes an increased basal metabolic rate in these children. Food intake that appears adequate for the child's weight is thus inadequate for normal growth. The degree of cyanosis or hypoxia does correlate with the degree of growth impairment. Decreased levels of IGF-I and IGFBP-3 and normal levels of GH and hepatic GHRs in chronically hypoxic newborn sheep suggest GHI distal to the GHR. Linear growth and pubertal maturation depend on left ventricular function in children and adolescents with complicated rheumatic heart disease.

Corrective surgery may restore normal growth, frequently after a phase of catch-up growth with normalization of energy expenditure. Surgery must, on occasion, be delayed until the infant reaches an appropriate size, resulting in the conundrum that surgery corrects growth failure but cannot be performed because the infant is too small. In these situations, meticulous attention to caloric support and alleviation of hypoxia and heart failure is necessary to promote growth prior to surgery. Fortunately, this problem has diminished because of operative successes in the neonatal period.

The nutritional management of these infants includes calorie-dense feedings, because of the need to restrict fluids; calcium supplementation, because of the use of...
diuretics that may cause calcium loss in the urine, and iron, to maintain an enhanced rate of erythropoiesis.

**Renal Disease**

All conditions that impair renal function can impair growth. Uremia and renal tubular acidosis can cause growth failure before other clinical manifestations become evident. The growth impairment results from multiple mechanisms, including inadequate formation of 1,25-dihydroxycholecalciferol (1,25 [OH]2 D) with resultant osteopenia, decreased calcitic intake, loss of electrolytes necessary for normal growth, metabolic acidosis, protein wasting, insulin resistance, chronic anemia, and compromised cardiac function as well as from impaired GH and IGF production and action. In nephrotic cytostasis, acquired hypothyroidism contributes to the inadequate growth. Sixty percent to 75% of patients with chronic renal failure treated prior to the GH therapeutic era had a final adult height more than -2 SD below the mean.

Children and adolescents have normal or elevated circulating levels of GH, depending on the degree of renal failure. In children with end-stage renal disease (ESRD) on dialysis or preterminal chronic renal failure (CRF), the half-life of GH is prolonged twofold. The number of secretory bursts was increased in ESRD by twofold to threefold over chronic renal failure and controls and mean levels of GH were 2.5-fold higher in ESRD than in chronic renal failure patients or controls. Overall, children with ESRD produced substantially more GH than either chronic renal failure patients or controls. Early reports of decreased serum IGF-I levels in uremia were an artifact due to inadequate separation of IGF from IGFBPs prior to assay. Decreased hepatic IGF production is possibly due to low hepatic irinocortisol gene expression. Additionally, the uremic state may cause a postreceptor defect in GH signal transduction by diminishing phosphorylation and nuclear translocation of GH-activated STAT proteins. Serum IGF-I and -II levels are, however, usually normal, but increases in serum IGFBPs, especially IGFBP-1, may inhibit GH action.

In patients with nephrotic syndrome, serum levels of IGF-I and IGFBP-3 are low because of urinary loss of IGF-IGFBP complexes. Chronic glucocorticoid therapy for a variety of renal disorders can exacerbate growth retardation by diminishing GH release and blunting IGF-I action at growth plates. Chronic renal disease, especially ESRD, with increased GH levels and production, low levels of IGF-I, and poor growth is thus a state of relative resistance to GH and, in some instances, to IGF-I. Overall, serum IGF-I levels in ESRD may be normal. In the large cohort of patients in the North American Pediatric Renal Transplant Cooperative Study, the mean height increased after transplantation by only 0.11 SD in the first 4.5 years. The youngest age group (≤2 years) had the largest deficit and the most catch-up growth (0.94 SD); children between 6 and 17 years of age showed no improvement in SD score. Although the mortality rate in the younger age group was 16%, more recent studies suggest that transplantation may have an acceptable risk in such patients. Growth-retarded post-transplantation children, whether receiving daily or alternate-day glucocorticoid treatment, have decreased GH secretion, normal levels of IGF-I and IGFBP-1, and increased levels of IGFBP-3. They differ from patients with ESRD in that IGFBP-1 levels are not strikingly elevated, perhaps because of altered glucose tolerance and hyperinsulinism due to chronic glucocorticoid therapy.

Even after successful renal transplantation, growth height may not be normal. After successful renal transplantation, mean height increased after transplantation by only 0.11 SD in the first 4.5 years. The youngest age group (<2 years) had the least growth. The number of secretory bursts was increased in ESRD by twofold to threefold over chronic renal failure and controls and mean levels of GH were 2.5-fold higher in ESRD than in chronic renal failure patients or controls. Overall, children with ESRD produced substantially more GH than either chronic renal failure patients or controls. Early reports of decreased serum IGF-I levels in uremia were an artifact due to inadequate separation of IGF from IGFBPs prior to assay. Decreased hepatic IGF production is possibly due to low hepatic irinocortisol gene expression. Additionally, the uremic state may cause a postreceptor defect in GH signal transduction by diminishing phosphorylation and nuclear translocation of GH-activated STAT proteins. Serum IGF-I and -II levels are, however, usually normal, but increases in serum IGFBPs, especially IGFBP-1, may inhibit GH action.

In patients with nephrotic syndrome, serum levels of IGF-I and IGFBP-3 are low because of urinary loss of IGF-IGFBP complexes. Chronic glucocorticoid therapy for a variety of renal disorders can exacerbate growth retardation by diminishing GH release and blunting IGF-I action at growth plates. Chronic renal disease, especially ESRD, with increased GH levels and production, low levels of IGF-I, and poor growth is thus a state of relative resistance to GH and, in some instances, to IGF-I. Overall, serum IGF-I levels in ESRD may be normal. In the large cohort of patients in the North American Pediatric Renal Transplant Cooperative Study, the mean height increased after transplantation by only 0.11 SD in the first 4.5 years. The youngest age group (≤2 years) had the largest deficit and the most catch-up growth (0.94 SD); children between 6 and 17 years of age showed no improvement in SD score. Although the mortality rate in the younger age group was 16%, more recent studies suggest that transplantation may have an acceptable risk in such patients. Growth-retarded post-transplantation children, whether receiving daily or alternate-day glucocorticoid treatment, have decreased GH secretion, normal levels of IGF-I and IGFBP-1, and increased levels of IGFBP-3. They differ from patients with ESRD in that IGFBP-1 levels are not strikingly elevated, perhaps because of altered glucose tolerance and hyperinsulinism due to chronic glucocorticoid therapy.

Overall, height SDS at the time of transplantation and use of alternate-day glucocorticoid therapy after transplant correlate positively with final height, and longer duration of reduced GFR and higher cumulative dose of prednisone have a negative impact; other factors, such as gender, age at transplant, diagnosis, and number of transplants, do not seem to have a significant impact on adult height. The importance of height at the time of transplantation in determining final adult height, despite complex post-transplantation health issues, confirms the value of improving growth velocity and absolute height prior to transplantation. Children who receive tacrolimus (FK-506)-based immunosuppression, allowing discontinuation of glucocorticoids, have normal growth and absence of obesity.

Although the growth failure of renal disease in the pretransplantation period is not specifically due to either GH or IGF deficiency, GH therapy accelerates skeletal growth and is approved by the U.S. Food and Drug Administration (FDA) for use. Such treatment probably increases the molar ratio of IGF peptides to IGFBPs and may override the inhibitory actions of IGFBPs. GH administration may also be useful for treatment of post-transplantation growth failure.

**Hematologic Disorders**

Chronic anemia, such as sickle cell disease, is characterized by growth failure. In general, the decrease in height and weight is greater in adolescent years than earlier because the onset of the adolescent growth spurt is delayed and menarche is late. The adolescent growth and final adult height in sickle cell disease, however, may be normal. The causes of growth retardation probably include impaired oxygen delivery to tissues, increased work of the cardiovascular system, energy demands of increased hematopoiesis, and impaired nutrition. The GH-IGF system probably does not have an important role in the growth impairment of sickle cell anemia.

In thalassemia, in addition to the consequences of chronic anemia, endocrine deficiencies can result from chronic transfusions and accompanying hemosiderosis. Despite vigorous efforts to maintain hemoglobin levels near normal and to avoid iron overload, growth failure is still a common feature of thalassemia, even in transfusion-dependent adolescents. The pathophysiological defects include disproportional growth with truncal shortening but normal leg length. It is likely that anemia, impaired IGF-I synthesis, hypothyroidism, gonadal failure, and hypogonadotropic hypogonadism all contribute to growth failure in this disorder. GH is suggested by generally adequate GH production with low IGF-I levels. In most patients, GH treatment increases growth at least initially. About half of the patients in the International Fanconi Anemia Registry have short stature. GH was demonstrated by provocative testing (22 of 48) or with assessment of endogenous secretion (13 of 13) in a group with mean height of -2.23 SDS. A longitudinal study of 46 children whose diabetes began before 10 years of age indicated that initial heights at diagnosis were normal and that the final height SDS was minimally reduced from that at onset. In boys, despite a delay of about 2.5 years in onset of puberty, total pubertal height gain was normal. In girls with diabetes mellitus, however, total pubertal height gain was diminished and the adolescent growth spurt was not delayed; the effects of increased insulin and IGF-I levels on ovarian function have not been assessed in such patients. Chronic metabolic control did not correlate with the pubertal height gain or with the normal final heights. Nevertheless, good glycomic control at certain maturational periods, such as puberty, may improve growth during those intervals.

**Diabetes Mellitus**

Although weight loss may occur immediately before the onset of clinically apparent IDDM, children with new-onset diabetes are frequently taller than their peer group, possibly because GH and insulin levels are increased during the preclinical evolution of the disease. Most children with IDDM grow quite normally, even those with marginal control, especially in prepubertal years, although growth velocity may decrease during puberty. Growth failure, however, can occur in diabetic children with long-standing disease. Growth failure in those with IDDM, severe growth failure, and hepatosplenomegaly may be due to excess hepatic glycol gen deposition. This type of growth retardation has become increasingly rare.

Many pathophysiologic processes, including malnutrition, chronic intermittent acidosis, increased gluco norticoid production, hypothyroidism, impaired calcium balance, and “end-organ uteropituitaryness” to either GH or IGF, may contribute to growth failure in those with IDDM. Reflecting acquired GHI, GHBP levels are decreased, supporting the concept of impaired GH function or number. Further, IGFBP-1 is normally suppressed by insulin, and hypoinsulinemia results in elevated serum IGFBP-1 levels, which may inhibit IGF action. In contrast to the situation in adolescents and adults, IGFBP-1 levels are not elevated in well-growing prepubertal children. On the contrary, increased IGFBP-3 proteolysis may enhance the bioactivity of the available IGF-I. Most children with IDDM, however, attain normal cellular nutrition and growth factor action despite intermittent hypoinsulinemia and derangements of peripheral indices of the GH-IGF system.

Although glycomic control is inversely correlated with IGF-I levels, the correlation between glycomic control and growth is weak. With conflicting reports as to the influence of glycomic regulation on growth, a longitudinal study of 46 children whose diabetes began before 10 years of age indicated that initial heights at diagnosis were normal and that the final height SDS was minimally reduced from that at onset. In boys, despite a delay of about 2.5 years in onset of puberty, total pubertal height gain was normal. In girls with diabetes mellitus, however, total pubertal height gain was diminished and the adolescent growth spurt was not delayed; the effects of increased insulin and IGF-I levels on ovarian function have not been assessed in such patients. Chronic metabolic control did not correlate with the pubertal height gain or with the normal final heights. Nevertheless, good glycomic control at certain maturational periods, such as puberty, may improve growth during those intervals.

**Inborn Errors of Metabolism**
Inborn errors of metabolism are often accompanied by growth failure that may be pronounced. Glycogen storage disease, the mucopolysaccharidoses, the glycogeninases, and the mucolipidoses are characterized by poor growth. Many inborn metabolic disorders are also associated with significant skeletal dysplasia. In a small number of patients with organic acids (e.g., methylmalonic and propionic acidurias), IGF-I levels are low and GH levels are normal, suggesting a possible state of GHR related to nutritional status. Preliminary data suggest that exogenous GH treatment may improve the metabolic status of such children.

**Pulmonary Disease**

Growth can be retarded in children with asthma who have not received glucocorticoid therapy. Mean height and growth velocity and degree of growth failure are related to the severity of the asthma. Delayed pubertal maturation in such patients is also associated with growth deceleration in early adolescence. Impaired nutrition and increased energy requirements along with chronic stress, especially with nocturnal asthma and enhanced endogenous glucocorticoid production, cause poor linear growth. The lowered growth in asthmatic children does not appear to be associated with abnormalities of the GH-IGF axis. Glucocorticoid therapy, generally given to more severely affected patients, further impairs growth throughout childhood. Synthetic glucocorticoids, such as prednisone or dexamethasone, may have a greater growth-suppressive effect than equivalent therapeutic doses of cortisol, presumably because the biopotency of the synthetic agents may be underestimated. Alternate-day or aerosolized glucocorticoid therapy often ameliorates growth retardation and can be associated with an accelerated catch-up phase.

The use of spacer devices and effective nebulizer solutions may permit use of inhaled glucocorticoids in young children without growth impairment. Clearly, however, sufficient glucocorticoid delivered by any route can diminish growth and impede the function of the adrenal gland. Nonetheless, judicious utilization of inhaled glucocorticoid results in normal adult height despite long-term exposure and an initial decrement of linear growth velocity. Indeed, examination of near adult heights of Swedish men with asthma demonstrated an improvement in the mean difference between “severe” asthmatics and normal controls in the era of inhaled corticosteroid use. Overall, normal adult height is usually achieved.

Bronchopulmonary dysplasia (BPD), a sequel of hyaline membrane disease and prematurity, is characterized by an incidence as high as 35% in very-low-birth-weight infants (<1500 g). The use of dexamethasone in the neonatal treatment of BPD causes a transient cessation of growth and has engendered long-term concern for neurodevelopment and somatic growth. Growth in surviving infants is poor through early childhood, but the defect generally disappears by 8 years of age. Long-term hypoxia, poor nutrition, chronic pulmonary infections, and reactive airway disease are responsible for the poor early growth.

In patients with cystic fibrosis (CF), chronic pulmonary infection with bronchiectasis, pancreatic insufficiency with exocrine and endocrine inadequacy, malabsorption, and malnutrition all contribute to decreased growth and late sexual maturation. In 17,857 patients with CF, mean height was at the 21st percentile and mean weight was at the 9th percentile. Early impairment of height and weight growth and retardation of skeletal maturation may progress or plateau during middle childhood years but become most marked in the preadolescent period when growth and maturational changes are delayed.

The degree of growth retardation is related most closely to the severity and variability of the pulmonary disease rather than to pancreatic dysfunction. The degree of steatorrhea does not correlate well with growth impairment, although improved nutrition programs enhance the overall clinical picture. Adult heights in surviving patients with CF approach the normal range. Endocrine abnormalities, such as failure of both alpha and beta islet cells with decreased glucagon and insulin production do not seem to influence prepubertal growth patterns in children with CF. The incidence of diabetes mellitus increases as patients live past the second decade. Alterations of vitamin D metabolism, although potentially affecting skeletal mineralization, do not diminish growth. Delayed sexual maturation in which GnRH administration evokes a prepubertal pattern of pituitary gonadotropin secretion in adolescent patients is similar to that in CDGM. The GH-IGF axis shows evidence for acquired GH with lowered mean IGF-I and elevated GH levels.

GHT treatment of prepubertal children with CF for 1 year resulted in anabolic effect with greater growth velocity and nitrogen retention and increased protein and decreased fat stores. Pulmonary function improved in most patients. A 4-year longitudinal study using the National Cystic Fibrosis Foundation Registry found that improved nutrition status and growth were associated with a slower age-related decrement of pulmonary function.

**Chronic Inflammation and Infection**

Poor growth is a characteristic feature of chronic inflammatory disease and recurrent serious infection. We have discussed impaired growth associated with disorders such as Crohn’s disease, CF, and asthma in which inflammatory processes may be significant. De Benedetti and co-workers, studying juvenile rheumatoid arthritis in humans and a transgenic murine model expressing excessive IL-6, demonstrated an IL-6-mediated decrease in IGF-I production to be a credible mechanism by which chronic inflammatory disease could lead to poor growth. The close relationship of the GHR to that of multiple cytokines makes this an interesting hypothesis. A complex cascade of cytokines, as part of the inflammatory response to acute and chronic infection, can impact the endocrine system at many levels, impairing mineral and nutrient metabolism and the growth and remodeling of bone.

Exposure to human immunodeficiency virus (HIV) in children and adolescents occurs through perinatal transmission, blood transfusions, drug usage and sexual contact, most commonly via perinatal transmission from HIV-infected mothers. Growth failure is a cardinal feature of childhood acquired immunodeficiency syndrome (AIDS). Mean height and weight measurements during early childhood years are at or below -1 SD below the mean, but weight-for-height data may be normal in contrast with the “wasting” syndrome described in adult patients with AIDS.

In a drug treatment study of 88 HIV-infected children (mean age, 3.1 years), more than 90% were below the 50th percentile for both height and weight. Only 44% of this group survived 4 years after initiation of the study. Height or growth velocity is not a useful predictive indicator for survival. In hemophilic boys with HIV growth impairment, delayed pubertal onset and progression, and lower skeletal age were common. Despite this delayed pubertal maturation, serum testosterone levels were not significantly decreased. In many developing countries, chronic infection with parasites (e.g., schistosomiasis, hookworm, roundworm) contributes to nutritional debilitation and growth failure.
Endocrine Disorders

Hypothyroidism

Growth may be retarded in children with hypothyroidism, but the development of newborn screening programs for congenital hypothyroidism has resulted in more prompt diagnosis and treatment of such newborns (1/4100 live births). Growth in appropriately treated infants and children with congenital hypothyroidism is normal for age.\(^{[900]}\) so that skeletal maturation approximates chronologic age.\(^{[901]}\) These data do demonstrate the essential role of T\(_4\) in linear growth during the first year of life.\(^{[902]}\) Pubertal growth and maturation and final adult height are normal in well-treated congenital hypothyroidism.\(^{[903]}\)

Many features of adult myxedema are present in children with hypothyroidism. The most prominent manifestation of acquired hypothyroidism is growth failure, which may be profound.\(^{[904]}\) In acquired hypothyroidism, growth retardation may take several years to become clinically evident; once present, however, it is typically severe and progressive. The poor growth is more apparent in height than in weight gain, so those children tend to be overweight for height.

Rivkees and associates\(^{[905]}\) reported a mean 4.2-year delay between slowing of growth and the diagnosis of hypothyroidism. At diagnosis, girls were 4.04 SD below and boys 3.15 SD below mean heights for age. (This is one of several situations in which the diagnosis of short stature is later in girls than in boys.) Body proportion is immature, with an increased upper body to lower body segment ratio. Skeletal age is usually markedly delayed. Although chronic hypothyroidism is usually associated with delayed puberty, precocious puberty and premature menarche can occur in hypothyroid children (see Chapter 24).

The diagnosis of primary hypothyroidism is usually straightforward. Serum levels of T\(_4\) are reduced, and thyrotropin levels are elevated. The presence of antithyroid antibodies (usually thyroperoxidase antibodies) is consistent with a diagnosis of Hashimoto's thyroiditis, the most common cause of acquired childhood hypothyroidism in the United States. Isolated secondary or tertiary hypothyroidism, due to thyrotrpin or thyrotropin-releasing hormone (TRH) deficiency, respectively, is a rare cause of hypothyroidism.

Replacement therapy results in rapid catch-up growth. Nevertheless, accelerated growth may not restore full growth potential because rapid skeletal maturation is rapid during the first 18 months of treatment.

In one study of profoundly hypothyroid children, those treated at a mean chronologic age of 11 years had adult heights approximately -2 SD below the mean, final heights that were lower than midparental and predicted adult heights.\(^{[906]}\) The deficit in adult stature correlated with the duration of hypothyroidism before initiation of treatment.

In a separate study of hypothyroid children treated at a mean age of 9 years and with a 3-year delay of bone age, mean height SDS for bone age fell from +0.59 to -0.55 in girls and from +1.6 to -0.87 in boys. Catch-up growth may be particularly compromised when therapy is initiated near puberty.\(^{[907]}\)

On the basis of these studies, it may be appropriate to use lower than usual replacement dosages of levothyroxine and to consider a pharmacologic delay of puberty in children treated at a young age.\(^{[908]}\) Replacement therapy results in rapid catch-up growth. However, accelerated growth may not restore full growth potential because rapid skeletal maturation is rapid during the first 18 months of treatment.

Cushing's Syndrome

Glucocorticoid excess impairs skeletal growth,\(^{[909]}\) interferes with normal bone metabolism by inhibiting osteoblastic activity, and enhances bone resorption.\(^{[910]}\) Regardless of whether Cushing's syndrome is due to ACTH hypersecretion, adrenal tumor, or glucocorticoid administration. The effects of glucocorticoids are probably at the level of the epiphysis because GH secretion and serum concentrations of IGF peptides and IGFBPs are usually normal.

GH treatment cannot completely overcome the growth-inhibiting effects of excess glucocorticoids, although short-term GH or IGF-I administration can diminish many of the catabolic effects.\(^{[911]}\) Linear growth in children receiving glucocorticoids falls during GH therapy if the exogenous prednisone dose is greater than 0.35 mg/kg per day.\(^{[912]}\) The “toxic” effects of glucocorticoids on the epiphysis may persist, in part, after correction of chronic glucocorticoid excess, and patients frequently do not attain target heights.\(^{[913]}\) The longer the duration and the greater the intensity of glucocorticoid excess, the less likely is catch-up growth to be completed. Therefore, exposure to glucocorticoids should be limited as much as the underlying condition allows, frequently by the use of alternate-day therapy.

Adrenal tumors secreting large amounts of glucocorticoids can produce excess androgen, which may mask growth-inhibitory effects of glucocorticoids. In addition, Cushing's syndrome in children may not cause all the clinical signs and symptoms associated with the disorder in adults and may present with growth arrest. However, Cushing's syndrome is an unlikely diagnosis in children with obesity, because exogenous obesity is associated with normal or even accelerated skeletal growth and growth deceleration is generally evident by the time other signs of Cushing's syndrome appear (Fig. 23-43). In a series of 10 children and adolescents treated for Cushing's disease with surgery and cranial radiation, mean final height was -1.36 SDS. Post-therapy GHD was common, and GH replacement contributed to a positive change of the difference between height and target height from diagnosis to final height (-0.72 to -0.93).\(^{[914]}\)\(^{[915]}\)

![Figure 23-43](image)

Figure 23-43 Growth curves of two boys with obesity. The boy depicted by the circles had cortisol excess related to Cushing's disease. An onset of rapid weight gain was associated with a decrease in linear growth velocity at 7 years of age. The diagnosis was established, and an adrenalectomy (arrow) was performed at the age of 8½ years, with an almost immediate increase in growth rate and striking catch-up. The boy whose growth is depicted by triangles had exogenous obesity. At the age of 9½ years, his weight was approximately the same as that of the patient with Cushing's disease, but his height was at the 97th percentile, reflecting the enhancement of linear growth in this individual with exogenous obesity.

Hypoparathyroidism

Pseudohypoparathyroidism (detailed in Chapter 26) is mentioned here because growth failure is a common feature.\(^{[916]}\) This condition typically combines growth failure, characteristic dysmorphic features, and hypocalcemia and hyperphosphatemia secondary to end-organ resistance to parathyroid hormone. Affected children are short and have truncal obesity with short metacarpals, subcutaneous calcifications, round facies, and mental retardation.

Rickets

In the past, hypovitaminosis D was a major cause of short stature often associated with other causes of growth failure, such as malnutrition, prematurity, malabsorption, hepatic disease, or chronic renal failure (see Chapter 27). In isolated vitamin D deficiency, breast-fed infants typically have poor exposure to sunlight and are not nutritionally supplemented with vitamin D. Characteristic skeletal manifestations of rickets include frontal bossing, craniolabes, rachitic rosary, and bowing
of the legs. Such children usually begin to synthesize 1,25(OH)\(_2\)D as they become older, broaden their diet, and have increased exposure to sunlight with amelioration of the transient early decrease of linear growth velocity.

The association of vitamin D receptor gene polymorphism with birth length, growth rate, adult stature, and bone mineral density.

<table>
<thead>
<tr>
<th>Mutant Gene</th>
<th>Inheritance</th>
<th>Phenotype</th>
<th>References</th>
<th>Murine Homologue</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>HESX1</td>
<td>AR</td>
<td>Septo-optic dysplasia; variable involvement of pituitary hormones</td>
<td>987, 988</td>
<td>Hesx1;Rpx</td>
<td>64, 987, 990</td>
</tr>
<tr>
<td>PROP1</td>
<td>AR</td>
<td>GH, PRL, TSH, LH, FSH deficiencies; Variable degree of ACTH deficiency</td>
<td>1097, 1098</td>
<td>Prop1 (Ames mouse)</td>
<td>1097, 1098</td>
</tr>
<tr>
<td>POU1F1 (Pit-1)</td>
<td>AR, AD</td>
<td>GH, PRL deficiency; variable degree of TSH deficiency</td>
<td>971, 1151, 1119, 1124</td>
<td>Pit1/Ghfl (Snell and Jackson mouse)</td>
<td>1118</td>
</tr>
<tr>
<td>RIEG</td>
<td>AD</td>
<td>Rieger syndrome; IGHD</td>
<td>532, 1126</td>
<td>Rieg P1k2</td>
<td>1127, 1128</td>
</tr>
<tr>
<td>IGHD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GHRHR</td>
<td>AR</td>
<td>IGHD, type IB form of IGHD</td>
<td>11311134</td>
<td>Ghrhr (little mouse)</td>
<td>1114, 1116, 11371139</td>
</tr>
<tr>
<td>GH1</td>
<td>AR</td>
<td>Type IA form of IGHD</td>
<td>971, 140</td>
<td>Gh (spontaneous dwarf rat)</td>
<td>1584</td>
</tr>
<tr>
<td></td>
<td>AR</td>
<td>Type IB form of IGHD</td>
<td>971, 1096, 1120</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AD</td>
<td>Type II form of IGHD</td>
<td>971, 1096, 1120</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>X-linked</td>
<td>Type III form of IGHD; hypogamma-globulinemia</td>
<td>1148</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AD</td>
<td>Bioactive GH molecule</td>
<td>1151, 1152</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GHI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GHR</td>
<td></td>
<td></td>
<td></td>
<td>Ghr</td>
<td>267</td>
</tr>
<tr>
<td>Extracellular</td>
<td>AD, AD</td>
<td>IGF-I deficiency, decreased or normal GHBP</td>
<td>169, 1178, 1179</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transmembrane</td>
<td>AR</td>
<td>IGF-I deficiency, normal increased GHBP</td>
<td>203, 1185</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intracellular</td>
<td>AD</td>
<td>IGF-I deficiency; normal increased GHBP</td>
<td>1186, 1187</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Primary Defects of IGF Synthesis**

| IGF1        | AR          | IGF-I deficiency; IUGR and postnatal growth failure | 270 | Igif | 277 |

ACTH, adrenocorticotropic hormone; AD, autosomal dominant; AR, autosomal recessive; FSH, follicle-stimulating hormone; GH, growth hormone; GHBP, GH-binding protein; GHD, GH deficiency; GHR, GH receptor; GHRH, GH-releasing hormone; GHRHR, GHRH receptor; IGF, insulin-like growth factor; IGHD, isolated GHD; LH, luteinizing hormone; PRL, prolactin; TSH, thyrotropin (thyroid-stimulating hormone).

Vitamin D-resistant (hypophosphatemic) rickets is an X-linked dominant disorder that is due to decreased renal tubular reabsorption of phosphate related to mutations in the phosphate-regulating endopeptidase gene (PHEX, Xq22.1) (see Chapter 27). The presence of high-affinity binding sites for the vitamin D receptor DNA-binding domain in the GH promoter suggests that the vitamin D receptor may actually modulate GH expression. Vitamin D-resistant rickets.

Treatment requires oral phosphate replacement, but such therapy may result in poor calcium absorption from the intestine. The addition of calcitriol to oral phosphate increases intestinal phosphate absorption and prevents hypocalcemia and secondary hyperparathyroidism. Such combined therapy does improve the rickets but does not necessarily correct growth. There is no clear association between endogenous GH secretion, IGF-I, or phosphate levels and height in this disorder. Nevertheless, GH therapy, in eight trials including 83 patients, has resulted in an enhancement of skeletal growth and improvement in bone mineral density.

IGF-I Deficiency

Because IGF-I is a major mediator of skeletal growth, its deficiency can result in severe growth failure. Causes of IGF-I deficiency include:

1. Central hypothalamic-pituitary dysfunction with failure of pituitary GH production (i.e., hypopituitarism or GHD). It may be impossible to discriminate between hypothalamic and pituitary dysfunction if both organs have the same pathologic process.
2. Primary or secondary GH insensitivity (GHI).

We use the term insulin-like growth factor I (IGF-I) deficiency syndrome to describe the generic condition, whether caused by GHD, dysfunction, or insensitivity, to illustrate this unifying concept.

It is not always possible to discriminate completely between hypothalamic and pituitary dysfunction because both organs may be involved in the same pathologic process. In addition, as described earlier, embryonic development of the hypothalamus and pituitary appears to be codependent. A number of factors produced in the developing ventral diencephalon function as molecular signals for initial formation and development of Rathke's pouch; subsequent differentiation of each of the various
anterior pituitary cell types appears to be primarily regulated by a strict temporal and spatial pattern of pituitary transcription factors. It is somewhat of a semantic issue as to whether to label some of the molecular defects either “hypothalamic” or “pituitary”; nevertheless, some arbitrary classification decisions have been made in the following discussions. Table 23-5 presents our current classification of molecular defects of the GH-IGF axis, whereas the sites of defined and likely sites of genetic defects are shown in Figure 23-44. The murine homologues of these gene defects are also listed in Table 23-5.

A potential system for analysis of the known and potential genetic errors in patients with IGF deficiency syndrome is shown in Figure 23-45. It is to be anticipated that significant development will be made in our understanding of these defects over the next few years and that this classification will require frequent updating and modification.

Clinical Features

IGF-I deficiency due to hypothalamic dysfunction with abnormalities of GHRHR, endogenous GHS or SRIF synthesis or secretion, primary or secondary decreased pituitary GH production, or GHI share a common phenotype. The similarity among these patients emphasizes the role of IGF-I in mediating most of the anabolic and growth-promoting actions of GH. This point is further supported by the capability of IGF-I therapy to correct growth in children with mutations of the GHR gene.

Accordingly, the typical clinical features of severe IGF deficiency are shared by all of these conditions. If GH or IGF deficiency is acquired, clinical signs and symptoms appear at a later age.

Birth size is normal or near normal in most children with IGF-I deficiency but low in severe congenital GHD and GHI and in the single case of a deletion of the IGF-I gene. Typically, birth length and weight are within 10% of normal, and severe IUGR is not part of typical IGF deficiency but is present in infants with profound IGF-I deficiency, confirming the critical role of IGF-I in intrauterine growth. Infants with early-onset GHD may have birth lengths of around 2 SD below the mean.

Although at least 50% of infants diagnosed before 2 years of age have birth lengths more than 2 SD below the mean (both in isolated GH and in multiple pituitary hormone deficiencies), mean birth weight is about -1 SD, lending an appearance of relative adiposity, even in the neonatal period. These data further support an intrauterine role of the GH-IGF system in growth regulation along with the high frequency of abnormalities of the hypothalamic-pituitary area defined by magnetic resonance imaging (MRI). Anatomic abnormalities include dysgenesis of the pituitary stalk, ectopic placement of the posterior pituitary inferior to the median eminence, and diminished volume of the anterior pituitary. There is high frequency of breech deliveries and perinatal asphyxia. Neonatal morbidity can include hypoglycemia and prolonged jaundice with direct hyperbilirubinemia due to cholestasis and giant cell hepatitis. When GHD is combined with deficiency of ACTH and thyrotropin, hypoglycemia may be severe. The combination of GHD with gonadotropin deficiency can cause microphallus, cryptorchidism, and hypoplasia of the scrotum. GHD (or GHI) should, therefore, be considered in the differential diagnosis of neonatal hypoglycemia and of microphallus/cryptorchidism.

Postnatal growth is abnormal in severe congenital IGF deficiency. Most surveys of GHD and GHI indicate that growth failure can occur during the first months of life.

Figure 23-44 The hypothalamic-pituitary-IGF axis: sites of established and hypothetical defects. The established defects are shown as Roman numerals in the gray-shaded circles or ovals, and the hypothetical defects are shown in the white circles or ovals. ALS, acid-labilesubunits; GH, growth hormone; GHB, growth hormone-binding protein; GHRH, growth hormone-releasing hormone; IGF, insulin-like growth factor; IGFBP, IGF-binding protein; JAK, Janus kinase; MAPK, mitogen-activated protein kinase; STAT, signal transducer and activator of transcription. (From Lopez-Bermejo A, Buckway CK, Rosenfeld RG. Genetic defects of the growth hormone-insulin-like growth factor axis. Trends Endocrinol 2000; 11:43.)

Figure 23-45 Decision tree for the investigation of genetic defects in patients with IGF deficiency. Hypothetical genetic defects are presented in parentheses. In these defects shown with a * sign, abnormalities in other organs and structures besides the hypothalamic-pituitary-IGF axis are expected to occur as a result of these genetic defects: ACTH, adrenocorticotropic hormone; CPHD, combined pituitary hormone deficiencies; FSH, follicle-stimulating hormone; GH, growth hormone; GHI, GH receptor; GHRHR, GH-releasing hormone receptor; IGF, insulin-like growth factor; IGFIR, IGF-I receptor; LH, luteinizing hormone; PRL, prolactin; STAT5, signal transducer and activator of transcription 5; TSH, thyrotropin (thyroid-stimulating hormone). [From Lopez-Bermejo A, Buckway CK, Rosenfeld RG. Genetic defects of the growth hormone-insulin-like growth factor axis. Trends Endocrinol 2000; 11:43.]

Figure 23-46 Magnetic resonance image of infundibular dysgenesis. A, T1-weighted sagittal and coronal images of the hypothalamic-pituitary area in a normal 8-year-old girl. The anterior pituitary (AP) and posterior pituitary (PP) lobes and the pituitary stalk (PS) are marked. B, T1-weighted sagittal and coronal images of the hypothalamic-pituitary area of a 17-year-old boy with isolated human growth hormone deficiency. The anterior pituitary (AP) lobe is hypoplastic, the posterior pituitary (PP) lobe is ectopic, and the pituitary stalk is absent. (From Rozd AJ, Martinez CR. Magnetic resonance imaging in patients with hypopituitarism. Trends Endocrinol Metab 1990; 3:293287.)

Figure 23-47 Height measurements for Ecuadorian children with insulin-like growth factor deficiency resulting from growth hormone insensitivity. (From Rosenfield RG, Rosenblom AL, Guevara-Aguilar J. Growth hormone [GH] resistance due to primary GH receptor deficiency. Endocr Rev 1994; 15:369390. © The Endocrine Society.)

Skeletal proportions tend to be relatively normal but correlate better with bone age than with chronologic age. Skeletal age may be delayed to less than 60% of the chronologic age but in the absence of hypothyroidism is similar to the height age. In acquired GHD, as from a CNS tumor that causes increased intracranial
pressure, bone age may approximate the chronologic age; delayed skeletal maturation, therefore, should not be required for the diagnosis of GHD.

Assessment of volumetric bone mineral density reveals decreased mineralization beyond that dependent on small size. Weight-to-height ratios tend to be increased, and fat distribution is often "infantile" or "doll-like" in pattern. Musculature is poor, especially in infancy and can cause delay in gross motor development and to lead the erroneous impression of mental retardation in an immature-appearing child. Facial bone growth may be particularly retarded, with an underdeveloped nasal bridge and frontal bossing. Fontanel closure is often delayed, but the overall growth of the skull is normal, leading to cephalofacial disproportion and the appearance of hydrocephalus. The voice is infantile because of hypoplasia of the larynx. Hair growth is sparse and thin, especially during early life; nail growth is also slow. Even with normal gonadotropin production, the penis is small and puberty is usually delayed.

Final height data in patients with untreated GHD are not plentiful. Wit and colleagues summarized data from 22 untreated men and 14 women with severe isolated GHD with a mean final height SDS of -4.7. In 19 patients with multiple pituitary hormone deficiencies, thus lacking in gonadal steroids, mean final height was -3.1 SDS.

Etiology

Two surveys of nearly 63,000 GH-treated patients, in databases managed by Pharmacia (Pharmacia International Growth Database [KIGS]) and Genentech (National Collaborative Growth Study [NCGS]), cared for by pediatric endocrinologists throughout the world include more than half of the internationally treated patients. The patients are diverse and include subjects with GHD and with Turner's syndrome and miscellaneous other disorders. About 59% of the total group (37,000 patients) had GHD (as defined by a stimulated GH level of < 10 µg/L) of whom 78% had "idiopathic" GHD and 22% had "acquired" or "organic" (neoplasms, trauma, inflammation, miscellaneous) causes of GHD. The latter group includes patients with congenital (developmental) GHD-associated syndromes. The organic/acquired group is probably underestimated because many of the patients classified as idiopathic had not had definitive imaging assessments of the hypothalamic-pituitary region and possibly have congenital structural abnormalities.

With the availability of synthetic IGF-I for treatment of patients with inherited abnormalities of the GHR, approximately 200 patients with primary GHD have been identified. This is an exceedingly small number of subjects, even with the addition of the potentially larger group of individuals with heterozygous abnormalities of the GHR. In contrast, patients with secondary GHI, including those with malnutrition or chronic systemic disease, must be considered as potentially a huge number, on a worldwide basis.

An incidence of GHD of 1:6,000 live births has been reported from the United Kingdom, and a survey of Scottish schoolchildren indicated prevalence as high as 1:4,000. The best estimate in the ultimate human population is approximately 1:3,480. It is likely, however, that childhood GHD is an overdiagnosed condition. In particular, the diagnosis of acquired, idiopathic, isolated GHD should always be suspect. Although one might argue that (1) destructive or inflammatory lesions of the hypothalamus or pituitary may affect only GH secretion or (2) isolated GHD due to a mild mutation or deletion of the GHRH receptor gene or GH gene may appear late or (3) combined pituitary hormone deficiencies (CPHDs) may first present with what appears to be isolated GHD, such circumstances appear to be rare. In the absence of anatomic abnormalities evident on imaging studies, or biochemical evidence of CPHD, the diagnosis of acquired, isolated, idiopathic GHD demands careful and thorough documentation, with greater skepticism as children approach adolescence. The entity of partial, transient GH insufficiency related to sex steroid deficiency in delayed puberty is particularly confounding.

Central Hypothalamic-Pituitary Abnormalities

Many of the disorders that affect hypothalamic regulation of GH synthesis and secretion also have a direct impact on pituitary function. Consequently, it is not always possible to definitively establish the primary of hypothalamic or pituitary dysfunction, hence the term idiopathic. Nevertheless, congenital (developmental) or functional abnormalities of the hypothalamus account for most idiopathic cases of hypopituitarism, and many such cases of GHD prove to have a molecular basis. Acquired structural damage to this area (such as from neoplasm and trauma) causes about 25% of GHD cases.

Hypothalamic Dysfunction: Genetic Abnormalities.

Hypothalamic factors involved in regulating GH synthesis and secretion include, but may not be limited to, GHRH, GHSs, endogenous GH-releasing peptides such as ghrelin, PACAP, galanin, and somatostatin. Mutations of the genes encoding these or other hypothalamic peptides may explain some cases of IG deficit due to hypothalamic dysfunction. To date, however, mutations of the genes encoding GHRH have not been identified. Targeted disruption of the murine homeobox gene Tebp, expressed in the ventral diencephalon but not in Rathke's pouch or in the pituitary during embryogenesis, results in early ablation of the pituitary primordium.

A Gsh-1 homeobox gene, expressed in varied parts of the developing murine CNS, plays an important role in pituitary development because mutant strains have impaired production of GHRH with anterior pituitary hypoplasia and GHD. The broad impact of defects of this gene on hypothalamic releasing factors may be similar to mutations of the Pit-1 gene because deficiencies of prolactin and LH also occur. Deletions of the murine dopamine transporter (DAT) result in increased dopaminergic tone, anterior pituitary hypoplasia, and dwarfism. This suggests an important role for hypothalamic dopamine in pituitary development.

Congenital Malformations Involving the Hypothalamus.

Hypothalamic dysfunction from congenital malformations of the brain or hypothalamus is a common cause of hypopituitarism. As noted earlier, patients with early diagnosed congenital GHD frequently have an abnormal pituitary stalk, ectopia of the posterior pituitary, and hypoplasia of the anterior pituitary. Anencephaly results in a pituitary gland that is small or abnormally formed and is frequently ectopic. Despite the loss of hypothalamic regulation, somatotrophs differentiate and proliferate diminished overall mass.
parallel to each other. GH GHD can occur by itself or in combination with deficiencies of thyrotropin, ACTH, and gonadotropins. About 50% of children with severe anatomic defects have hypopituitarism, and the diagnosis should be considered in any child with growth failure associated with pendular or rotary nystagmus or impaired vision and a small optic nerve disc. In some patients, hypoplastic or interrupted pituitary stalks and ectopic posterior pituitary placement have been identified by MRI. It is not clear whether this disorder is inherited, but there is an increased incidence in offspring of young mothers and in first-born children.

Mutations of HESX1, a paired-like homeodomain gene, expressed early in pituitary and forebrain development, are associated with familial forms of septo-optic dysplasia. Three of 228 patients with a broad spectrum of congenital pituitary defects ranging from pituitary hypoplasia and septo-optic dysplasia were found to have heterozygous mutations of the HESX1 gene. Transgenic mice lacking this gene exhibit a variety of anterior midline CNS defects (e.g., abnormalities of the corpus callosum and septum pellucidum and microphthalmia) and pituitary dysplasia.

In most patients, so-called idiopathic hypopituitarism or GHD is due to abnormalities of synthesis or secretion of the hypothalamic hypophysiotropic factors. In a number of reports, idiopathic GHD is associated with MRI findings of an ectopic neurohypophysis, pituitary stalk dysgenesis, and hypoplasia or aplasia of the anterior pituitary.

In multiple series involving 397 children with isolated GHD or with CPHDs, 54% had the characteristic MRI findings; 93% of CPHD patients were abnormal in contrast with 32% of patients with isolated GHD.

Abrahams and co-workers studied 35 patients with idiopathic GHD and found that those with MRI abnormalities could be divided into two groups: (1) 43% had an ectopic neurohypophysis (neurohypophysis located near the median eminence), absent infundibulum, and absence of the normal posterior pituitary bright spot; and (2) 43% had a small anterior pituitary, either as an isolated finding or combined with an ectopic neurohypophysis. Overall, those patients with the most striking abnormalities of the hypothalamic-pituitary region, those with MRI abnormalities of the hypothalamic-pituitary region, had the smallest anterior pituitary glands.

Patients with more severe deficiencies of GH have greater frequency of significant morphologic abnormalities.

Although the increased incidence of breech presentation and birth trauma with neonatal asphyxia in congenital idiopathic hypopituitarism has led some to suggest an etiologic role for these occurrences, the syndrome of pituitary stalk dysgenesis with congenital hypopituitarism is probably due to abnormal development, and the perinatal difficulties are likely the consequence rather than the cause of the abnormalities. Findings of a similar MRI appearance in patients with septo-optic dysplasia, in association with type I Arnold-Chiari syndrome and syringomyelia, and perhaps also in holoprosencephaly and the occurrence of microcephaly with this syndrome, all support the concept that congenital hypopituitarism is a genetic or developmental malformation, not a birth injury.

A single report of an apparent autosomal dominant mutation of early brain development with infundibular dysgenesis associated with GH deficiency supports the primary nature of the abnormality. Further indirect evidence in studies of isolated, complete anterior pituitary aplasia indicates that hypothalamic hypopituitarism and breech delivery are consequences of congenital midline brain defects, although perinatal residua of breech delivery may exacerbate ischemic damage to the hypothalamic-pituitary unit.

The MRI findings described earlier for patients with an early diagnosis of hypopituitarism are also found in children diagnosed at a later age. Most of these children have hypothalamic dysfunction as the cause of diminished pituitary hormone secretion. In the older group, as in the infants, structural, acquired hypothalamic, stalk, or pituitary abnormalities must be considered.

Patients with myeloencephalocele with long-term survival have growth failure, decreased bone mineral density, and pubertal abnormalities. Diminished growth is due to maldevelopment of the vertebral-skeletal system and to diverse midline CNS developmental anomalies such as hydrocephalus and Arnold-Chiari malformation. Many have hypothalamic-pituitary dysfunction including GHD and precocious puberty. GH treatment does improve growth in these patients, although most of the growth is in the trunk and arms in children with shunted hydrocephalus, nearly two thirds have endocrine abnormalities. As a group, prepubertal patients were about 1 SDS below control populations and had a higher body mass index. Sixteen (30%) of 54 patients had inadequate GH production.

pituitary heights found on MRI. Early and accelerated pubertal maturation contributes to reduced final height.

Trauma of the Brain and Hypothalamus.

Head trauma may cause isolated GHD or multiple anterior pituitary deficiencies, and some series of patients with GHD indicate an increased incidence of birth trauma, such as breech deliveries, extensive use of forceps, prolonged labor, or abrupt delivery. Although GHD may be a consequence of a difficult delivery or hypoxic perinatal period, it is more commonly an associated developmental abnormality deficiency (see earlier) or due to head trauma later in life. In a series of 22 head-injured adolescents and adults, nearly 40% had some degree of hypopituitarism.

Inflammation of the Brain and Hypothalamus.

Bacterial, viral, or fungal infections may result in hypothalamic or pituitary insufficiency, and the hypothalamus, pituitary gland, or both may also be involved in sarcoidosis.

Tumors of the Brain and Hypothalamus.

Brain tumors are a major cause of hypopituitarism insufficiency, especially midline brain tumors such as germinomas, meningiomas, gliomas, ependymomas, and optic nerve gliomas. Although short stature and GHD are most often associated with suprasellar lesions in neuroblastoma, they may also exist without such lesions; whether growth impairment antedates the pathologic findings is not clear. Metastases from extracranial carcinomas are rare in children, but hypothalamic insufficiency can result from local extension of craniopharyngeal carcinoma or Hodgkin’s disease of the nasopharynx.

The laboratory diagnosis of GHD in children with brain tumors may be difficult because levels of both IGF-I and IGFBP-3 are poor predictors, especially in pubertal patients. Cranial radiotherapy and histiocytosis can cause hypothalamic dysfunction, as described elsewhere ("Pituitary Growth Hormone Deficiency").

Radiation of the Brain and Hypothalamus.

Cranial radiation appears to be an increasing cause of hypothalamic-pituitary dysfunction. Taken in aggregate, there may be as many as 4000 pediatric cancer survivors whose GH has resulted from the broad range of cancer treatments.

Radiation may impair both hypothalamic and pituitary function, and often it is not easy to discriminate between damage at the two levels. The hypothalamus is more radiosensitive than the pituitary and is more often the site of damage, especially in the dose range usually given to children with malignancy. Thyroidal and gonadal function may also be directly impaired by certain radiation therapies. The degree of pituitary dysfunction is relative to the dose of radiation received. Low doses typically cause isolated GHD, and higher doses cause multiple pituitary deficiencies. GH develops in most long-term survivors, with the adverse effect of radiotherapy directly related to the biologically effective dose to the hypothalamus. Within 5 years of radiation, nearly 100% of children receiving more than 30 Gy over 3 weeks to the hypothalamic-pituitary axis had subnormal GH responses to provocative tests, whereas GHD may not become apparent for a decade or more after lower doses (18 to 24 Gy).

The degree of pituitary deficiency is also a function of the length of time after radiation; children who test normally at 1 year post-therapy may develop pituitary deficiencies later. Prior to development of GH secretory deficiency, GHI with low levels of IGF-I, IGFBP-3, and GHB (presumably caused by the malignancy and the intensive chemotherapy and radiation therapy regimens) may decrease growth velocity.
Chemotherapy regimens by themselves may impair final adult height, although not nearly to the extent seen after radiation. When such therapy is stopped, prepubertal children experience some degree of catch-up growth but also have persistent abnormalities of the IGF/IGFBP system. Even when serum GH responses to provocative testing are normal, spontaneous GH secretion may be blunt at x-ray doses as low as 18 to 24 Gy. Although acquired GHD impairs final height, the relation of diminished GH production to levels of the GH-dependent peptides, IGF-I, and IGFBP-3 is variable. With long-term follow-up, however, correlations have been found between nocturnal GH secretion, levels of IGF-I and IGFBP-3, and pituitary size.

Poor linear growth from decreased GH secretion may be exacerbated by the impact of radiation itself, with inadequate pubertal acceleration of spinal growth. Surprisingly, cranial radiation can result in precocious puberty, especially in children undergoing radiation therapy at young ages, causing early epiphyseal fusion. Sexual precocity appears to occur more frequently with low doses of radiation and gonadotropin deficiency is likely at high doses. The rate of pubertal progression, however, does not appear to be accelerated.

Treatment with GnRH analogues may be necessary to suppress the hypothalamic-pituitary-gonadal axis in an attempt to attain normal final height. Three possibilities must be considered in following children after craniospinal radiation: (1) evolving hypopituitarism, (2) decreased spinal growth potential, and (3) early puberty with premature epiphyseal fusion. Children with documented GHD and growth failure are candidates for exogenous GH treatment. There is no evidence for enhanced relapses of the primary neoplasm in patients receiving GH, but there appears to be a variable growth response to GH. Spinal growth impairment, inadequate or delayed treatment, and sexual precocity may limit linear growth.

Bone marrow transplantation (BMT) for patients with inborn errors of metabolism, aplastic anemias, and malignancies requires preparative regimens that include total body irradiation, often with chemotherapy and sometimes including cranial radiation. Children whose clinical condition required modest treatment programs before BMT experienced minimal loss of growth after BMT. In children who underwent cranial radiation followed by high-dose chemotherapy and total body radiation, especially in a single dose, as preparative regimens, growth failure was almost inevitable 2 to 5 years after BMT. Fetal growth was most affected. If the total body radiation was fractionated and if cranial radiation was not previously needed, growth velocity and height were not compromised 3 years after BMT.

In the absence of cranial radiation, there was poor correlation between GH production and levels of IGF-1 or IGFBP-3 and growth in children after BMT suggesting the importance of factors such as nutrition and radiation-induced vertebral dysplasia or hypothyroidism. Growth failure may be evident at birth and is more impressive postnatally, with mean adult heights more than 2 standard deviations below the mean. Mean final heights in 28 long-term survivors of BMT were about 1 SDS lower than at the time of the transplantation but still within the normal range in all but one patient. Such data suggest a conservative approach with regard to exogenous hormonal treatment.

Psychosocial Dwarfism.

An extreme form of failure to thrive is termed psychosocial dwarfism or emotional deprivation dwarfism. Most cases of failure to thrive can be traced back to a poor home environment and inadequate parenting, with improved weight gain and growth on removal of the infant from the dysfunctional home. Some children have dramatic behavioral manifestations beyond those in the typical infant with failure to thrive, namely bizarre eating and drinking habits, such as drinking from toilets, social withdrawal, and primitive speech.

Hyperphagia and abnormalities of GH production are associated. GH secretion is low in response to pharmacologic stimuli but returns to normal on removal from the home. Concomitantly, eating and behavioral habits returned to normal and a period of catch-up growth ensued. Careful assessment of endogenous GH secretion showed reversal of the GH insufficiency within 3 weeks, including enhancement of GH pulse amplitude and a variable increase of pulse frequency. The reversibility of GH secretion and the later growth increment in the context of the clinical findings described earlier confirm the diagnosis of psychosocial dwarfism.

The neuroendocrinologic mechanisms involved in psychosocial dwarfism remain to be elucidated. GH secretion is abnormal, and ACTH and thyrotropin levels may also be low, although some patients have high plasma cortisol levels. Even when GH secretion is reduced, treatment with GH is not usually beneficial until the psychosocial situation is improved. Management of the emotional causes of the growth failure is imperative and often associated with substantial growth. In our experience, although psychosocial dysfunction is a common cause of failure to thrive in infancy, the constellation of bizarre behaviors described in psychosocial dwarfism is rare.

The fact that GH production is impaired in adults with varied psychiatric disorders and the presence of growth aberrations of functional GHD with psychosocial dwarfism suggest that children with emotional problems may have impaired GH secretion and growth. Indeed, depression in children, as in adults, can lower GH production and anxiety disorders in girls predict a modest height loss in adults.

Growth Hormone Neurosecretory Dysfunction.

Because tests of GH secretion following pharmacologic provocation may not accurately reflect normal GH secretion, it has been argued that a subset of children with GH neurosecretory dysfunction may be identified by frequent or continuous serum sampling over a 12- to 24-hour period. This condition is characterized by short stature and poor growth, normal serum GH response to provocative testing, but reduced IGF-I and 24-hour serum GH levels. Prior cranial radiation may be the most common cause of these findings. Patients with ISS (see later) do not appear to have diminished 24-hour GH production rates, especially when the very broad range of data in normal and short normal children are considered. There appears to be little doubt that some of children with GH neurosecretory dysfunction secrete insufficient amounts of GH, even if they pass provocative GH testing; whether they should be identified by 24-hour GH sampling or by determination of the GH-dependent peptides is unclear.

Prader-Willi Syndrome.

Prader-Willi syndrome is a genetically determined syndrome complex, with a frequency of 1 in 10 to 25,000 live births. Profound neonatal hypotonia and subsequent developmental delay and strength are seen. Growth failure may be evident at birth and is more impressive postnatally, with mean adult heights more than 2 SD below the mean and almost always below the midparental height target range. Cryptorchidism and microphallus are present in the neonate and hypogonadotropic hypogonadism may persist into adult life. With advancing age, hyperphagia and obesity become prominent.

The genetic defect in Prader-Willi syndrome is a functional deletion of the paternal allele within chromosome 15q11-13. Most patients with the syndrome have deletions of the long arm of the paternal derived chromosome 15, in some both copies of 15q may be maternally derived (uniparental disomy), whereas rarely there may be mutations of the imprinting center of chromosome 15q.

The probable cause of the short stature in Prader-Willi syndrome is deficient GH production due to as yet undefined hypothalamic dysfunction. MRI assessment of the hypothalamic-pituitary area does not yield evidence for congenital structural abnormalities. The body habitus and composition are similar to those in classic GHD, including small hands and feet, increased fat mass, and low muscle mass. Low mean serum GH levels or inadequate responses after provocative testing may reflect the impact of obesity, but serum levels of GH-dependent peptides are low in Prader-Willi syndrome, in contrast with the findings in obesity, a condition in which these factors are produced normally despite diminished GH production. Thus, Prader-Willi syndrome is an IGF deficiency condition due to inadequate GH production, although the possibility of failure of up-regulation of GH action, as seen in obesity, may yet be found to play a role.

GH treatment of growth failure in patients with Prader-Willi syndrome is now an FDA-approved indication for GH use. Treatment results in improved growth velocity, normalization of final height potential, increased muscle mass and strength, and decreased fat mass. In view of the risk for development of obesity-related insulin resistance and diabetes mellitus, glycemic status must be monitored closely during GH therapy.
In most pediatric endocrine centers, many children receiving GH are classified as having acquired, idopathic, isolated GHD. As noted earlier, this diagnosis should always be considered somewhat suspect, especially prepubertally, although some patients may actually have undiagnosed gene defects in GH production or secretion or may be exhibiting the first manifestation of CPHD.

Multiple studies have reported data on retesting patients with GH-treated GHD during or after cessation of therapy. All the vagaries of different GH assays and GH provocative tests, varied “cut-off” levels of GH normality, diagnostic categorization of patients, and radiologic interpretation of MRI findings certainly affect these evaluations. Nonetheless, several clear conclusions emerge. In 464 patients with isolated GHD, 207 (44%) had normal GH levels during provocative retesting. In contrast, approximately 96% of 148 patients with CPHD, with or without structural abnormalities of the hypothalamic-pituitary area, had sustained GHD. The presence of multiple anterior pituitary hormone deficiencies or structural disease would seem to obviate the need for subsequent retesting. Whether these results simply cast doubt on the validity of the initial GH tests (or GH provocative testing in general) or whether children with earlier GHD may truly normalize is not clear. The entity of partial, transient GHD associated with delayed puberty certainly may be an example of this latter situation.

**Pituitary Growth Hormone Deficiency.**

As discussed earlier, many of the disease processes that impair hypothalamic regulation of GH secretion also impair pituitary function. Another group of abnormalities specifically affects pituitary somatotroph development and function.

As many as 3% to 30% of patients with GHD have an affected parent, sibling, or child, and multiple genetic causes of GHD have been recently described. We discuss inborn errors of genes for nuclear transcription factors affecting hypothalamic-pituitary development, the GHRH receptor, and the GH gene, each of which can cause GHD and IGF deficiency.

**Genetic Abnormalities Resulting in Combined Pituitary Hormone Deficiency:**

Septo-optic dysplasia and its relationship to HESX1 have been discussed earlier. The gene PRO1 (denoting prophet of Pit1) encodes a paired-like 226amino acid homeodomain protein, which is involved in the early determination and differentiation of multiple anterior pituitary cell forms and is necessary for POU1F1 (Pit1) expression. Mutation of this gene is responsible for a form of murine pituitary-dependent dwarfism, the Ames mouse. Abnormalities of human PRO1 result in CPHD, characterized by variable and often age-dependent degrees of deficiency of GH, prolactin, thyrotropin, FSH, LH, and, occasionally, ACTH.

Gonadotropin abnormalities are particularly diverse because about 30% of patients undergo spontaneous pubertal development, including menarche, before ultimately developing hypogonadotropic hypogonadism. Striking variability has been described in pituitary size, with very large glands having a hyperintense T1-weighted signal occasionally demonstrated by MRI. These pituitary glands may then undergo involution, leaving a large empty sella in a patient with complete anterior hypopituitarism, including ACTH deficiency. At least eight PRO1 (chromosome 5q35, OMIM01536) gene abnormalities have been identified and include missense, frameshift, and splicing mutations. A GH repeat in exon 2 (295-CGA-GAG-AGT-303) has been reported to be a “hot spot” in the PRO1 gene; any combination of a GA or AG deletion in this repeat region results in a frameshift in the coding sequence and premature termination at codon 109.

Similar abnormalities result from homozygous lesions at other sites on exon 2 affecting codons 73, 88, and 149. Further, compound heterozygosity for two mutations was detected in 36% of children from four families because two different common deletions both led to a stop codon at position 109. These mutations all result in loss of the DNA-binding and C-terminal trans-activating domains of PRO1. There does not appear to be strong correlation between phenotype and genotype. Large-scale screening of patients with CPHD has found 54% with PRO1 mutations in two series of patients with multiple affected individuals, however, PRO1 mutations accounted for all of the 25 siblings. Such frequencies far exceed those reported for POU1F1 mutations and emphasize a central role for PRO1 in pituitary cellular differentiation.

The POU1F1 (Pit-1, GHF-1) gene (chromosome 3p11, OMIM 173110) encodes Pit1, a member of a large family of transcription factors, referred to as POU-domain proteins, and is responsible for pituitary-specific transcription of genes for GH, prolactin, thyrotropin, and the GHRH receptor. Additionally, the 280amino acid Pit-1 protein activates transcription of genes that regulate differentiation, proliferation, and survival of somatotrophs, lactotrophs, and thyrotrophs. Both Snell (dw/dw) and Jackson (dw/dwJ) dwarf mice have GH, prolactin, and thyrotropin deficiency and mutations or rearrangements of the murine Pit-1 gene. Many different mutations (at least 12 point mutations and deletions in the POU1F1 gene) have been found internationally in families with GHD and prolactin deficiency and variable defects in thyrotropin secretion. The most common mutation is an R271W substitution affecting the POU homeodomain protein, which is involved in the early determination and differentiation of multiple anterior pituitary cell forms and is necessary for POU1F1 (Pit1) expression.

**Molecular Defects of GHRH.**

Despite extensive assessment, mutations of the gene encoding GHRH, which would cause the IGF deficiency phenotype, have not been identified. Although expression of GHRH is suppressed in mutations of murine neural genes GnF1 and DAT1, the failure to demonstrate a GHRH mutation remains a surprise because this gene would appear to be a likely candidate for familial GHD. Abnormalities of the endogenous "GHSs" or their receptors remain to be identified. Mutations causing constitutive or enhanced ligand-mediated activation of the G protein-related somatostatin receptor to yield chronic inhibition of GH also have not yet been reported.

**Molecular Defects of the GHRH Receptor.**

Although no defects of the GHRH gene have yet been reported, multiple kindreds have been found with homozygous mutations of the GHRH receptor gene. Wajnrajch and associates reported the first human cases of a mutation in the GHRH receptor gene (chromosome 7) in two cousins with IGF deficiency and growth failure. These patients had a severely truncated GHRH receptor protein that lacked the membrane-spanning regions and the G protein-binding site. The affected children exhibited undetectable GH release during standard provocative tests and after exogenous GHRH administration but responded to GH treatment.

Another series of 18 patients with the same point mutation was found in Pakistan ("dwarfish of Sind") and in two members of a Tamoulean family from Sri Lanka. The largest kindred with a mutation of GHRH receptor has been identified in Brazil; a donor splice mutation in position 1 of exon 1 also results in a severely truncated GHRH protein.

The patients in all of the groups have striking short stature (often more than -5 SDS), lack other features of GHD such as microphallus, truncal obesity, and...
hypoglycemia, but have profound abnormalities of the GH-IGF axis. The absence of the GHRH receptor in the tests does not preclude fertility. The patients respond well to exogenous GH without antibody formation. Heterozygotes may have minimal height deficits but do show moderate biochemical deficiencies of the GH-IGF axis.

Despite extensive study, the geographic separation and ethnic differences do not suggest recent (>200 years) contact among the families from the Indian subcontinent. The most probable explanation for all four families is that of a "founder effect," or one-time mutations in each group with propagation within geographically isolated gene pools. In an analysis of 30 families with isolated GHD type IB, Salvatori and colleagues found new missense mutations in transmembrane and intracellular domains of GHRH receptor in three families (10%) with two affected members in each. Transfection experiments indicated normal cellular expression of these mutant receptors.

Mutation of the gene for GHRH receptor in its ligand binding domain has also been identified in the little mouse (little). leading to dwarfism and decreased numbers of somatotrophs. This in model, the fetal somatotroph mass is normal and hypoplasia, but not absence, of the somatotrophs is evident only after birth. Such data suggest that GHRH is not an essential factor for fetal differentiation of the somatotrophs and that GHRH-independent cells persist or that mutation does not cause total loss of GHRH function.

**Genetic Abnormalities of Growth Hormone Production and Secretion Resulting in Isolated Growth Hormone Deficiency.**

Four isomeric GHD due to errors of the GH gene have been reported (see Table 23-5). The gene encoding GH (GH1) is located on chromosome 17q23 in a cluster that includes two genes for hpl: (1) a pseudogene for hpl and (2) the hpl gene that encodes placental hpl. GH1 and GH2 differ in mRNA splicing pattern: GH1 generates 20- and 22-kd proteins (of approximately equal bioactivity), whereas GH2 yields a protein differing from GH1 in 13 amino acid residues.

Isolated GHD type IA (GH1-IA) results primarily from large deletions, with rare microdeletions and single base-pair substitutions of the GH1 gene that prevent synthesis or secretion of the hormone. GH1-IA is inherited as an autosomal recessive trait and affected individuals have profound congenital GHD. Because GH is not produced even in fetal life, patients are immunologically intolerant of GH and, typically, develop anti-GH antibodies when treated with either pituitary-derived or recombinant DNA-derived GH. When antibodies prevent patients from responding to GH, GH1-IA can be viewed as a form of GHI, and such patients are candidates for IGF-I therapy.

The less severe form of autosomal recessive GHD (isolated GHD type IB) also may result from mutations or rearrangements of the GH1 gene. These mutations cause production of an aberrant GH molecule that retains some function or at least generates immune tolerance. Patients usually respond to exogenous GH therapy without antibody production. The very low frequency (1.7%) of GH1 gene mutations in familial type IB isolated GHD suggests the importance of studying the GH1 gene promoter region.

In a group of 65 children with isolated GHD-IB, the GHRH receptor gene was normal in domains coding for the extracellular region, but more recently, mutations in the transmembrane and intracellular gene domains were found in 10% of families with isolated GHD-IB.

Isolated GHD-IL is inherited as an autosomal dominant trait. Such patients may have splice site, intronic, and missense mutations of the GH1 gene. The most common cause appears to be those mutations that inactivate the 5’ splice donor site of intron 3, resulting in skipping of exon 3. It is likely that they function in a dominant-negative manner with the GH mutant suppressing intracellular accumulation and secretion of wild-type GH.

In patients with missense mutations in exon 4 or 5, clinical presentation is quite variable with some evidence for reversibility of the impairment of intracellular GH storage and secretion by GH treatment.

Type III GHD, transmitted as an X-linked trait with associated hypogammaglobulinemia, has not yet been related to a mutation of the GH1 gene.

**Bisubunit Growth Hormone.**

Serum GH exists in multiple molecular forms, the consequences of alternative post-transcriptional or post-translational processing of the mRNA or protein, respectively. Some of these forms are presumed to have defects in the amino acid sequences required for binding of GH to its receptor and different molecular forms of GH may have varying potencies for stimulating skeletal growth, although this remains to be rigorously proven.

Short stature with normal GH immunoreactivity but reduced biopotency has been suggested, but the molecular abnormalities have been characterized only in a few such situations. In one child with extremely short stature (-6.1 SDS), a mutant GH caused by a single missense heterozygous mutation (cys to arg, codon 77 of GH1 gene) bound with greater affinity than normal to GHBP and the GHR and inhibited the action of normal GH. The child grew more (6 versus 3.9 cm/year) during a period of therapy with exogenous GH in moderate dosage. Strangely, the father had the same genetic abnormality but did not express the mutant hormone. In the second patient, with marked short stature (-3.6 SDS), a heterozygous A-to-G substitution on exon 4 of GH leads to a glycine to arginine substitution. This mutation is located in site 2 of GH molecular binding with its receptor and leads to failure of appropriate sequential receptor dimerization and subsequent diminished tyrosine phosphorylation and the GH-mediated intracellular cascade of events. Bioactivity determined in a mouse B cell lymphoma line was about 33% of immunoreactivity. Exogenous GH substantially increased growth velocity (4.5 to 11.0 cm/year).

There remain other patients, however, in whom diminished bioactivity by sensitive in vitro assays is not reflected by comparable immunoreactivity, but who do not have GH1 mutations, suggesting the importance of abnormal post-translational modifications of GH or other peripheral mechanisms.

**Trauma.**

See earlier topics.

**Inflammation.**

See earlier topics.

**Tumors Involving the Pituitary Gland.**

Many tumors that impair hypothalamic function also impair pituitary secretion of GH. In addition, craniopharyngiomas are a major cause of pituitary insufficiency. These tumors arise from remnants of Rathke’s pouch, the diverticulum of the roof of the embryonic oral cavity that normally gives rise to the anterior pituitary. Genetic defects in this condition, although certainly reasonable to suspect, have not yet been identified. This tumor is a congenital malformation present at birth and gradually grows over the ensuing years. The tumor arises from nests of squamous cells at the junction of the adenohypophysis and neurohypophysis, and it forms a cyst as it enlarges, which contains degenerated cells and may calcify but does not undergo malignant degeneration. The cyst fluid ranges from a “machinery oil” to a shimmering cholesterol-laden liquid, and the calcifications may be microscopic or gross. About 75% of craniopharyngiomas arise in the suprasellar region, the remainder resembling pituitary adenomas.

Craniopharyngiomas can cause manifestations at any age from infancy to adulthood but usually in middle childhood. The most common presentation is due to increased intracranial pressure, including headaches, vomiting, and ocular motor abnormalities. Visual field defects result from compression of the optic chiasm and papilledema or optic atrophy may be present. Visual and ophthalmic hallucinations have been reported, as have seizures and dementia. Most children with craniopharyngiomas show evidence of growth failure at the time of presentation.

GH and the gonadotropins are the most commonly affected pituitary hormones in children and adults, but deficiency of thyrotropin and/or ACTH may also occur; diabetes insipidus is present in 25% to 50%. Fifty percent to 80% of patients have abnormalities of at least one anterior pituitary hormone at diagnosis.
The initial report of primary GHI by Laron and colleagues described abnormalities of the Growth Hormone Receptor. Antibodies to GH; and (4) antibodies to the GHR. Illness-related GHI is discussed in the specific text sections.

Secondary GHI domain; (2) postreceptor abnormalities of GH signal transduction; (3) primary defects of IGF-I biosynthesis; and (4) genetic insensitivity to IGF-I action.

Primary GHI describes patients with the phenotype of GHD but with normal or elevated serum GH levels and diminished production of IGF-I.

Growth Hormone Insensitivity.

Partial GH insensitivity. GH insensitivity in the absence of dysmorphic features described by Laron et al.

IGF-I, insulin-like growth factor I; GH, growth hormone.

To supplement this revised classification, the following definitions are proposed:

GH insensitivity: Clinical and biochemical features of IGF-I deficiency and resistance to exogenous GH, associated with GH secretion that would not be considered abnormally low.

GH insensitivity syndrome: GH insensitivity associated with the recognizable dysmorphic features described by Laron et al.

Partial GH insensitivity. GH insensitivity in the absence of dysmorphic features described by Laron et al.

Operative intervention either via craniotomy or transsphenoidal resection may result in partial or almost complete removal of the lesion. Postoperative radiation, especially when tumor resection is incomplete, is commonly used. In some patients, especially those who become obese, a syndrome of normal linear growth without GH may occur. The circulating growth-promoting substances in this condition include insulin and other poorly characterized mitogens. The long-term childhood and adolescent consequences of craniohypophysealoma are substantial, with many quality of life issues exacerbating the hypopituitarism.

Pituitary adenomas (see Chapter 8) are infrequent during childhood and adolescence, accounting for fewer than 5% of operated patients at large centers. Nearly two thirds of tumors immunochemically stain for prolactin, and a small number stain for GH. GH-secreting pituitary adenomas are exceedingly unusual in youth. There is a variable experience as to the invasive nature of pituitary adenomas, although the prevailing opinion is that they are less aggressive in children than in adults. In 56 patients at the Mayo Clinic with non-ACTH secreting adenomas removed transsphenoidally, macroadenomas were about one third more frequent than microadenomas, with girls outnumbering boys 3.3 to 1. The incidence of hypopituitarism in patients with macroadenomas was about 50%, compared with none in patients with microadenomas; long-term cure rates were 55% to 65% for both tumor sizes.

These syndromes are characterized by an infiltration and accumulation of Langerhans cells in the involved areas, such as skull, hypothalamic-pituitary stalk, CNS, and viscera. Although these disorders, especially Hand-Schüller-Christian disease, are classically associated with diabetes insipidus, approximately 50% to 75% of patients in selected series have growth failure and GHD at the time of presentation. These individuals clearly have IGF-I deficiency.

Operative intervention either via craniotomy or transsphenoidal resection may result in partial or almost complete removal of the lesion. Postoperative radiation, especially when tumor resection is incomplete, is commonly used. In some patients, especially those who become obese, a syndrome of normal linear growth without GH may occur. The circulating growth-promoting substances in this condition include insulin and other poorly characterized mitogens. The long-term childhood and adolescent consequences of craniohypophysealoma are substantial, with many quality of life issues exacerbating the hypopituitarism.

The localized or generalized proliferation of mononuclear macrophages (histiocytes) characterizes Langerhans cell histiocytosis, a diverse disorder occurring at all ages, with peak incidence at ages 1 to 4 years. Endocrinologists are more familiar with the term histiocytosis X, which includes three related disorders: (1) solitary bony disease (eosinophilic granuloma), (2) Hand-Schüller-Christian disease (chronic disease with diabetes insipidus, exophthalmos, and multiple calvarial lesions), and (3) disseminated histiocytosis X (Letterer-Siwe, with widespread visceral involvement).

These syndromes are characterized by an infiltration and accumulation of Langerhans cells in the involved areas, such as skull, hypothalamic-pituitary stalk, CNS, and viscera. Although these disorders, especially Hand-Schüller-Christian disease, are classically associated with diabetes insipidus, approximately 50% to 75% of patients in selected series have growth failure and GHD at the time of presentation. In contrast, only 1% of unselected children with Langerhans cell histiocytosis living in Canada during a 15-year period had GHD. Isolated GHD or deficiencies of other anterior pituitary hormones may occur.

TABLE 23-7 -- Clinical Features of Growth Hormone Insensitivity

<table>
<thead>
<tr>
<th>Birth weight: near normal</th>
<th>Primary GHI insensitivity (hereditary defects)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. GH receptor defect (may be positive or negative for GH-binding protein)</td>
<td></td>
</tr>
<tr>
<td>(1) Extracellular mutation</td>
<td></td>
</tr>
<tr>
<td>(2) Cytoplasmic mutation</td>
<td></td>
</tr>
<tr>
<td>(3) Intracellular mutation</td>
<td></td>
</tr>
<tr>
<td>b. GH signal transduction defect (distal to cytoplasmic domain of GH receptor)</td>
<td></td>
</tr>
<tr>
<td>c. Insulin-like growth factor I (IGF-I) synthetic defect (IGF-I gene deletion)</td>
<td></td>
</tr>
<tr>
<td>d. IGF-I transport defect</td>
<td></td>
</tr>
<tr>
<td>e. IGF-I receptor defect</td>
<td></td>
</tr>
<tr>
<td>f. Bioinactive GH molecule</td>
<td></td>
</tr>
<tr>
<td>2. Secondary GH insensitivity (acquired defects)</td>
<td></td>
</tr>
<tr>
<td>a. Circulating antibodies to GH that inhibit GH action</td>
<td></td>
</tr>
<tr>
<td>b. Antibodies to the GH receptor</td>
<td></td>
</tr>
<tr>
<td>c. GH insensitivity caused by conditions such as malnutrition, liver disease, catabolic states</td>
<td></td>
</tr>
<tr>
<td>d. Other conditions that cause GH insensitivity</td>
<td></td>
</tr>
</tbody>
</table>

To supplement this revised classification, the following definitions are proposed:

GH insensitivity: Clinical and biochemical features of IGF-I deficiency and resistance to exogenous GH, associated with GH secretion that would not be considered abnormally low.

GH insensitivity syndrome: GH insensitivity associated with the recognizable dysmorphic features described by Laron et al.

Partial GH insensitivity. GH insensitivity in the absence of dysmorphic features described by Laron et al.

TABLE 23-6 -- Proposed Classification of Growth Hormone Insensitivity Syndromes

<table>
<thead>
<tr>
<th>1. Primary GH insensitivity (hereditary defects)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. GH receptor defect (may be positive or negative for GH-binding protein)</td>
</tr>
<tr>
<td>(1) Extracellular mutation</td>
</tr>
<tr>
<td>(2) Cytoplasmic mutation</td>
</tr>
<tr>
<td>(3) Intracellular mutation</td>
</tr>
<tr>
<td>b. GH signal transduction defect (distal to cytoplasmic domain of GH receptor)</td>
</tr>
<tr>
<td>c. Insulin-like growth factor I (IGF-I) synthetic defect (IGF-I gene deletion)</td>
</tr>
<tr>
<td>d. IGF-I transport defect</td>
</tr>
<tr>
<td>e. IGF-I receptor defect</td>
</tr>
<tr>
<td>f. Bioinactive GH molecule</td>
</tr>
<tr>
<td>2. Secondary GH insensitivity (acquired defects)</td>
</tr>
<tr>
<td>a. Circulating antibodies to GH that inhibit GH action</td>
</tr>
<tr>
<td>b. Antibodies to the GH receptor</td>
</tr>
<tr>
<td>c. GH insensitivity caused by conditions such as malnutrition, liver disease, catabolic states</td>
</tr>
<tr>
<td>d. Other conditions that cause GH insensitivity</td>
</tr>
</tbody>
</table>

To supplement this revised classification, the following definitions are proposed:

GH insensitivity: Clinical and biochemical features of IGF-I deficiency and resistance to exogenous GH, associated with GH secretion that would not be considered abnormally low.

GH insensitivity syndrome: GH insensitivity associated with the recognizable dysmorphic features described by Laron et al.

Partial GH insensitivity. GH insensitivity in the absence of dysmorphic features described by Laron et al.

TABLE 23-7 -- Clinical Features of Growth Hormone Insensitivity

<table>
<thead>
<tr>
<th>Birth weight: near normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary GHI insensitivity (hereditary defects)</td>
</tr>
<tr>
<td>a. GH receptor defect (may be positive or negative for GH-binding protein)</td>
</tr>
<tr>
<td>(1) Extracellular mutation</td>
</tr>
<tr>
<td>(2) Cytoplasmic mutation</td>
</tr>
<tr>
<td>(3) Intracellular mutation</td>
</tr>
<tr>
<td>b. GH signal transduction defect (distal to cytoplasmic domain of GH receptor)</td>
</tr>
<tr>
<td>c. Insulin-like growth factor I (IGF-I) synthetic defect (IGF-I gene deletion)</td>
</tr>
<tr>
<td>d. IGF-I transport defect</td>
</tr>
<tr>
<td>e. IGF-I receptor defect</td>
</tr>
<tr>
<td>f. Bioinactive GH molecule</td>
</tr>
<tr>
<td>2. Secondary GH insensitivity (acquired defects)</td>
</tr>
<tr>
<td>a. Circulating antibodies to GH that inhibit GH action</td>
</tr>
<tr>
<td>b. Antibodies to the GH receptor</td>
</tr>
<tr>
<td>c. GH insensitivity caused by conditions such as malnutrition, liver disease, catabolic states</td>
</tr>
<tr>
<td>d. Other conditions that cause GH insensitivity</td>
</tr>
</tbody>
</table>

To supplement this revised classification, the following definitions are proposed:

GH insensitivity: Clinical and biochemical features of IGF-I deficiency and resistance to exogenous GH, associated with GH secretion that would not be considered abnormally low.

GH insensitivity syndrome: GH insensitivity associated with the recognizable dysmorphic features described by Laron et al.

Partial GH insensitivity. GH insensitivity in the absence of dysmorphic features described by Laron et al.

TABLE 23-6 -- Proposed Classification of Growth Hormone Insensitivity Syndromes

<table>
<thead>
<tr>
<th>1. Primary GH insensitivity (hereditary defects)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. GH receptor defect (may be positive or negative for GH-binding protein)</td>
</tr>
<tr>
<td>(1) Extracellular mutation</td>
</tr>
<tr>
<td>(2) Cytoplasmic mutation</td>
</tr>
<tr>
<td>(3) Intracellular mutation</td>
</tr>
<tr>
<td>b. GH signal transduction defect (distal to cytoplasmic domain of GH receptor)</td>
</tr>
<tr>
<td>c. Insulin-like growth factor I (IGF-I) synthetic defect (IGF-I gene deletion)</td>
</tr>
<tr>
<td>d. IGF-I transport defect</td>
</tr>
<tr>
<td>e. IGF-I receptor defect</td>
</tr>
<tr>
<td>f. Bioinactive GH molecule</td>
</tr>
<tr>
<td>2. Secondary GH insensitivity (acquired defects)</td>
</tr>
<tr>
<td>a. Circulating antibodies to GH that inhibit GH action</td>
</tr>
<tr>
<td>b. Antibodies to the GH receptor</td>
</tr>
<tr>
<td>c. GH insensitivity caused by conditions such as malnutrition, liver disease, catabolic states</td>
</tr>
<tr>
<td>d. Other conditions that cause GH insensitivity</td>
</tr>
</tbody>
</table>

To supplement this revised classification, the following definitions are proposed:

GH insensitivity: Clinical and biochemical features of IGF-I deficiency and resistance to exogenous GH, associated with GH secretion that would not be considered abnormally low.

GH insensitivity syndrome: GH insensitivity associated with the recognizable dysmorphic features described by Laron et al.

Partial GH insensitivity. GH insensitivity in the absence of dysmorphic features described by Laron et al.

The initial report of primary GHI by Laron and colleagues described...
Osteopenia

Birth length: may be slightly decreased

Postnatal growth: severe growth failure

Bone age: delayed, but may be advanced relative to height age

Genitalia: microopenis in childhood; normal for body size in adults

Puberty: delayed 37 years

Sexual function and fertility: normal

**Craniofacies**

- Hair: sparse before age 7 years
- Forehead: prominent; frontal bossing
- Skull: normal head circumference; craniofacial disproportion due to small facies
- Faces: small
- Nasal bridge: hypoplastic
- Orbit: shallow
- Dentition: delayed eruption
- Scleras: blue
- Voice: high-pitched

**Musculoskeletal/Metabolic/Miscellaneous**

- Hypoglycemia: in infants and children; fasting symptoms in some adults
- Walking and motor milestones: delayed
- Hips: dysplasia; avascular necrosis of femoral head
- Elbow: limited extensibility
- Skin: thin, prematurely aged
- Osteopenia

"three siblings with hypoglycemia and other clinical and laboratory signs of GHD, but with abnormally high levels of immunoreactive serum GH." To date, approximately 250 cases have been identified worldwide, most from the Mediterranean region or from Ecuador (probably from Spanish Conversos, or Jews who converted to Christianity during the Inquisition). These individuals do not respond to exogenous GH, in terms of growth, metabolic changes, or increases in serum levels of IGF-I and IGFBP-3. Cellular unresponsiveness to GH was demonstrated in vitro by the failure of GH to stimulate erythroid progenitor cells from the peripheral blood of patients. Direct evidence of receptor dysfunction was provided by the demonstration that microsomes obtained by liver biopsy do not bind radiolabeled GH. GHBP activity is usually (in 75% to 80% of cases) undetectable in the sera of patients with this disorder.

Studies of the GHR gene in Israeli patients indicate that some, but not most, contained gene deletions, and a wide variety of homozygous point mutations in this gene (missense, nonsense, and abnormal splicing) have been identified subsequently. Most of the mutations are in the extracellular (GH binding) domain of the GHR; at least one mutation of the extracellular domain does not affect GH binding but prevents dimerization of the receptor. Mutations of the intracellular and transmembrane domains are much less frequent.

One patient had two separate amino acid substitutions in the intracellular domain, but because both mutations are on the same allele, which comes from the unaffected mother, the diagnosis of GHI in this patient was not certain. Chujo and associates, additionally, found that a specific heterozygous missense exon 10 (which codes for most of the GHR intracellular domain) mutation was present in 14 of 96 volunteers; there was no significant effect on stature, suggesting that this mutation represents a normal polymorphism. Another subject with compound heterozygous mutations in exon 10 was extremely short (-4 SDS) but grew in response to a very high dose of GH. At this time, data are inadequate to determine whether these intracellular domain substitutions represent genuine mutations or innocent polymorphisms.

Another profoundly short patient, fully resistant to GH but highly responsive to IGF treatment, had coexistent heterozygous mutations affecting exons 6 and 9. A similarly short girl (-8 SDS) had compound heterozygosity in exons 8 and 10.

Woods and colleagues described two cousins with severe GHR and homozygous mutations at the 5' splice donor site of intron 8, resulting in a mutant GHR without functional transmembrane or intracellular domains. A similar defect was found in a Druse girl with a mutation of the 3' acceptor site of intron 7. Serum levels of GHBP were elevated because the mutant receptor protein apparently becomes detached from the cell receptor surface.

Two defects directly affecting the intracellular domain have been reported to result in dominantly inherited GHI. A, B The
patients and their mother had a heterozygous point mutation that disrupted the 5’ splice donor site of intron 9, causing skipping of exon 9 and the appearance of a premature stop codon in exon 10 and resulting in the same GH 1277 receptor molecule as described by Ayling and co-workers. Under in vitro conditions, the Japanese mutation has been shown to result in a GHR molecule that behaves in a dominant negative manner, inhibiting GH-induced tyrosine phosphorylation of STAT5.

Heterozygosity for defects of the GHR may cause relative GHRI, with modest growth occurring only in response to high doses of GH. Such observations raise the important question of whether heterozygosity for GHRI can result in a clinically important phenotype and whether some children labeled as “diopathic short stature” (ISS) may harbor such mutations. Given the requirement for dimerization of the GHR, there is the potential for an abnormal protein to have varying degrees of dominant negative effect. Ross and associates described a truncated (1-279) GHR splice variant whose differential production could act to regulate GHBP production and, more importantly, to modulate GHR signaling in a negative fashion.

In summary, the clinical features of GHR due to GHR deficiency are identical to those of other forms of severe IGFI deficiency, such as congenital GHDI. As with GHDI, however, there is a wide range of clinical phenotypes. Basal serum GH levels are typically elevated in children but may be normal in adults. Most patients have decreased serum GHBP levels, but a normal or even elevated serum GHBP concentration does not exclude the diagnosis of GHRD because mutations of the GHR dimerization domain and in the intracellular domain have been described. Patients with measurable GHBP tend to be taller. Serum IGFI-I, IGFI-II, and IGFBP-3 levels are profoundly reduced, but partial clinical and biochemical phenotypes have been described, typically but not always related to milder mutations of the GHR gene, resulting in only a modest reduction in binding activity and receptor action.

Primary Defects of IGFI Biosynthesis.

Woods and colleagues described a 15-year-old boy with a partial deletion of the IGFI-gene yielding a truncated IGFI-molecule. Severe prenatal and postnatal (-6.7 SDS) growth retardation and insensitivity to exogenous GH were consistent with the expectations of the phenotype of IGFI-deficiency. Somnolineal deafness, mental retardation, and microcephaly suggest a role for prenatal IGFI in CNS development. Hyperinsulinism and insulin resistance were presumably due to overproduction of GH. IGFI-I levels were exceedingly low, but IGFBP-3 and GHBP levels were normal. The patient was homozygous for deletions of exons 4 and 5 of the IGFI-gene, with both parents being heterozygous carriers and perhaps mildly affected themselves. Although unresponsive to GH therapy, the patient was able to achieve accelerated growth velocity, improved body composition, increased bone mineralization, and decreased insulin resistance on treatment with IGFI-I.

Primary Defects of IGFI Transport and Cleavage.

In fibroblasts from a single short child of 127 studied, Tollesen and co-workers found a marked resistance to IGFI-stimulated -amininosobutyric acid uptake and thymidine incorporation. And IGFI-variant with 60-fold lower binding affinity for IGFBPs stimulated the fibroblasts of this child and normal children, thus eliminating a primary IGFI-receptor defect. This patient's fibroblasts secreted more IGFBPs than normal and had a 10-fold increase in a cell surface protein similar in size though not immunoreactivity to IGFBP-1.

Barreca and associates studied a short boy (-6 SDS) with increased GH, normal IGFI-I, and 20- to 30-fold elevated IGFBP-1 levels. Growth failure seemed due to inhibition of IGFI-action by IGFBP-1. Short-term treatment with GH led to suppression of IGFBP-1, increased ternary-complexed IGFI-I, and a marked increase of growth rate.

Primary Defects of IGFI-Receptor Production or Responsiveness.

In mouse knockout models, homozygous mutations of the IGFI-receptor result in profound growth failure and neonatal mortality. Heterozygous mutations are phenotypically similar to wild-type mice. In the African Efe pygmies, a series of studies demonstrated extreme insensitivity to the in vitro growth-enhancing effects of IGFI-I, Reduced IGFI-I receptor transplants and sites with resultant diminished tyrosine phosphorylation and postreceptor signaling, although no definable receptor mutation, are suggested explanations.

In leprochaunism, a syndrome of growth failure and insulin receptor dysfunction, IGFI-insensitivity is variable. The profound abnormality of the insulin receptor suggests that heterodimeric insulin-receptor and IGFI-receptor combinations could possibly lead to failed activation of the IGFI-signaling cascade. As the IGFI-receptor gene resides at 15q26.3, deletions of the distal long arm of chromosome 15 or ring chromosome 15 may lead to hemizygosity for the IGFI-receptor. Although such patients may have IGFR1 and skipping postnatal growth failure, lack of a biologic response to IGFI-I has not been conclusively demonstrated. Whether growth failure in such patients is due to altered levels of IGFI-receptor or represents the net effect of the loss of other genes located on 15q remains to be determined.

Figure 23-51 Serum levels of insulin-like growth factor I (IGFI-I), IGFI-II, IGFI-binding protein 2 (IGFBP-2) and IGFBP-3 in patients with GH receptor deficiency from Ecuador. (From Rosenfield RG, Rosenbloom AL, Guevara-Aguirre J. Growth hormone [GH] resistance due to GH receptor deficiency. Endocr Rev 1994; 15:369-390. © The Endocrine Society.)
TABLE 23-8—Tests to Provoke Growth Hormone Secretion*

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Dosage</th>
<th>Times Samples Are Taken (minutes)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exercise</td>
<td>Step climbing: exercise cycle for 10 min.</td>
<td>0, 10, 20</td>
<td>Observe child closely when on the steps</td>
</tr>
<tr>
<td>Levodopa</td>
<td>&lt;15 kg: 125 mg</td>
<td>0, 60, 90</td>
<td>Nausea, rarely emesis</td>
</tr>
<tr>
<td></td>
<td>1030 kg: 250 mg</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&gt;30 kg: 500 mg</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clonidine</td>
<td>0.15 mg/kg/m²</td>
<td>0, 30, 60, 90</td>
<td>Tiredness, postural hypoponosion</td>
</tr>
<tr>
<td>Arginine HCl (IV)</td>
<td>0.5 g/kg (max 30 g)</td>
<td>0, 15, 30, 45, 60</td>
<td></td>
</tr>
<tr>
<td>Insulin (IV)</td>
<td>0.050–1.0 unit/kg</td>
<td></td>
<td>Hypoglycemia, requires close supervision</td>
</tr>
<tr>
<td>Glucagon (IM)</td>
<td>0.03 mg/kg (max 1 mg)</td>
<td></td>
<td>Nausea, occasional emesis</td>
</tr>
<tr>
<td>GHHR (IV)</td>
<td>1 g/kg</td>
<td></td>
<td>Flushing, metallic taste</td>
</tr>
</tbody>
</table>

GH deficiency, growth hormone deficiency; GHRH, growth hormonereleasing hormone; IM, intramuscular; IV, intravenous.

*Tests should be performed after an overnight fast. Many investigators suggest that prepubertal children should be "primed" with gonadal steroids, e.g., a 5 mg Prenalin orally the night before and the morning of the test with 500 mg of 10 mg/kg ethinyl estradiol orally 3 consecutive days before testing. This, of course, alters the patient's steady state and performs the provocative test in a steroid-rich environment. Patients must be euthyroid at the time of testing.

Insulin-induced hypoglycemia is a potential risk of this procedure, which is designed to lower the blood glucose by at least 50%. Documentation of appropriate lowering of blood glucose is recommended. If GHD is suspected, the lower dosage of insulin is usually administered, especially in infants. Glucagon should be available.

involves at least two hypothalamic factors—GHRH and somatostatin—and multiple other peptides and neurotransmitters. Spontaneous GH secretion varies with gender, age, pubertal stage, and nutritional status, all of which must be factored into the evaluation of GH production.

Between normal pulses of GH secretion, serum GH levels are low (often <0.1 µg/L), below the limits of sensitivity of most conventional assays (usually <0.2 µg/L). Accordingly, measurement of a random serum GH concentration is virtually useless in diagnosing GHD but may be useful in the diagnosis of GHI and GH excess. Measurement of GH "secretory reserve," therefore, relies on the use of physiologic or pharmacologic stimuli, and such provocative tests have been the basis for the diagnosis of GHD for more than 30 years. Physiologic stimuli include fasting, sleep, exercise, and pharmacologic stimuli include levodopa, clonidine, glucagon, propranolol, arginine, and insulin.

Stimulation tests have often been divided into screening tests (exercise, fasting, levodopa, clonidine), which are characterized by ease of administration, low toxicity, low risk, and low specificity, and definitive tests (arginine, insulin, glucagon). To improve specificity, provocative tests are customarily combined or given sequentially. We have, for example, often used fasting plus oral clonidine as a screening test, to be followed by fasting plus intravenous sequential arginine, oral clonidine, levodopa, or insulin as a definitive test. It is generally accepted that a child must "fail" provocative tests with at least two separate stimuli to be considered GHD. Standard provocative GH tests are summarized in Table 23-8.

Although provocative GH testing has been the foundation for the diagnosis of GHD since GH assays first became available, they have come under criticism for a number of reasons:

1. Provocative GH testing is nonphysiologic. None of the standard pharmacologic provocative tests satisfactorily mimic the normal secretory pattern of pituitary GH. Even when normal secretory regulatory peptides are used for stimuli, the routine, route of administration, and interactions with other regulatory factors are artificial. Furthermore, because most endocrine centers use several different stimulation tests, there is no validated means of resolving conflicting data from two or more provocative tests. To emphasize this point, Guyda reported on 6373 GH stimulation tests performed on 3233 short French children; 11 different pharmacologic tests were employed, with 62 of the possible 66 pairs employed at least once and the most frequent combination of tests used only in 12.7% of patients.

2. Arbitrary definitions of "subnormal" response to provocative tests. Different centers vary in the definition of a "normal" response to stimulation tests. Although early reports generally employed a cut-off level of 2.5 µg/L, this cut-off was gradually increased to 7 µg/L and with the availability of recombinant DNA-derived GH, increased to 10 µg/L, although there are no data for validating higher arbitrary cut-off values. The initial levels of GH that were used to define GHD were based on the study of patients with profound classic findings or organic destruction of the adenohypophysis. The lack of documentation of defined normal responses can be seen in the use of vague terminology such as "lack of adequate endogenous growth hormone secretion" and "inadequate secretion of normal endogenous growth hormone." There are multiple new GH assays that measure GH immunoreactivity at 33% to 50% of earlier assays, but there has not been a systematic reassessment of "new normal" GH cut-off levels nor recognition by many endocrinologists of which assay their center might be using.

3. Age dependency and use of gonadal steroids. Serum GH levels typically rise during puberty, typically because of an increase in pulse amplitude rather than an increase in pulse frequency. GH secretion may normally be so low as to blur the differentiation between GHD and CDGM. Many children who "fail" provocative testing before the onset of puberty prove to have normal GH secretion after puberty or after administration of exogenous gonadal steroids. In a placebo-controlled comparison of estrogen priming in children with GHD and ISS, supramaximal provocative tests performed 3 days before the tests performed (using 9 µg/L as a cut-off with a polyclonal GH assay) had a diagnostic efficiency of 98%. A well-controlled study of provocative GH testing in children of normal stature documented the inherent problems of such testing and the need for standardization of gonadal steroid administration during stimulation tests in peripubertal children. When exercise and arginine-insulin stimulation tests were administered to these normal children, the lower limit of normal (-2 SD) for peak serum GH concentration in prepubertal children was only 1.9 µg/L, whereas in children of Tanner stage 5 puberty, this level was 9.3 µg/L. When estrogen was administered prior to provocative testing, the lower 95% confidence limit for the normal serum GH range rose to 7.2 µg/L. When estrogen was not administered, the serum GH level did not rise above 7 µg/L, during three provocative tests. These normally growing children could, potentially, be erroneously labeled as GHD in 61% of normal prepubertal children. Furthermore, the finding of similar GH values in slow-growing, short children emphasizes the difficulty of basing this important diagnosis on provocative test data that use a "magic" number as an arbitrary cut-off for normal.

4. Variability of GH assays limit discriminatory power. Several studies have demonstrated as much as threefold variability in the measurement of serum GH levels among established laboratories. This is explained, at least in part, by the presence of several molecular forms of GH in serum and by the use of different monoclonal antibodies in contrast with older polyclonal antibodies and variations in the choice of standards, labeling techniques, and assay buffers (matrix). The consequence is that children labeled as GHD by one assay are considered normal by another. This is an unacceptable situation for clinicians, who must remain aware of the type and source of GH assay being used by a given laboratory. A highly sensitive immunofunctional GH assay has now been developed that measures concentrations of GH capable of binding to GHBP. It is not clear, however, that such assays necessarily have any advantages over standard radioimmunoassays for routine GH measurements. When arginine/levodopa or arginine/insulin stimulation tests were used, approximately 50% of normal children had peak GH concentrations below 7 ng/mL and 30% were below 5 ng/mL, whether GH was measured by immunofunctional assay or ELISA.
5. Expense, discomfort, and risks of provocative GH testing. Provocative testing typically requires multiple timed blood samples and the parental administration of drugs. The resulting expense and discomfort to the patient are self-evident. In addition, tests involving insulin administration carry the risk of hypoglycemia and seizures and should be performed only by experienced medical personnel under appropriate supervision. Deaths have been reported from insulin-induced hypoglycemia and from its overly vigorous correction with parental glucose. 1029

6. Poor reproducibility of provocative tests. The reproducibility of provocative GH tests has never been adequately documented, even when GH levels are measured with the same assay. 1030

Another diagnostic approach involves measurement of spontaneous GH secretion. This can be done either by multiple sampling (every 5 to 30 minutes) over a 12- to 24-hour period or by continuous blood withdrawal over 12 to 24 hours. 1031 The former method allows one to evaluate and characterize GH pulsatility, whereas the latter permits only determination of mean GH concentration. Both approaches are subject to many of the same limitations as provocative GH testing. The expense and discomfort of such testing are obvious, and although it was thought that this approach is more reproducible than provocative GH tests, variability is a problem. 1032 The ability of such tests to discriminate between GHD and normal short children is also an issue.

Rose and colleagues reported that measurement of spontaneous GH secretion identified only 57% of children with GHD as defined by provocative testing. Similarly, Lanes reported that one fourth of normally growing children had low overnight GH levels, and a longitudinal study of normal boys through puberty demonstrated a wide intersubject variance, including many "low" 24-hour GH production rates, despite fully normal growth. 1033

Given the problems with GH testing, it is not surprising that provocative tests and 24-hour GH profiles do not always correlate. It is likely that 12- to 24-hour GH profiles can identify most children with GHD and is superior, both in sensitivity and specificity, to provocative GH testing. "Neurosecretory dysfunction" probably does exist in children after cranial radiation and likely does characterize a subgroup of children with GHD and IGF deficiency. The expense and discomfort of such tests is likely to outweigh their value.

The measurement of GH levels in urine is an alternative means of estimating integrated GH secretion (or, at least, excretion). This technique requires timed urine collections and anti-GH antibodies of high affinity, because urinary GH levels are normally low. Adequate age- and gender-related standards have not been developed, and the diagnostic use of urinary GH determination remains to be defined.

An alternative means of diagnosing GHD involves assessment of IGF-I and IGF-II and their binding proteins. Bone loss is not uncommon in children with GHD, and normal serum IGF-I levels had reduced IGFBP-3 levels, and 10 (43%) of 23 normal short children had decreased serum IGFBP-3. For example, in one study, the sensitivity of the IGFBP-3 assay in complete GHD (peak GH <5 µg/L) was 93%; however, it was only 43% in partial GHD (peak GH 5 to 10 µg/L). The utility of IGFBP-3 assays in the diagnosis of GHD was evaluated by Blum and colleagues, who found that serum IGFBP-3 levels were below the 5th percentile in 80% of children with GHD and normal serum IGF-I levels had reduced IGFBP-3 levels, and 10 (43%) of 23 normal short children had decreased serum IGFBP-3. In the latter study, 18% of patients with low provocative GH levels had IGFBP-3 concentrations in the normal range, but only 4% of "GHD" patients had normal serum levels of both IGF-I and IGF-II. Serum levels of IGF-I and IGF-II were both reduced in only 0.5% of normal children and in 11% of normal short children.

The assay of GH-dependent IGFBP-3, normally the major serum carrier of IGF, is an additional means of diagnosing IGF deficiency caused by GHD because the concentrations of IGFBP-3 correlate with the sum of the levels of IGF-I and IGF-II:

1. The immunoassay of IGFBP-3 is technically simple and does not require separation of the binding protein from IGF peptides.
2. Normal serum levels of IGFBP-3 are high, typically in the range of 1 to 5 mg/L, so that assay sensitivity is not an issue. However, IGFBP-3 assays do have the following potential limitations:
   1. IGFBPs may interfere with radioimmunoassays, RIA, and bioassays. These binding proteins either must be removed by acid gel chromatography (which is labor-intensive) or must be blocked by the addition of excess IGF-II (which requires a high-affinity, high-specificity antibody for IGF-I). An alternative approach is to employ a radiolabeled IGF-1 analogue with reduced affinity for IGFBPs.
   2. Serum IGF-I levels are age-dependent, being lowest in young children (<5 years of age), during which one must wishes to have an accurate diagnostic test.
   3. Serum IGFBP-3 levels may be low in conditions other than GHD, such as primary GH (Laron's syndrome) and secondary GHI (e.g., malnutrition, liver disease).
   4. Serum concentrations of IGF-I (and IGFBP-3) are often normal in adult-onset GHD and in children with GHD resulting from brain tumors and cranial radiation. 1034
   5. Interlaboratory differences of absolute IGF values, although not as striking as with GH, may be substantial.

Even when these caveats are considered, the correlation between serum IGF-I levels and provocative or spontaneous GH measurements is imperfect. In a group of children younger than 10 years of age, IGF-I levels were below -2 SD in only 8 of 15 children with a diagnosis of GHD based on provocative testing (53.3% sensitivity) and normal in 47 of 48 children with a normal GH response (97.9% specificity). In one study, 18% of patients with low provocative GH levels had IGFBP-3 concentrations in the normal range, but only 4% of "GHD" patients had normal serum levels of both IGF-I and IGF-II. Serum levels of IGF-I and IGF-II were both reduced in only 0.5% of normal children and in 11% of normal short children.

The correlation between IGF-I and IGFBP-3 levels and assessments of spontaneous GH secretion is also imperfect. Even in normal children, the correlation between 24-hour GH secretion and serum IGF-I and IGFBP-3 levels is modest (r = 0.78 and 0.62, respectively). It is not possible to resolve fully conflicts between assay of IGF-I levels and measurement of GH secretion, since there is no definitive way to diagnose GHD; however, studies of patients with GHI have developed useful methods to diagnose GHD and to estimate the degree of GHI. Although such patients may have normal or elevated serum GH levels, mutations or deletions of the GHI gene render them unresponsive to GH, making them functionally deficient. In approximately 70 instances of GHI gene mutations, all had markedly reduced serum levels of both IGF-I and IGFBP-3. Even so, both IGF-I and IGFBP-3 correlated significantly with height. Measurements of IGF-I and IGFBP-3 levels have been used in other studies to establish the diagnosis of GHD.
velocity. Figure 23-52 provides an algorithm for the evaluation of the child with growth failure.

The many illness-related causes of diminished growth have been discussed in prior sections. The possibility of hypothalamic or pituitary dysfunction should always be considered in children with documented growth deceleration, particularly in the face of known or suspected CNS pathology (e.g., tumors, radiation, malformations, infection, trauma, blindness, nystagmus). Similarly, the neonate with hypoglycemia or microphallus warrants evaluation of pituitary function (including MRI), and children with documented thyropin, ACTH, antidiuretic hormone, or gonadotropin deficiency are candidates for GHD. For children with proportional short stature and documented growth deceleration, assessment of serum IGF-I and IGFBP-3 is warranted, and, based on the results, the possibilities of hypothalamic dysfunction, pituitary insufficiency, and GHI can be investigated.

Recent recommendations by the Growth Hormone Research Society (GRS) for defining GHD recognize no "gold standard" for that diagnosis and propose that in a child with slow growth, whose history and auxology suggest GHD (Table 23-9), testing for GH/IGF-I deficiency requires the measurement of IGF-I and IGFBP-3 levels as well as GH provocation tests (after hypothryoidism has been excluded). In cases of suspected isolated GHD, two GH provocation tests (sequential or on separate days) are required: in those with defined CNS pathology, history of radiation, CPHD, or a genetic defect, one GH test will suffice. In patients who have had cranial radiation or malformations of the hypothalamic-pituitary unit, GHD may evolve over years, and its diagnosis requires serial testing. Some patients with auxology suggestive of GHD, however, may have IGF-I and/or IGFBP-3 levels below the normal range on repeated tests but GH responses in provocation tests above the cut-off level. Such children do not have classic GHD but nonetheless may have an abnormality of the GH-IGF axis and, after the exclusion of systemic disorders affecting the synthesis or action of IGF-I, might be considered for GH treatment. A cranial MRI, with particular attention to the hypothalamic-pituitary region, should be carried out in any child given a diagnosis of GHD. These recommendations stress the importance of rational clinical judgment rather than specific tests in characterization of childhood GHD.

Ultimately, the diagnosis of GHD (or IGF deficiency) should be made by on the basis of combined clinical and laboratory criteria. Short children who have well-documented normal height velocities do not, generally, require evaluation of GH secretion, and the finding of normal serum levels of IGF-I and IGFBP-3 is confirmatory. Children with Turner's syndrome and short stature should not be required to undergo GH testing to qualify for GH therapy, because such treatment is not predicated on abnormal GH secretion. On the other hand, the child with documented growth deceleration requires further evaluation, even if tests of GH secretion appear normal. Documentation of decreased serum IGF-I and IGFBP-3 levels would then substantiate the diagnosis of IGF deficiency, and the differential diagnoses of GHD and GHI would need to be considered.

The child with a history of cranial irradiation, decreased height velocity, and reduced serum levels of IGF-I and IGFBP-3 should be considered to have GHD (or GHI), even in the face of normal provocative tests. Alternatively, such patients may have normal IGFBP-3 levels with low GH levels in pharmacologic tests. This approach still leaves a place for measurements of GH secretion. Such determinations are critical for distinguishing between GHD and GHI as causes of IGF deficiency. Documentation of abnormal pituitary GH secretion raises the possibility of intracranial tumors and the potential for deficiency of other pituitary hormones. Evaluation

TABLE 23-9 – Key History and Physical Examination Findings in Growth Hormone Deficiency (Growth Hormone Research Society)

<table>
<thead>
<tr>
<th>Cranial radiation</th>
<th>Head trauma or central nervous system infection</th>
<th>Consanguinity and/or an affected family member</th>
</tr>
</thead>
<tbody>
<tr>
<td>Craniofacial midline abnormalities</td>
<td>Severe short stature (&lt; -3 SD)</td>
<td>Height &lt; -2 SD and a height velocity over 1 year &lt; -1 SD</td>
</tr>
<tr>
<td>A decrease in height SD of &gt;0.5 over 1 year in children &gt;2 years of age</td>
<td>A height velocity &lt; -2 SD over 1 year</td>
<td>A height velocity &gt;1.5 SD below the mean sustained over 2 years</td>
</tr>
<tr>
<td>Signs indicative of an intracranial lesion</td>
<td>Signs of multiple pituitary hormone deficiency</td>
<td>Neonatal symptoms and signs of growth hormone deficiency</td>
</tr>
</tbody>
</table>


for GHD permits concomitant assessment of ACTH/cortisol secretion during insulin-induced hypoglycemia.

The diagnosis of GHD in a newborn is especially challenging. The presence of microopenis in a male newborn should always lead to an evaluation of the GH-IGF axis. A GH level must be measured in the presence of neonatal hypoglycemia occurring in the absence of a metabolic disorder, such as hyperammonemia or carnitine deficiency syndromes. A level below 20 mg/L in a polyclonal radioimmunoassay suggests GHD. The use of standard GH stimulation tests, except for the glucagon test, is not recommended in neonates. Normative data are not available for stimulated serum GH levels, but a cut-off of 25 ng/mL is probably appropriate and stimulated values lower than 20 ng/mL certainly should raise suspicion. MRI is essential when the diagnosis is suspected, and useful clinical information defining developmental abnormalities of the hypothalamic-pituitary area may be available sooner than GH assay data. An IGFBP-3 level is of value for the diagnosis of neonatal GHD, but IGF-I levels are rarely helpful. In fact, serum IGFBP-3 should be performed as the test of choice in suspected neonatal GHD.

In summary, a child should be considered a candidate for GH therapy if he or she meets one of these auxologic criteria, supported by biochemical evidence of GHD based on sex steroidprimed provocative tests or evidence of IGF deficiency based on measurement of IGF-I and IGFBP-3 concentrations. Such patients should also

---

**TABLE 23-9 – Key History and Physical Examination Findings in Growth Hormone Deficiency (Growth Hormone Research Society)**

<table>
<thead>
<tr>
<th>Cranial radiation</th>
<th>Head trauma or central nervous system infection</th>
<th>Consanguinity and/or an affected family member</th>
</tr>
</thead>
<tbody>
<tr>
<td>Craniofacial midline abnormalities</td>
<td>Severe short stature (&lt; -3 SD)</td>
<td>Height &lt; -2 SD and a height velocity over 1 year &lt; -1 SD</td>
</tr>
<tr>
<td>A decrease in height SD of &gt;0.5 over 1 year in children &gt;2 years of age</td>
<td>A height velocity &lt; -2 SD over 1 year</td>
<td>A height velocity &gt;1.5 SD below the mean sustained over 2 years</td>
</tr>
<tr>
<td>Signs indicative of an intracranial lesion</td>
<td>Signs of multiple pituitary hormone deficiency</td>
<td>Neonatal symptoms and signs of growth hormone deficiency</td>
</tr>
</tbody>
</table>

undergo MRI studies of the hypothalamus-pituitary and assessment of other pituitary hormone deficiencies. It is understood that this approach will result in GH treatment of some children with idiopathic, isolated GHD or IGF deficiency and that such cases require careful monitoring of both pituitary status and responsiveness to GH treatment. The latter can be assessed relatively to recently developed predictive models [1251,1252] and the diagnosis of GHD reconsidered in the child with idiopathic, isolated GHD, normal MRI findings, and a subnormal clinical response to GH.

**Diagnosis of IGF Deficiency Syndrome: Growth Hormone Insensitivity**

The combination of decreased serum levels of IGF-I, IGF-II, and IGFBP-3 plus increased serum levels of GH suggests a diagnosis of GHI. [1253] The possibility of GHR deficiency is supported by a family history consistent with autosomal recessive transmission. Savage and associates [1254,1255] devised a scoring system for evaluating short children for the diagnosis of GHR deficiency, based on five parameters:

1. Basal serum GH higher than 10 µL (5 µg/L).
2. Serum IGF-I below 50 µg/L.
3. Height SDS below -3.
4. Serum GHBP less than 10%, based on binding of (125)GH.
5. A rise in serum IGF-I levels after GH administration of less than twofold the intra-assay variation (10%).

Blum and colleagues [1256] proposed that these criteria could be strengthened by:

1. Evaluating GH secretory profiles, rather than isolated basal levels.
2. Employing an age-dependent range and the 0.1 percentile as the cut-off level for evaluation of serum IGF-I concentrations.
3. Using highly sensitive IGF-I immunosassays and defining a failed GH response as the inability to increase serum IGF-I levels by at least 15 µg/L.
4. Measuring both basal and GH-stimulated IGFBP-3 levels.

These criteria fit well with the population of patients with GHR deficiency in Ecuador, but that is a homogeneous population with severe GHI. [1257] The applicability of these criteria elsewhere remains to be evaluated. An important biochemical marker is the response of IGF-I (and, possibly, IGFBP-3) to GH stimulation. Normal ranges and age-defined responses of serum IGF-I levels have not been established. [1258,1259]

Decreased serum levels of GHBP suggest the diagnosis of GHR deficiency, but some individuals with GHR deficiency have normal serum concentrations of GHBP. [1260,1261,1262] Such cases represent mutations in the dimerization site or in the intracellular domain of the receptor or abnormalities of postreceptor signal transduction mechanisms. On the other hand, polymorphisms of the GHR gene, without associated reductions in levels of IGF-I or IGFBP-3, should not be considered examples of GHI. At this point, definitive diagnosis of GHI requires (1) the classic phenotype, (2) decreased serum levels of IGF-I and IGFBP-3, and (3) identification of an abnormality of the GHR gene.

**Idiopathic Short Stature**

Many children and early adolescents are short (<3rd percentile), with slowed linear growth velocity (<25th percentile). They may have delayed skeletal maturation and an impaired or attenuated pubertal growth spurt, with or without a family history manifesting some or all of these clinical features, and have no chronic illnesses or apparent endocrinopathies. Such children usually have normal GH secretory dynamics, although provocative test results may be blunted under some circumstances; GH-dependent peptides are lower than expected on a chronologic, although usually not skeletal, age basis; treatment with exogenous GH usually augments linear growth.

Such children are usually considered to have variants of normal growth and achieve a final adult height within the range considered acceptable for the family. In most of these children, the cause of the slowed childhood growth and commonly delayed pubertal spurt has not been established. Because this is the largest group of short children, continuing efforts are under way to develop a rational categorization and the means of distinguishing between these children from children with an abnormality of the GH-IGF axis. Several groups of patients, including those with CDGM, genetic or familial short stature, and heterozygous abnormalities of the GHR, are described later. Additional causes of ISS will likely be identified at each level of the hypothalamic-pituitary-IGF axis. [1263]

---

**Constitutional Delay of Growth and Maturation**

The term constitutional delay [1264,1265,1266] describes children with a normal variant of maturational tempo; characterized by short stature but relatively normal growth rates during childhood; delayed puberty with a late and attenuated pubertal growth spurt; and attainment of normal adult height. Most children with constitutional delay begin to deviate from the normal growth curve during the early years of life, and they are at or slightly below the 5th percentile for height by age 2 years. [1267] During middle childhood years, height SDS may gradually drift lower, but this does not appear to affect adult height outcome. [1268] Final height, although usually within the normal population range, is often in the lower part of the parental height target zone, [1269,1270] with few patients exceeding that target height. The predicted final height, especially when the skeletal age is extremely delayed, is greater than that usually achieved. [1271,1272,1273]

The delayed growth spurt may adversely affect growth of the spine and mineralization of the vertebrae, which is not overcome when the pubertal growth acceleration finally occurs, thus limiting the final height. [1274] The osteopenia reported in men with a history of delayed puberty may be due to a profound alteration of normal pubertal bone mineral accretion or to a prepubertal and continuing deficit in bone mass that is an intrinsic part of CDGM. [1275,1276]

GH secretion may be decreased with transient partial GHD at the time of the delayed pubertal growth spurt, apparently the consequence of inadequate production of gonadal steroids. [1277,1278] Such children would be expected to have delayed skeletal ages, normal or slightly low serum IGF-I but usually normal IGFBP-3 levels for skeletal age, and normal GH provocative tests (if pretreated with gonadal steroids). Overnight GH secretion is generally normal in these children when control groups are carefully matched. [1279] By definition, children with pure CDGM should have bone ages sufficiently delayed to result in normal predicted adult heights (>163 cm in males and >150 cm in females) (Table 23-10), although the correlation between predicted and final height is imperfect and must be viewed with caution. [1280,1281] When CDGM occurs in the context of familial short stature (see later), however, children may experience both a delayed adolescent growth spurt and a short final height.

As stated earlier, some have attributed the diminished growth in the peripubertal period in CDGM to a transient GHD or to a “lazy” pituitary, a concept that is probably due to the delay in the expected growth spurt.

**TABLE 23-10 -- Criteria for Presumptive Diagnosis of Constitutional Delay of Growth and Maturation**

<table>
<thead>
<tr>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. No history of systemic illness</td>
</tr>
<tr>
<td>2. Normal nutrition</td>
</tr>
<tr>
<td>3. Normal physical examination, including body proportions</td>
</tr>
<tr>
<td>4. Normal thyroid and GH levels</td>
</tr>
<tr>
<td>5. Normal CBC, sedimentation rate, electrolytes, BUN</td>
</tr>
<tr>
<td>6. Height at or below the 3rd percentile, but with annual growth rate above the 5th percentile for age</td>
</tr>
<tr>
<td>7. Delayed puberty</td>
</tr>
<tr>
<td>a. Males: failure to achieve Tanner G2 stage by age 13.8 years or P2 by 15.6 years</td>
</tr>
<tr>
<td>b. Females: failure to achieve Tanner B2 stage by age 13.3 years</td>
</tr>
<tr>
<td>8. Delayed bone age</td>
</tr>
</tbody>
</table>
9. Normal predicted adult height

a. Males: >163 cm (64 inches)

b. Females: >150 cm (59 inches)

BUN, blood urea nitrogen; CBC, complete blood count; GH, growth hormone.

to the inadequacies of GH testing, especially to the failure to pretreat patients with a brief course of gonadal steroids. Low serum levels of IGF-I and IGFBP-3 or a poor GH response to provocative testing (after priming with gonadal steroids) should mandate an investigation for underlying pathology, such as intracranial tumors.

Genetic (Familial) Short Stature

The control of growth in childhood and the final height attained are polygenic in nature. For this reason, familial height affects an individual's growth, and evaluation of a specific growth pattern must be placed in the context of familial growth and stature. Formulas have been developed for determination of parental target height, and growth curves that relate a child's height to parental height are available. As a general rule, a child who is growing at a rate that is inconsistent with that of siblings or parents warrants further evaluation.

Furthermore, many organic diseases characterized by growth retardation are genetically transmitted. This list includes multiple causes, such as GHI due to mutations of the GHR gene, GH gene deletions, mutations of the Pprop1 or Pouf1 gene, pseudohypoparathyroidism, diabetes mellitus, and some forms of hypothyroidism. Inherited nonendocrine diseases characterized by short stature include osteochondrodysplasias (see earlier), dysmorphic syndromes associated with IUGR (see earlier), inborn errors of metabolism, renal disease, and thalassemia (see later). Identifying short stature as inherited thus does not, by itself, relieve the physician of responsibility for determining the underlying cause of growth failure.

Nonetheless, a constellation of clinical findings describes a normal variant referred to as genetic short stature (GSS) (or familial short stature) that differs from the syndrome of CDGM discussed earlier. In GSS, childhood growth is at or below the 5th percentile but the velocity is generally normal. The onset and progression of puberty are normal or even slightly early and more rapid than normal, so that skeletal age is concordant with chronologic age. Parental height is short (both parents are often below the 10th percentile), and pubertal maturation is normal. Final heights in patients with GSS are short and in the target zone for the family. The GHIGF system is normal, but exogenous GH therapy during middle childhood years may increase linear growth velocity substantially without disproportionate augmentation of skeletal maturation. Whether long-term GH treatment enhances final height outcome, however, is not clear.

Heterozygous Mutations of the Growth Hormone Receptor

The level of the GHR may be genetically determined, although modulated by such factors as nutritional status. GH production appears to be inversely related to GHR/GHBP levels. Accordingly, GHBP levels have been assessed in subjects with ISS. Serum levels of GHBP in 90% of children with ISS are lower than the normal mean, 20% being below the normal range, especially a subgroup with low IGF-I and higher mean 12-hour levels of GH. Such data raise the possibility that an abnormality of GHR content or structure might impair GH action.

The inverse relationship of GHBP levels to GH production is consistent with this hypothesis. In a small group of patients with growth failure, low levels of IGF-I, and poor response to exogenous GH, heterozygous GHR mutations were present in 28%. In contrast with the rarity of homozygous GHR mutations in GHR, heterozygosity is more common and may be a frequent cause of short stature. In heterozygotes, protein from the mutant allele may disrupt the normal dimerization that occurs when GH interacts with its receptor, leading to diminished GH action and growth impairment.

The IGF-I/IGFBP-3 generation test following 4 days of GH administration may reveal individual patients with findings of low basal and provoked peptides and modestly elevated GH levels that might represent partial GHI. Biochemical confirmation of insensitivity is mandatory in such cases.
TREATMENT OF GROWTH RETARDATION

When growth failure is the result of a chronic underlying disease, such as renal failure, CF, or malabsorption, therapy must be directed at treatment of the underlying condition. Although growth acceleration may occur in such children with GH or IGF-I therapy, complete catch-up requires correction of the primary medical problem. If treatment of the underlying condition involves glucocorticoids, growth failure may be profound and is unlikely to be correctable until steroids are reduced or discontinued.

Correction of growth failure associated with chronic hypothyroidism requires appropriate thyroid replacement. As discussed earlier, thyroid therapy causes dramatic catch-up growth but also markedly accelerates skeletal maturation, potentially limiting adult height. More gradual thyroid replacement or the use of gonadotropin inhibitors to delay puberty, or both, may be necessary to obtain maximal final height.

Treatment of Constitutional Delay

CDGM is a normal variant, with (by definition) potential for a normal (although delayed) pubertal maturation and a normal (albeit diminished for target zone) adult height. Most subjects can be managed by careful evaluation to rule out other causes of abnormal growth and delayed puberty combined with appropriate explanation and counseling. The skeletal age and Bayley-Pinneau table are often helpful in explaining the potential for normal growth to the patient and parents. A family history of constitutional delay is also a source of reassurance. On occasion, however, the stigmata of short stature and delayed maturation may be psychologically disabling for the preadolescent or teenager.

Some adolescents with delayed puberty have poor self-images and limited social involvement. In such patients and in some in whom pubertal delay is predicted on the basis of the overall clinical picture, there is a role for the judicious use of short-term gonadal steroids.

Two aspects of this syndrome are addressed by androgen treatment: short stature, especially in boys between ages 10 and 14, and delayed puberty after age 14 years. In the younger group, in whom CDGM is apparent, the orally administered, synthetic androgens, oxandrolone, has been used extensively. In several controlled studies, oxandrolone therapy for 3 months to 4 years increased linear growth velocity of 3 to 5 cm/year without adverse effects or decreasing either actual or predicted final height. The growth-promoting effects of oxandrolone appear related to its androgenic and anabolic effects rather than to augmentation of the GH-IGF axis. Currently recommended treatment is 0.1 mg/kg orally per day. In older boys, in whom delayed pubertal maturation is unbearable and anxiety-provoking, testosterone enanthate has been administered intramuscularly with success.

Criteria for therapy of such adolescents should include:

1. A minimal age of 14 years.
2. Height below the 3rd percentile.
3. Prepubertal or early Tanner G2 stage with a early morning serum testosterone lower than 3.5 nmol/L (<1 ng/mL).
4. A poor self-image that does not respond to reassurance alone.

Therapy consists of intramuscular testosterone enanthate, 50 to 200 mg every 3 to 4 weeks, for a total of four to six injections. Patients typically show early secondary sex characteristics by the fourth injection and grow an average of 10 cm in the ensuing year. Despite attempts to choose subjects carefully for treatment programs in CDGM, a spectrum of activation of the reproductive system is inevitable; growth responses to short courses of therapy are best in the boys who have early pubertal gonadotropin secretory patterns. Testosterone enhances growth velocity by direct actions and increases GH production. Brief testosterone regimens do not cause overly rapid skeletal maturation, compromise adult height, or suppress pubertal maturation. It is important to emphasize to the patient that he is normal, that therapy is short-term and designed to provide some pubertal development earlier than he would on his own, and that treatment will not increase adult height. In such situations, the combination of short-term androgen therapy, reassurance, and counseling helps the boys with constitutional delay to cope with a difficult adolescence.

The availability of several new forms of testosterone, which are approved for adults with hypogonadism, provides adolescents with an opportunity for a choice among different androgen replacement therapies. Although effectiveness of these preparations has not been demonstrated in children with constitutional delay, we have personal experience with their successful use, finding an equivalent response to that obtained with testosterone injections. Testosterone gel is painless and easy to apply and has proven popular since its release. Testosterone patches also avoid the need for injections, but they work best when applied to the scrotum and are often accompanied by complaints of itching. The dosing of these alternative forms of therapy in children and adolescents has not yet been established. In view of the important role of estrogen in the process of skeletal maturation, aromatase inhibitors might be used in conjunction with androgen therapy to prevent an acceleration of bone age and further enhance final adult height.

Patients must be reevaluated to ensure that they enter "true" puberty. One year after testosterone treatment, boys should have testicular enlargement and a serum testosterone in the pubertal range. If this is not the case, the diagnosis of hypothalamic-pituitary insufficiency or hypogonadotropic hypogonadism should be considered. Although the diagnosis of constitutional growth delay remains most likely in such patients, some eventually prove to be gonadotropin-deficient, especially if they are still prepubertal late in adolescence.

Referrals for constitutional delay are more common in boys than girls, undoubtedly reflecting our cultural values. When constitutional delay is a problem in girls, short-term estrogen therapy can be employed, but the advancement of bone is a greater hazard at doses that enhance growth velocity and sexual maturation. The use of GH in patients with constitutional delay is discussed in the following sections.
Treatment of Growth Hormone Deficiency

Nomenclature and Potency Estimation

The nomenclature for the various biosynthetic GH preparations reflects the source and the chemical composition of the product. Somatropin refers to GH of the same amino acid sequence as that in naturally occurring human GH. Somatropin from human pituitary glands is abbreviated GH or pit-GH; recombinant-origin somatropin is termed recombinant GH (rGH). Somatrem, abbreviated met-rGH, refers to the methionine derivative of rGH. Although the latter preparation is a more antigenic preparation, that propensity is not clinically relevant; despite the presence of anti-GH antibodies, growth responses to met-rGH are similar to those seen in patients treated with rGH.1209,1210 We refer to these biosynthetic preparations as GH in the subsequent discussions.

The biopotency of commercially available biosynthetic GH preparations, expressed as International Units per milligram of the new WHO rGH reference reagent 88/624 for somatropin, is 3 IU/mg.1229 It was necessary to standardize the early GH preparations by bioassay because of variable production techniques (such as extraction and column purification). The most common bioassays have been the hypophysectomized rat weight gain assay, the tibial width assay, and the more sensitive Nb2 rat lymphoma proliferation assay.1213,1214,1215 With the availability of purified and essentially equivalent rGH products, the requirement for bioassays has become an FDA requisite to substantiate biologic activity rather than to assess potential differences between preparations. It is likely that the bioassays will be replaced by in vitro binding assays using GHRs or GHBP derived from molecular techniques.1229

Historical Perspective

Because untreated patients with IGF deficiency syndrome have profound short stature (averaging nearly -5 SDS1216,1217,1218), the clinical urgency to use GH therapy as soon as it was available has been apparent.1219 The action of GH is highly species-specific, and humans do not respond to animal-derived GH.1220,1221,1222,1223 Unlike most other hormones, the only GH that is biologically active in humans is primate GH. For many years, human cadaver pituitary glands were the only practical source of primate GH for treatment of GHD, and more than 27,000 children with GHD worldwide were treated with pit-GH.1224 The limited supplies of pit-GH, low doses, and interrupted treatment regimens resulted in incomplete growth increments; usually therapy was discontinued in boys whose height reached 5 feet, 5 inches in height and in girls who reached 5 feet in height. Nonetheless, this treatment did increase linear growth and in many patients enhanced final adult height. The dose-response relationship and the relation of age to GH response were recognized during this period.1225

Distribution of pit-GH was halted in the United States and most of Europe in 1985 because of concern about a causal relationship with Creutzfeldt-Jakob disease (CJD), a rare and fatal spongiform encephalopathy that had been previously reported to be capable of iatrogenic transmission through human tissue.1226,1227 In North America and Europe, the incidence of this disorder is approximately 1 case per million in the general population; it is exceedingly rare before age 50 years. To date, more than 100 young adults who had received human cadaver pituitary products have been identified as cases of CJD, with the sad likelihood that all will die of the disease.1228,1229,1230 In patients in the United States, the onset of CJD was 14 to 33 years after starting treatment, whereas the large cohort of French patients had a median incubation period approximately 5 years shorter.1231,1232 Vigilant surveillance for this dreadful complication continues.

Fortunately, by the time the risks of pituitary-derived GH were discovered, rGH was being tested for safety and efficacy.1233,1234,1235 The original rGH included an N-terminal methionine, added as a start signal for transcription (met-rGH). This preparation mimicked pit-GH in regard to both anabolic and metabolic actions. Subsequent rGH preparations do not contain the additional methionine. rGH has universally replaced pit-GH as the treatment for children with GHD.

Treatment Regimens

The recommended therapy starting dose of GH in GHD is 0.18 to 0.35 mg/kg of body weight per week, administered in seven daily doses, with the mean dose in the United States being 0.3 mg/kg.1236 Alternative regimens include a 6-day/week or 3-day/week schedule, with the same weekly dosage, but they are not as successful. In general, the growth response to GH is a function of the log-dose given, so that increasing dosages further enhance growth velocities,1237,1238 but daily dosing may be the most important treatment parameter.1239 Either subcutaneous or intramuscular administration has equivalent growth-promoting activity;1240 the former is now used almost exclusively. GH is available in several vehicles, and multiple systems are now available for administering GH (Table 23-11).

The standard preparation is lyophilized GH that is highly water-soluble, so that it may be brought into solution with a small volume of diluent. An aqueous solution is ready to use and has 28-day stability. A sustained-release preparation of GH with protein integrity in a poly(lactide-coglycolide) polymer that is biocompatible and biodegradable permits once-monthly or twice-monthly treatments.1241 The pharmacokinetics of this system show an early release of GH and then a sustained release over 12 to 28 days. Average exposure to GH and IGF-I following administration of current treatment regimens of depot GH is approximately half that of daily GH therapy. Either reconstituted or liquid GH is administered in insulin syringes with ultralfine needles that are almost pain-free in skilled hands. Pen devices with internal reconstitution of the GH are frequently used because of ease, accuracy, and "hidden" needles. Needle-free, jet injector systems are available and do yield a normal serum immunoreactive and bioactive GH profile.1242 The sustained-release GH is administered through a short, larger-bore needle, but the injection pain is balanced against the low frequency of treatments. At this time, all of the GH preparations yield comparable short-term growth outcomes, except long-acting GH, in which mean first year growth rates are about 2 cm lower per year.1243,1244

GH treatment should be continued after growth ceases, because GH has other important metabolic effects, including support of normal gonadal function1245,1246 and attainment of normal adult bone mineral density.1247,1248 A report from the Drug and Therapeutics committee of the Lawson Wilkins Pediatric Endocrine Society summarized the society's views on the use of GH in children with diverse syndromes of short stature.1249

Growth responses to exogenous GH vary, depending on the frequency of administration, dosage, age (greater absolute gain in a younger child, though not necessarily of growth velocity SDS), weight, and GH amount, as assessed by serum GHBP levels.1250,1251,1252 On this general regimen, nonetheless, the typical

<table>
<thead>
<tr>
<th>Table 23-11 – New Modalities for Treatment of Growth Hormone Deficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liquid formulations</td>
</tr>
<tr>
<td>Pen-type delivery devices</td>
</tr>
<tr>
<td>Needle-less devices</td>
</tr>
<tr>
<td>Oral secretagogues</td>
</tr>
<tr>
<td>Long-acting GH formulations</td>
</tr>
<tr>
<td>GH-releasing hormone</td>
</tr>
<tr>
<td>Inhaled GH delivery systems</td>
</tr>
</tbody>
</table>

GH, growth hormone.
GHD child accelerates growth from a pretreatment rate of 3 to 4 cm/year to 10 to 12 cm/year in year 1 of therapy and 7 to 9 cm/year in years 2 and 3. Progressive waning of GH efficacy occurs and is poorly understood. The importance of dosage frequency is illustrated (Fig. 23-53 and Fig. 23-54) by data from a carefully done assessment of growth responses of prepubertal GHD children randomly assigned to receive thrice-weekly or daily GH at the same total weekly dose (0.30 mg/kg per week). The mean total height gain during this period was 9.7 cm greater in the daily-treated patients (38.4 versus 28.7 cm. P < 0.0002) with similar increments in skeletal maturation and no acceleration of the onset of puberty. Mean height SDS at the end of 4 years was +0.2, or at the midpoint of normal for age. At a dosage of 0.30 mg/kg per week, the approximate current cost of GH therapy for a 20-kg child is $12,000 to $15,000 annually.

Sophisticated mathematical models have examined many laboratory and auxologic parameters that influence response to GH therapy. Because age at onset of treatment is inversely correlated with growth responses and the smaller, lighter child requires less GH (with marked economic benefit), growth data in early treated children are important to assess. In short-term studies of 134 patients, treated prior to 3 years of age, marked early catch-up occurred with a mean height gain of around 3 SDS by 4 years of therapy, allowing most children to reach the normal height range by middle childhood. In one study, mean height reached -0.4 SDS after 8 years of treatment. Near adult height data in 13 patients treated before 5 years did not differ from the midparental target height (-0.9 versus -0.7 SDS). If future long-term outcome studies also show excellent growth responses with achievement of genetic target height and adherence to treatment regimen were possible in very young children, strongly worded recommendations for early treatment would surely be appropriate.

**Longer-Term Final Height Results**

Much information on growth has been reported about pGH-treated children, generally thrice-weekly and administered intramuscularly. Five-year data are available from Bundak and co-workers on a group of 58 prepubertal and 20 pubertal children with GHD. The younger group increased its height from -3.6 SDS to -2 SDS, whereas the pubertal children grew to -2.3 SDS. The height SDS for bone age, however, did not increase; thus, further loss of adult height was prevented, but there was no increase in the adult height prediction. The importance of early initiation of treatment is stressed by such data. Similarly, Libber and associates found that GH therapy increased mean height from about -4.2 SDS to -2.3 SDS. Patients treated with daily (QD) and thrice-weekly injections (TIW). The mean annual growth velocity in the QD group was significantly greater during each year, although significance diminished from year 1 to year 4. (From MacGillivray MH, Baptista J, Johanson A, and Genentech Study Group. Outcome of a four-year randomized study of daily versus three times weekly somatropin treatment in prepubertal growth hormone-deficient children. J Clin Endocrinol Metab 1996; 81:18061809; reproduced by permission of M. H. MacGillivray.)

![Figure 23-54](image)

Figure 23-54: Height standard deviation score (SDS) (mean + SD) for prepubertal patients with growth hormone deficiency (GHD) before and during 4 years of growth hormone (GH) treatment, contrasting results with daily (QD) and thrice-weekly (TIW) injections. The mean SDS in the QD group was significantly greater throughout the treatment period. Younger patients had the greatest increase in height SDS, and the effect of age was more marked in the QD group. (From MacGillivray MH, Baptista J, Johanson A, and Genentech Study Group. Outcome of a four-year randomized study of daily versus three times weekly somatropin treatment in prepubertal growth hormone-deficient children. J Clin Endocrinol Metab 1996; 81:18061809 (1328). Reproduced by permission of M. H. MacGillivray.)

Mean height reached -0.7 SDS, which was equivalent to the midparental target height. Even in these closely followed patients, however, a -0.4 to -0.6 SDS difference from midparental targeted height still occurred. The achievement of the genetic target is possible, however, as a Swedish subgroup (in KIGS) of consistently treated patients reached a median final height SDS of -0.32, which was equivalent to the midparental target height. By multiple regression analysis, factors found to correlate with enhanced adult height were baseline height, younger age at onset of treatment, longer treatment duration, and a greater growth velocity during the first year of treatment. Although the development of rGH

### TABLE 23-12: Adult Height in Children with Growth Hormone Deficiency Treated with Biosynthetic Growth Hormone (GH)

<table>
<thead>
<tr>
<th>Study</th>
<th>Sex</th>
<th>No. of Children</th>
<th>Dose</th>
<th>Duration (yr)</th>
<th>Age (yr)</th>
<th>Height SDS</th>
<th>Height SDS</th>
<th>Height vs. MPH</th>
</tr>
</thead>
<tbody>
<tr>
<td>KIGS</td>
<td>M</td>
<td>154</td>
<td>0.16</td>
<td>8.3</td>
<td>16.9</td>
<td>-0.9</td>
<td>+1.8</td>
<td>-0.3</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>115</td>
<td>0.18</td>
<td>8.0</td>
<td>17.1</td>
<td>-1.2</td>
<td>+1.6</td>
<td>-0.7</td>
</tr>
<tr>
<td>NCGS</td>
<td>M</td>
<td>2095</td>
<td>0.28</td>
<td>5.2</td>
<td>18.2</td>
<td>-1.1</td>
<td>+1.4</td>
<td>-0.7</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>1116</td>
<td>0.29</td>
<td>5.0</td>
<td>16.7</td>
<td>-1.3</td>
<td>+1.6</td>
<td>-0.9</td>
</tr>
</tbody>
</table>

KIGS, Pharmacia International Growth Database; MPH, midparental target height; NCGS, Genentech National Cooperative Growth Study.

*GH dose is mg/kg per week.*

Patients treated largely with biosynthetic GH have improved actual or near-final adult height SDS, with average final height in more than 1400 patients approximating -1.3 SD below the mean. Even in these closely followed patients, however, a -0.4 to -0.6 SDS difference from midparental targeted height still occurred. The achievement of the genetic target is possible, however, as a Swedish subgroup (in KIGS) of consistently treated patients reached a median final height SDS of -0.32, which was equivalent to the midparental target height.
has solved the problem of supply experienced in the pituitary GH era, delays in diagnosis and initiation of therapy have still compromised adult height.

In an effort to increase final height of patients with GHD, the use of high-dose GH during puberty has been studied based on the rationale that GH secretion normally rises two-fold to four-fold during the pubertal growth spurt with dramatic concomitant increases in serum IGF-I levels and that the pubertal growth spurt normally accounts for approximately 17% of adult male height and 12% of adult female height. Earlier studies by Stanhope and colleagues indicated that little difference in height gain could be observed when adolescent patients were given 30 versus 15 IU/m² of GH weekly (0.04 versus 0.02 mg/kg per day). Mauers and colleagues, however, evaluated higher pubertal GH doses (0.1 versus 0.043 mg/kg per day) and found that the higher dosage resulted in a 4.6 cm increase in near-final height. Mean height SDS achieved in the 0.043 mg/kg per day group (as in the earlier report) was 0.7 ± 0.9 but 0.0 ± 1.2 in the group receiving 0.1 mg/kg per day. The higher GH dosage did not result in more rapid acceleration of skeletal maturation.

Important factors in these improved adult height outcomes include:

1. The use of higher doses of GH.
2. The ability to treat until growth cessation.
3. Early initiation of treatment, progressive weight-related dose increments.
4. Attention to compliance with daily administration.
5. Appropriate thyroid hormone and glucocorticoid replacement therapy.

Since final height correlates with height at the onset of puberty in patients with GHD, every effort must be made to enhance growth velocity during prepuberty.

In data from NCGS and KIGS, the height gained during puberty in patients with GHD was generally comparable to that in healthy children with delayed bone ages. The pubertal height gain is negatively correlated with the age of pubertal onset. When normal or precocious puberty limits the response to GH, it may be appropriate to delay puberty by the use of a GnRH analogue. Use of this strategy in pubertal GHD patient groups, however suggestive, is not yet clearly documented to enhance final height. Nevertheless, the earlier the age of pubertal onset, the lower the final height outcome, and patients with GHD with delayed puberty or hypogonadotrophic hypogonadism have a taller adult height.

Growth Hormone Treatment of Prader-Willi Syndrome

GH treatment of growth failure in Prader-Willi syndrome is now an FDA-approved indication, with a recommended dose of approximately 1 mg/m² per day or 0.24 mg/kg per week. Numerous clinical trials document the efficacy of GH and help to confirm that the characteristic IGF deficiency state is due to GHD. After 5 years of GH therapy, Lindgren and co-workers (at a dose of 0.23 mg/kg per week) showed that mean height SDS approached 0.5, with a gain of nearly 2 SDS. Other shorter-term (6- to 24-month) treatment programs have found increased growth rates but also provide important information on the changes in abnormal metabolic parameters during GH treatment. Such data demonstrate reduction in body fat mass and percentage, increased fat-free mass, improved muscle strength and agility, and increased fat oxidation. Respiratory muscle weakness, which is found in Prader-Willi syndrome, was improved after GH treatment. In view of the association of insulin resistance and type 2 diabetes mellitus with obesity, glycemic status should be monitored in GH-treated patients with Prader-Willi syndrome.

Combined Pituitary Deficiencies

If GHD is part of a multiple pituitary insufficiency, it is necessary to address each endocrine deficiency both for general medical reasons and to ensure maximal effect of GH therapy. Thyrotropin deficiency is often "unmasked" during the initial phase of therapy, and thyroid function should be assessed both before the onset of therapy and during the first 3 months of GH treatment and at least on an annual basis thereafter.

Monitoring Growth Hormone Therapy

Although most pediatric endocrinologists simply document changes in growth velocity as the signal parameter of therapeutic efficacy, this may not be sufficient. Treatment models that predict growth rate with quite narrow confidence limits provide quantitative estimates of whether the individual patient is responding appropriately to GH. A model explaining 61% of growth response variability for the first year of therapy includes inverse relationships with maximum GH response during provocative testing, age and height SDS minus midparental height SDS, and positive correlation with body weight SDS, GH dose, and birth weight SDS. The most important predictive factor for years 2 through 4 is the first-year height velocity. Clearly, after age at diagnosis, GH dose management is the variable most affected by the physician. Changes in levels of the GH-dependent peptides, IGF-I and IGFBP-3, ALS, and the aggregate ternary complex, as well as leptin, correlate with growth responses. Measurement of these may give added information on the growth-promoting and fat-mobilizing actions of GH as well as of the spectrum of childhood responsiveness to exogenous GH. Safety monitoring should include yearly assessment of IGF-I, IGFBP-3, and fasting glucose/insulin ratios.

Poor Growth Responses

The growth response to GH typically attenuates after several years but should continue to be equal to or greater than

<table>
<thead>
<tr>
<th>TABLE 23-13 – Elements of Monitoring Growth Hormone Therapy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Close follow-up with a pediatric endocrinologist every 34 months</td>
</tr>
<tr>
<td>Determination of growth response (change in height Z-score)</td>
</tr>
<tr>
<td>Annual measurement of serum IGF-I and IGFBP-3 levels and fasting glucose/insulin ratio</td>
</tr>
<tr>
<td>Screening for potential adverse effects</td>
</tr>
<tr>
<td>Evaluation of compliance</td>
</tr>
<tr>
<td>Consideration of dose adjustment based on IGF values, growth response, and comparison to growth prediction models</td>
</tr>
<tr>
<td>IGF-I, insulin-like growth factor I; IGFBP-3, IGF-binding protein 3.</td>
</tr>
</tbody>
</table>

the normal height velocity for age throughout treatment. Using the statistical growth treatment models may prove valuable in judging therapeutic efficacy.
A suboptimal response to GH can be due to several causes:

1. Poor compliance.
2. Improper preparation of GH for administration or incorrect injection techniques.
3. Subclinical hypothyroidism.
5. Excessive glucocorticoid therapy.
7. Epiphyseal fusion.

Although 10% to 20% of recipients of GH develop anti-GH antibodies, growth failure is rarely due to such antibodies. Maximal growth response to GH can usually be obtained by early diagnosis and initiation of therapy and by careful attention to compliance and psychological support. Although in many earlier studies, large numbers of boys and especially of girls with idiopathic GHD did not achieve normal adult heights, it is our belief that normal height (i.e., reaching the family-specific target height) can be reached in most cases. The referral of short girls and their ultimate treatment with GH remain less frequent than those of boys.

Despite the efficacy of GH in accelerating growth in GHD children and to bring adult height into the normal range if treatment is begun sufficiently early, several studies have indicated that the long-term prognosis for such patients is guarded, although there is much more follow-up assessment needed in the modern era. The educational, vocational, and social outlook for adults who had childhood GHD is frequently suboptimal. Whether this reflects subtle intellectual deficits or the consequences of lower expectations of patients, families or teachers remains to be determined. In any case, patients with GHD should be followed carefully throughout life.

**Growth Hormone Treatment during the Transition to Adulthood and in Adulthood**

Studies have focused on the clinical consequences of GHD in adults and at the potential benefits of GH therapy in such patients. Signs and symptoms of adult GHD include reduced lean body mass and muscle mass, increased body fat, reduced bone mineral density, reduced exercise performance, and increased plasma cholesterol. GHD adults have been found to have impaired psychological well-being and quality of life, characterized by depression, anxiety, reduced energy and vitality, and social isolation. Several placebo-controlled studies have demonstrated that GH therapy of adult GHD patients results in marked alterations in body composition, fat distribution, bone density, and sense of well-being. Whether these effects of GH therapy will be sustained and, if so, what the optimal GH regimen will be remain to be determined.

On the basis of the data in adults, which suggest profound metabolic derangements associated with untreated GHD, continuation of GH treatment in late adolescent patients who show GHD on retesting has become a clinical issue. In nearly 500 patients with isolated GHD, 207 (44%) had normal GH levels during provocative retesting. In contrast, approximately 96% of patients with CPHD, with or without structural abnormalities of the hypothalamic-pituitary area, had sustained

The presence of multiple anterior pituitary hormone deficiencies or structural disease would seem to obviate the need for subsequent retesting.

One unresolved matter is the strict pharmacologic definition of GHD in that age group: There are no convincing normative data, and it does not seem reasonable to use the much lower GH responses to testing that characterize older adults. This remains to be worked out. Many patients do not wish to continue daily injections, but data support their necessity. Loss of energy and strength is frequent. Total body and abdominal fat in untreated patients increases significantly, whereas lean body mass is lost relative to controls or comparable GH-treated patients or those after reinstatement of therapy. Similarly, resting metabolic rates decreased within 6 months and were predictive of increased fat mass at 1 year after stopping GH treatment. Because bone mass accrual is not completed until the third decade, late adolescence is an important time for GH sufficiency to prevent later osteopenia.

In young adults (mean age, 24 years) in whom GH therapy had been discontinued since late adolescence, 57% had average spine bone mineral density more than -1 SD below the mean. Two years of GH treatment resulted in a 5.2% increase after a transient decrement at 6 months of therapy. After 1 to 2 years off GH therapy, IGFI-1 levels decrease to about 35% and IGFBP-3 to 75% of treated values. Resumption of GH normalizes these levels. These studies affirm the necessity of continuing GH treatment in late adolescence, albeit at non-childhood growth-promoting doses to prevent development of adverse cardiovascular risk, diminished bone mineralization, and an overall lowering of energy level.

Our current method of assessing the late adolescent patient is based on the retesting data discussed earlier and on the recommendations of the Growth Hormone Research Society. An algorithm to guide this transition is shown in Figure 23-55. On completion of skeletal growth, GH therapy should be halted for approximately 3 months and the patient is then retested. When an insulin tolerance test is mandatory for the patient to qualify for further GH therapy, this test should be performed; other pituitary hormones and serum IGFI-1 and IGFBP-3 levels are also measured. The opportunity should be taken to assess body composition, bone mineral density, fasting lipids, insulin, and quality of life, before and after discontinuation of GH therapy. These studies establish baseline data prior to long-term GH treatment and are also used for longitudinal assessment of untreated patients. The likelihood of GHD persisting into adult life in patients with CPHD, structural abnormalities of the hypothalamic-pituitary, or documented hypothalamic-pituitary molecular defects far exceeds that in patients with idiopathic, isolated GHD.

In summary, a conservative approach would suggest that all children classified as having GHD should be retested by insulin-provocative tests on completion of skeletal growth and before a commitment is made for long-term adult treatment. An argument can be made, however, that patients with a high likelihood of persistent GHD do not necessarily require retesting or, at most, should undergo determinations of IGFI-1 and IGFBP-3 concentrations. On the other hand, the child who has carried a diagnosis of idiopathic, isolated GHD should always be retested. When the diagnosis of adult GHD is established, continuation of GH therapy is strongly recommended. One should exercise caution when considering whether to continue GH therapy when there is a known risk of diabetes mellitus or malignancy. The tr guesses to close collaboration between the pediatric and adult endocrinologists, who should discuss the reintroduction of treatment with the patient.
survey of 251 pediatric endocrinologists suggested that 30% to 50% would consider treating short children with a wide array of diagnoses, including Russell-Silver and Noonan's syndromes, IUGR, and steroid-induced growth suppression. Indeed, children with growth failure due to multiple different disorders have received GH and have grown substantially.

Although these data arise from large, uncontrolled NCGBS and KGBS studies, certain trends do emerge. On purely auxologic grounds, it is difficult to discriminate between responses, over at least 4 years, to GH in children with a diagnosis of GHD by current clinical and laboratory criteria or in children with Turner's syndrome from responses in many children with "other conditions." If such data are taken in the context that responsiveness to GH (rather than an arbitrary diagnosis of GHD) should determine appropriateness of GH therapy, prospective evaluations in many clinical states of poor growth should ensue.

The questions of whether such therapy is safe and justifies the cost and potential risks are more complicated. Additionally, questions have been raised about the appropriateness of "cosmetic" hormonal therapy. These are not issues that are presently answered. Guidelines have been developed by the Lawson Wilkins Pediatric Endocrine Society and the Growth Hormone Research Society to provide a rational framework in this area.

The 2001 FDA-approved indications for GH treatment in childhood and adolescence are GHD, Prader-Willi syndrome, chronic renal failure, Turner's syndrome, and growth failure due to prior IUGR. GH insufficiency need not be documented in the latter four conditions of poor growth.

### Chronic Renal Failure

GH accelerates growth in children with chronic renal failure at least over several years of therapy (Fig. 23-56 and Fig. 23-57). Using a GH dosage of 0.05 mg/kg per day.

![Figure 23-56 Annual growth velocity in 20 growth-retarded prepubertal patients with chronic renal insufficiency who were treated with growth hormone. (From Fine RN, Kohaut E, Brown D, et al. Long-term treatment of growth-retarded children with chronic renal insufficiency with recombinant human growth hormone. Kidney Int 1996; 49:781785; reproduced by permission of R. N. Fine.)](image1)

![Figure 23-57 Height standard deviation score (SDS) (mean ± SD) in 20 growth-retarded prepubertal patients with chronic renal insufficiency. Note that the basal height is outside the normal range (at -2.5 SDS); enters the normal range within 1 year of treatment, and does not differ from the mean by the 5th year of growth hormone therapy. (From Fine RN, Kohaut E, Brown D, et al. Long-term treatment of growth-retarded children with chronic renal insufficiency with recombinant human growth hormone. Kidney Int 1996; 49:781785; reproduced by permission of R. N. Fine.)](image2)

Fine and associates reported a mean first-year growth rate of 10.7 cm in GH recipients and 6.5 cm in the placebo group; in the second year GH-treated patients had a mean growth rate of 7.8 cm/year versus 5.5 cm/year in placebo recipients; this resulted in an improvement of height SDS from -2.9 to -1.5. Twenty patients who were treated for 5 years reached a normal height SDS of -0.7, having had a mean height increase of 40 cm. The youngest patients (<2.5 years of age) had the most impressive growth response to GH therapy (14.1 cm/year). Deleterious effects on renal function or progression of osteodystrophy were not observed.

This treatment regimen does not adversely affect renal graft function after transplantation, nor is there significant "catch-down" growth following the transplantation.

The final height in 38 German children treated with GH for an average of 5.3 years was 1.6 ± 1.2 SDS, an increase of 1.4 SDS over the pretreatment baseline. The final height of an untreated control group was 2.1 ± 1.2 or 0.6 SDS below baseline.

Long-term GH treatment has also been shown to be safe and effective for extremely short (<4.0 SDS) children with nephropathic cystinosis and should be considered if nutrition and cysteamine treatment do not prevent growth failure. Because children are often short at the time of renal transplantation, hormone findings of relative GHI, and receive chronic prednisone therapy, GH is sometimes administered in the post-transplantation period. Data for 1 to 2 years of treatment of such children and adolescents indicate a large increment of growth velocity at year 1 and a smaller benefit at year 2. As with GH treatment of chronic renal failure, the pharmacologic regimen overcomes the relative GHI. Considerable assessment must yet be undertaken to demonstrate whether there is increased height growth over a longer term, that renal function does not deteriorate during therapy, and that the risk of rejection is not enhanced. Initial data suggest that GH treatment does not cause an accelerated decline of allograft function or changes in histopathologic findings but that the exacerbation of chronic rejection by GH therapy remains a possibility. Use of nonsteroid-based immunosuppressive regimens may obviate the need for post-transplantation GH treatment.

### Turner's Syndrome

Patients with Turner's syndrome have a final height of about 143 cm in the United States, which is about 20 cm lower than the mean final height of normal women. Extensive trials of therapy have involved androgens, estrogens, and GH. Androgens have either no effect or cause modest gains in final adult height, despite frequent reports of short-term improvement in growth velocity. Although estrogen deficiency is probably not involved in the growth failure of Turner's syndrome, especially that in infancy and childhood years, low-dose therapy has been unsuccessfully used in an attempt to enhance growth. These studies consistently show modest, transient increments of growth velocity, invariably accompanied by advancement of skeletal maturation. Final height outcome is not improved and often is impaired. Planning for the needed estrogen replacement for ovarian failure must take into account its effects on growth outcome.

Prior to the availability of rGH, data were conflicting concerning the efficacy of pit-GH in this disorder, but the ability of GH to accelerate growth has now been
demonstrated in multiple reports. In 1983, a randomized, controlled North American study of GH (at a dose of 0.375 mg/kg per week), with or without added oxandrolone, was initiated, with mean age of onset of treatment approximately 9 years. Analysis of all 62 girls enrolled in the study at near final height shows a stature of 152.1 cm in the GH plus oxandrolone group (a gain of 10.3 cm, compared with Lyon height predictions), whereas girls receiving GH alone are at 150.4 cm (a gain of 8.4 cm) (Fig. 23-35). In another arm of this study, addition of estrogen to the GH regimen prior to age 15 years lowered the final height gain from 8.4 to 5.1 cm. In a reassessment of North American data in NCGS, early initiation of GH treatment was shown to allow estrogen administration at a physiologic age without loss of adult height.

Several other studies have used higher doses of GH, have shown even greater gains in adult height outcomes. Sas and colleagues, in a multicenter trial using a maximum GH dose of approximately 0.63 mg/kg per week for 4.8 estrogen-free GH treatment years, beginning at mean age 8.1 years, found a gain of 16 cm over the modified Lyon projection. In their group receiving a similar GH dose to the American studies, a height gain of 12.5 cm was achieved by age 16 with 4.8 estrogen-free GH treatment years starting at 7.9 years. Carel and colleagues, using 0.7 mg/kg per week in a group that received 5.1 estrogen-free GH treatment years beginning at 10.2 years, gained 10.6 cm over Lyon projections. Their conventional dose group (0.3 mg/kg per week) gained only 5.2 cm with 3.0 estrogen-free GH treatment years starting at 11 years. The substantial variations in the GH-induced growth increments in these studies are presumably related to GH dose, duration of estrogen-free GH treatment years, the age of initiation of GH and estrogen administration, as well as the population and parental adult heights. In these higher dose treatment studies, hyperinsulinism with presumed insulin resistance was evident, though reversible.

Some uncertainty still exists, however, because none of the studies were placebo controlled to adult height, and many studies have yielded much poorer height outcomes. Nonetheless, in light of historical data on natural growth in Turner’s syndrome, the available treatment results do provide convincing support for the belief that GH can both accelerate growth and increase adult height.

We recommend seeking the diagnosis vigorously in short girls of any age and initiating therapy at that young age (i.e., at diagnosis). We consider the addition of oxandrolone to the regimen in girls with a late diagnosis. We initiate estrogen administration at an appropriate physiologic age in girls who began GH therapy at a young age, but we delay estrogen administration as long as possible in girls who began GH therapy at a late age. The diminished areal bone density in Turner’s syndrome is enhanced by GH, but estrogen therapy is needed to normalize volumetric density (i.e., related not simply to size). Use of statistical prediction models for long-term growth in Turner’s syndrome may permit a more quantitative assessment of the individual therapeutic efficacy.

### Down’s Syndrome

The encouraging results of GH trials in Turner’s syndrome have led to studies of GH in Down’s syndrome. In several preliminary studies, GH accelerated growth in such patients, although ethical issues have been raised concerning the appropriateness of such therapy. In the uncontrolled NCGS experience, 23 children experienced a 1.3 SDS height gain over the first 4 years of GH therapy. No convincing data exist that GH improves neurologic or intellectual function in such patients. The increased risk of diabetes mellitus and leukemia in children with Down’s syndrome might be augmented by GH therapy.

### Intrauterine Growth Retardation

A number of studies employing GH have been performed in short children with IUGR but are hampered by the heterogeneity of this group of patients, whose poor growth may reflect maternal factors, chromosomal disorders, dysmorphic syndromes, toxins, and others. Indeed, the low levels of IGF-I and IGFBP-3 in infants with IUGR, apparently related to fetal malnutrition, do not seem to predict the degree of subsequent growth impairment. Short children with IUGR make up a substantially portion of growth-retarded patients seen in pediatric endocrine practices. Because the height of these children may be in the range seen in growth deficiency syndrome, therapeutic attempts certainly are appropriate if it can be assumed that the insulin resistance noted in these children does not become a clinical issue. Encouraging short-term and long-term responses have been obtained with GH treatment.

A randomized, placebo-controlled, double-blind, two-dose (0.13 and 0.4 mg/kg per week) study on 95 children of GH therapy was serially reported during 3 years of treatment and then after 1 year off GH. After this period, mean height was normal in the high-dose group (1.76 ± 0.17 SDS) and higher than in the low-dose group (2.46 ± 0.39 SDS); skeletal maturation in the two groups advanced 5.3 ± 0.4 and 4.6 ± 0.7 years, respectively, over the 4-year study, but remained 1 to 2 years below chronologic age.

Several other 2-year multicenter trials using doses ranging from 0.23 to 0.7 mg/kg per week also showed enhanced growth velocity (gaining 5 to 10 cm more than control groups), but bone age advanced 2.7 years. These data suggest that height growth may catch up in children with IUGR who are very short during middle childhood years. Perhaps an approximation to the growth velocity of their peers would have enough benefit to justify GH treatment, even if long-term data do not indicate an appreciable enhancement of final adult height. Subsequently, the questions regarding long-term efficacy of GH therapy on final height have been answered.

Sas and co-workers contrasted prepubertal SGA children receiving 0.23 (n = 23) to 0.47 mg/kg per week (n = 16) over a 5-year period. Mean height SDS increment of 3.3 ± 0.7 in the higher dose group exceeded that in the lower (2.4 ± 0.5). Despite a mean skeletal age increase of 7 years during the study, predicted adult height increased 9.1 ± 2.8 and 14.0 ± 5.5 cm in the two groups.

de Zegher and colleagues reported results from the longest GH treatment of SGA children. They used several different regimens, including continuous treatment with 0.23 to 0.45 mg/kg per week for 6 years and a discontinuous program of 0.23 to 0.7 mg/kg per week for 2 years followed by no therapy or one additional 2-year course at these varied doses.

The mean dose in the discontinuous treated children was 0.22 mg/kg per week over the 6 years, but they received 46% fewer injections. Mean height SDS gains were approximately 2.0 cm in these regimens, with mean parent adjusted height SDS at about -0.5. Bone age advanced 7.3 ± 0.1 years over the 6-year period and remained slightly delayed. At 8 years of therapy (personal communication from de Zegher F, 2000), the children are still growing and have now reached approximately 12 cm over predicted adult height. Onset of puberty was not accelerated in either of these studies. It appears that GH therapy for this group of patients, which accounts for about 20% of short children, may become a quantitatively important use as this treatment is now an FDA-approved indication. No evidence of meaningful insulin resistance was found in these study patients, in contrast to its high prevalence in children with IUGR who are growing well and potentially evolving into the adult syndrome.
Osteochondrodysplasias

GH therapy has been studied in several skeletal dysplasias. The largest published study in achondroplasia involved 40 children. During the first year of treatment, height velocity increased from 3.8 to 6.6 cm/year; in year 2, height velocity decreased to approximately 5 cm/year. A modest improvement was seen in the ratio of lower limb length to height. Although GH was well tolerated, atlantoaxial dislocation during GH therapy has been reported in one patient. In another study, normal growth velocity was achieved for up to 6 years in 35 subjects with a significant increment in height SDS for at least 4 years. In this study, vertebral growth was disproportionately greater than limb growth.

Bridges and Brook reported on the effects of GH therapy of 27 patients with hypochondroplasia; response was maximal during the first year of treatment, but substantial benefit was seen through 4 years of treatment in pubertal subjects. Much of the growth response represents an increase of spinal length; with leg-lengthening procedures, some patients may achieve adult height within the normal range. Experience with GH treatment is limited in other skeletal disorders, such as dyschondrosteosis, hereditary multiple exostoses, osteogenesis imperfecta, and Ellis-van Creveld syndrome.

Noonan's Syndrome

Most experience with GH therapy of short stature in Noonan's syndrome has been limited to small, uncontrolled studies in which few patients have reached final height. The clinical diagnosis of a dysmorphic syndrome potentially makes the treatment groups heterogeneous. Overall, treatment results for 3 to 4 years are similar to those attained in Turner's syndrome, with growth velocities improving from baseline rates of approximately 4 cm/year to 8, 7.7, and 6 cm/year over the first 4 years of therapy, gaining from about -3.5 to -2 SDS without inordinate advancement of bone age. Although initial anecdotal experience suggested progression of ventricular hypertrophic cardiomyopathy, such has not been found in larger carefully monitored studies.

Idiopathic Short Stature (or, Possibly Normal Short Children)

A heterogeneous group that includes children with constitutional growth delay, genetic short stature, possible heterozygous mutations of the GHR gene, and an array of conditions still called idiopathic short stature has been very difficult to characterize. Even though these children are placed in one category, this group includes some children who have normal provocative GH responses but are, nevertheless, relatively IGF-deficient, reflecting the inadequacies of GH testing. This group also includes children with undiagnosed syndromes, chronic illnesses, or endocrine disorders. For example, children with heterozygous mutations of the GHR have subnormal growth, diminished IGF and GHBP levels, and impaired responsiveness to exogenous GH administration. These children often experience stressful circumstances, although studies suggest variable relationships of behavioral problems to the short stature. Accordingly, although the specific causes of these disorders are often unknown, GH treatment has been used widely.

The failure to report levels of IGF-I, IGFBP-3, and GHBP in many studies and differing interpretations of endogenous GH secretion studies (according to factors such as assay variance and control group size), in addition to the heterogeneity of the patient groups, confound data assessment. Furthermore, published clinical trials have not contained long-term control groups and have reported variable growth responses. Most normal short children treated with GH experience growth acceleration (catch-up) that generally is sustained over the first several years of therapy, although attenuation of the response occurs as in all other instances of GH treatment. It appears that slower pretreatment growth velocity and higher weight/height ratio, factors suggesting GHD, and a lesser degree of bone age retardation are associated with better early growth responses.

Longer-term data are now available to begin to determine the impact of therapy on adult height. Nearly 3000 children were classified as having ISS in KIGS, with having reached final height in 1999. GH treatment (0.2 to 0.25 mg/kg per week) resulted in achievement of target height in familial (genetic) short stature (FSS) patients, although at a short stature (-1.7 SDS in males and -0.2 SDS in females) with a mean gain during therapy of 0.6 to 0.9 SDS. In non-FSS children, the mean final height was greater in males (-1.4 SDS than in females (-2.3 SDS), with mean gains 1.3 and 0.9 SDS, respectively. These latter heights are, nonetheless, distant from the midparental target heights that were near 0 SDS.

Hintz and associates assessed adult height in 80 North American children with ISS treated for up to 10 years at a GH dose of 0.3 mg/kg per week. Mean height SDS at conclusion was -1.4 with a gain of 1.3 SDS, similar to the broader KIGS experience. Although this study was not placebo-controlled, the data were compared to predicted and actual final heights of two groups of short children followed for similar periods. Treated boys achieved 9.2 cm and girls 5.7 cm greater gain in final height than untreated controls. Careful monitoring did not reveal any discernible metabolic side effects.

Similarly, unpublished data from Sweden showed a mean final height gain of 1.5 SDS, in contrast with 0.6 SDS in an untreated group of short children.

McCaughy and colleagues showed that GH treatment (0.34 mg/kg per week) of 8 prepubertal girls led to a mean height SDS gain of 1.28 after 6.2 years of treatment. This was a 7.6 cm greater gain than in a control group whose mean height SDS did not change during the study.

Taken together, these studies show that GH treatment of prepubertal children with ISS does increase growth velocity and final height, but that the range of results reflects the heterogeneity of the study populations. Girls may be less responsive than boys, and early treatment may enhance growth outcome.

In a collaborative study of 229 untreated children from nine European countries, contrasting data, which challenge the value of GH treatment in such patients, showed final heights of -1.5 SDS in boys and -1.6 SDS in girls. Price found generally similar results in reviewing six studies but stressed the inaccuracy of height predictions and the failure to attain target heights.

Concerns had been raised that GH treatment might accelerate pubertal onset and progression resulting in failure to improve height SDS for bone age, thereby offsetting the positive responses observed during early years of GH treatment of ISS. This has not been substantiated by the studies described earlier. Nonetheless, several small studies assessed the value of delaying puberty with a GnRH agonist in GH-treated children with ISS. Although two trials showed considerable gain (6 to 10 cm) between predicted and final height, others did not. At this time, the added cost along with the potential negative impact of stopping puberty in a child already shorter and less mature than his or her peers diminishes our enthusiasm for such a delaying regimen.

Important questions have been raised about the financial, ethical, and psychosocial impact of GH therapy of normal short children. Given the cost, the financial implications of treating normal short children (whether at the bottom 5%, 3%, 1%, or 0.1%) is considerable. The point is well taken that 5% of the population will always be below the 5th percentile, whether or not we treat with GH, and focusing on short stature potentially handicaps an otherwise normal child, psychologically or socially. No convincing data have been presented to date that GH treatment of normal short children improves psychological, social, or educational function. Furthermore, the final adult height in children with CDGM (probably the most frequent discrimination) will be adequate without any treatment.

Finally, the treatment risks of GH therapy, both known and unknown, must be considered when treatment of otherwise normal children is an issue. There can be no discernible risk associated with therapy in this situation. Yet, given (1) the current limitations in the definitive discrimination between GHD and normal short stature, (2) the inadequate understanding of neurosecretory defects of GH secretion, (3) the inadequate recognition of “partial” GHD or GHI, and (4) the need to move to a more global concept of IGF deficiency, it seems unfair to prevent GH therapy of short children who do not meet a narrow definition of GHI (i.e., provocative testing) that we recognize as inadequate.

As noted earlier, many of these children are comparable clinically and in responsivity to GH to classic GHD patients. Accordingly, we recommend the following approach:

1. Controlled therapeutic trials of normal short stature must continue to be carried out to adult height.
2. Appropriate evaluation should include extensive analysis of the GH-IGF axis (with GHBP levels, IGF responses to GH treatment, and possibly GHRP/GHRH testing) before labeling a short child as “normal.”
3. Proper assessment of pretreatment growth velocity should be over a minimum of a 6-month period and preferably for 12 months.
4. In the otherwise normal child with severe short stature (at least 2.5 SD below the mean for age) and a poor height velocity (e.g., <25th percentile for age)
cm/year before age 4 years; <5 cm/year at age 4 to 8 years; <4 cm/year anytime before puberty), the possibility of a trial of GH therapy should be discussed with the patient and family. This discussion should include an assessment of normal growth patterns, familial growth patterns, and the predicted pubertal and statural development. The inconveniences, discomforts, and potential risks of GH treatment should be fully described. It is the physician's responsibility to ensure that expectations of the child and the parents are realistic in regard to short-term growth and ultimate height. Where appropriate, counseling and psychological support should be provided.

5. If a trial of GH therapy is desired, treatment should be for a minimum of 6 months. There are no perfect guidelines for dosage; we recommend 0.05 mg/kg per day.

6. Therapy with GH should be continued beyond 6 months only if growth is accelerated (an increase in height velocity of at least 2 cm/year). Efficacy of treatment requires continuous monitoring, especially in patients with partial GHR deficiency, in whom the possibility of IGF-I therapy is a future alternative. The use of complex, nonlinear multivariate models to assess growth response is recommended.

7. Growth acceleration with GH treatment does not relieve the physician of child's growth failure. Appropriate studies should be repeated, when indicated.

8. Treatment must be carefully monitored for side effects of GH treatment.

9. Continued psychological support should be provided for the child and family. This includes guiding the patient through puberty and providing post-treatment follow-up.

The evaluation and treatment of a "normal short child" with GH should be conducted by physicians trained in the management of growth disorders.

Miscellaneous Causes of Growth Failure

In addition to the clinical conditions described earlier, GH has been employed to treat short stature associated with various conditions related to postnatal growth failure. In general, such trials have been uncontrolled and have not included sufficient numbers of subjects for efficacy to be evaluated. Continuing examination of such treatment should be noted in the large international databases.

Normal Aging and Other Catabolic States

Detailed consideration of the potential use of GH in normal aging is beyond the scope of this chapter. The rationale for such therapy is based on the concept of the "somatopause," referring to the fact that GH secretion normally declines progressively after 30 years of age, as reflected in decreasing IGF-I levels. Aging can be viewed as a catabolic state, with the potential that GH therapy must reverse or retard the loss of muscle mass and strength and the decrease in bone density with aging. Clinical studies are in progress.

Growth failure, often with impaired final adult height, is a characteristic clinical finding in endogenous or exogenous Cushing's syndrome. Excess glucocorticoids cause a catabolic state characterized by increased proteolysis, decreased protein synthesis, lowered osteoblastic and increased osteolytic activity, and insulin resistance. GH treatment blunts some of these catabolic actions but increases the insulin resistance.

Mauras and Beaufreere found that IGF-I therapy, similarly, induces an anabolic response despite excess glucocorticoids but does not cause insulin resistance. GH treatment in the post-transplantation period and in other glucocorticoid-treated children causes some height increments but not as good a response as in individuals not on glucocorticoids. GH does enhance bone formation and increases osteoblastic activity in such children. GH therapy is also being investigated in catabolic states, such as burns, tumor cachexia, major abdominal surgery, AIDS, sepsis, metabolic acidosis, and situations requiring total parenteral alimentation.
Side Effects of Growth Hormone

For a quarter of a century, pituitary-derived human GH had an enviable safety record but proved to be the agent for transmission of the fatal spongiform encephalopathy, CJD. Although pit-GH was removed from use in the United States in 1985 and later throughout the world, more than 100 patients with GH-derived CJD have been identified, and cases are likely to continue to be discovered over the next several decades. Although this risk does not exist with recombinant DNA-derived GH, the experience with pituitary GH serves as a grim reminder of the potential toxicity that can reside in "normal" products and "physiologic replacement."

Extensive experience with rGH over more than 20 years has been encouraging. Every attempt has been made to seek physiologic replacement rather than pharmacologic therapy, but this is often not possible. Concerns have been raised about a number of potential complications, which clearly require continued follow-up and assessment. This evaluation has been greatly facilitated by the extensive databases established by GH manufacturers, in particular Genentech (NCGS) and Pharmacia (KIGS).

Development of Leukemia

The development of leukemia as a complication of GH therapy was first reported in five cases from Japan in 1988, and to date more than 50 cases of leukemia have been reported in GH-treated cases. Many of the cases are from Japan, but some are from the United States. One difficulty in assessing the role of GH treatment in this disorder is that many children with GHD have conditions that may predispose the development of leukemia, such as a history of malignancy, radiation, or syndrome associated with the development of leukemia (Bloom's syndrome, Down's syndrome, Fanconis anemia). GH-treated patients who develop leukemia do so at a later age than the normal population. Patients have included recipients of both pit-GH and rGH, and leukemia has occurred both during treatment and following termination of therapy. Calculations of relative risk are imprecise but vary from sevenfold in Japan to twofold to fourfold in the United States. Leukemia has been reported in GHD individuals without any history of GH therapy, suggesting that the GH state alone might be a predisposing factor.

We cannot be certain whether GH is a causative agent in the development of leukemia. If it is, the increase in risk appears modest and may arise from the underlying state rather than from GH therapy. The number of cases of new leukemia worldwide in children treated with GH but with no known risk factors is approximately what would be expected on a patient-year basis. This issue should be discussed with all potential recipients of GH, but it appears that the risk is limited to children with several risk factors. Particular care should be used in prescribing GH therapy for children with a history of leukemia or lymphoma or other disorder that conveys an increased risk of leukemia.

In a study of more than 600 children with prior leukemia and who received GH, the relapse rate was within the expected range, consistent with no effect of GH replacement therapy on recurrence of leukemia. In addition, data from 59,158 GH-treated patients with more than 193,000 patient-years at risk did not reveal an increased risk of nonleukemic extracranial neoplasms.

Recurrence of Central Nervous System Tumors

As many recipients of GH therapy have acquired GHD because of CNS tumors or their treatment, the possibility of tumor recurrence with therapy is of obvious importance. Estimates of CNS tumor recurrence rates in non-GH-treated children and adolescents are difficult to obtain, bearing in mind the vast array of treatment programs used in the past three decades.

Of a total of 1363 patients, documented in 11 reports, not treated with GH, 209 (19.3%) had recurrences. Such data in a heterogeneous group, including craniopharyngiomas, gliomas, ependymomas, medulloblastomas, and germ cell tumors, provide a background for assessing recurrence rates in GH-treated youth. Reports from nine centers, encompassing 390 patients, indicate recurrence in 64 (16.4%) at the time of publication, not much different from the much larger number of untreated patients.

In a particularly well-conducted comparative study at three pediatric neuro-oncology centers with data on 1071 patients with a brain tumor (180 received GH for a mean period of 6.4 years, and 31 were observed for more than 10 years), the relative risk of recurrence or death was similar in both groups. Extensive analysis of 4410 patients with brain tumor or craniopharyngioma histories prior to GH therapy in the NCGS and KIGS experiences showed a similar lack of increased tumor recurrence. In the NCGS series, recurrence rates of the most common CNS neoplasms, craniopharyngioma (6.4%), primary neuroectodermal tumors (medulloblastoma, ependymoma) (7.2%), and low-grade glioma (18.1%) were lower or similar to those reported in children not given GH. Despite these comforting data, the relatively short median follow-up times, even in the huge international databases, must temper the willingness to eliminate any relationship of GH therapy to recurrence of intracranial neoplasia.

Pseudotumor Cerebri

Pseudotumor cerebri (idiopathic intracranial hypertension) has been reported in GH-treated patients. The disorder may develop within months of starting treatment or as long as 5 years into the course; it appears to be more frequent in patients with renal failure than in those with GHD. The mechanism for the effect is unclear but may reflect changes in fluid dynamics within the CNS. Pseudotumor has also been described following thyroid hormone replacement in hypothyroidism.

In any case, physicians should be alert to complaints of headache, nausea, dizziness, ataxia, or visual changes. Significant fluid retention with edema or hypertension is rare. Because of the possible association of pseudopapilledema with GHD, perhaps representing a variant of optic nerve hypoplasia, careful ophthalmologic evaluation should be undertaken in patients with suspected GH therapy-associated pseudotumor cerebri to avoid overdiagnosis and invasive treatments.

Slipped Capital Femoral Epiphysis

Slipped capital femoral epiphysis (SCFE) is associated with both hypothyroidism and GHD. Whether GH therapy plays a role in this condition has been difficult to determine, in part, because the incidence of SCFE varies with age, sex, race, and geographic locale. It has been reported in 2 to 142 cases per 100,000. The data in the KIGS and NCGS studies are in this range. Accordingly, although SCFE cannot be attributed to GH therapy per se, complaints of hip and knee pain or limp should be evaluated carefully.

Diabetes Mellitus

The association of GH treatment with insulin resistance is well documented. A retrospective analysis of the KIGS experience found 43 of 23,333 children with abnormalities of glucose regulation, including 11 with type 1 and 18 with type 2 diabetes mellitus. The heterogeneity of this patient group and the failure to corroborate these findings with a similar retrospective analysis of NCGS data (type 2, 6.2/100,000) put the report into question. Nonetheless, the reduction of insulin sensitivity by GH is a concern that demands close assessment of high-risk patients, such as those with Prader-Willi or Turner's syndromes and a history of IUGR. It seems most likely that the relationship of the development of diabetes mellitus in childhood and adolescent GH recipients is due to a common genetic linkage rather than to a GH side effect.
Miscellaneous Side Effects

Other potential side effects of GH therapy include prepubertal gynecomastia, pancreatitis, growth changes, scoliosis and kyphosis, worsening of neurofibromatosis, hypertrophy of tonsils and adenoids, and sleep apnea. One report of reduced testicular volume and elevated gonadotropin levels in four young adult men previously treated with GH for ISS had not been confirmed by a double-blind, placebo-controlled study or in the international databases. This list, obviously, is only partial. It is best for the clinician to remember that GH and the peptide growth factors it regulates are potent mitogens with diverse metabolic and anabolic actions. All patients receiving GH treatment, even as replacement therapy, must be carefully monitored for side effects.

For the most part, side effects from GH therapy are minimal and rare. When they occur, a careful history and physical examination are adequate to identify their presence. Management of these side effects may include either transient reduction of dosage or temporary discontinuation of GH. In the absence of other risk factors, there is no evidence that the risk of leukemia, brain tumor recurrence, SCFE, or diabetes mellitus are increased in recipients of long-term GH treatment. Any patient receiving GH who has a second major medical condition, such as being a tumor survivor, should be observed in conjunction with an appropriate specialist, such as an oncologist and a neurosurgeon. Although GH does increase the mortality of critically ill patients in intensive care units, there is no evidence that GH replacement therapy needs to be discontinued during intercurrent illness in children with GHD.

The Question of Long-Term Cancer Risk

Several epidemiologic studies suggest an association between high serum IGF-I levels and incidence of malignancy. The calculated risk of cancer in those studies was also increased for patients with low IGFBP-3 levels. Although additional studies are being conducted to verify or disprove these associations, the role of GH should also be carefully examined.

IGF-I levels have not been statistically associated with cancer risk, but the combination of high IGF-I and low IGFBP-3 appears to be related to a heightened risk. Because GH positively influences production of both peptides, this casts doubt on its role as a driving force in the IGF-cancer relationship. Epidemiologic studies assessing the risk of malignancy in patients with acromegaly found differing results, with some, but not others, identifying significant associations between acromegaly and colon cancer risk. Small size, uncontrolled retrospective nature, and multiple possible sources of bias make these reports difficult to interpret.

The largest study, a review of more than 1000 patients, indicated no overall increased cancer incidence in acromegaly. Although colon cancer risk was also not increased in that study, mortality was higher in this population, suggesting an effect of GH or IGF-I on established tumors.

A recent prospective analysis of colon and colonic polyps in patients with acromegaly did not observe an association between these two diseases when using either autopsy series or prospective colonoscopy screening series for the control population. Acromegaly is associated with a marked increase in incidence of benign hyperplasia of several organs, including colonic polyps. Such findings suggest that the GH-IGF axis may lead to symptomatic benign proliferative disease, which might be associated with symptoms such as rectal bleeding that would then lead to a potential detection (or ascertainment) bias.

Children receiving GH do not have a greater risk of de novo or recurrent tumors. No increased incidence of cancer was found in GH recipients among adults who were treated for GHD. These reports represent imperfect, uncontrolled studies, but the experience gained through them strongly suggests that GH therapy is not associated with future development of neoplasms in the absence of other risk factors.

The use of IGF-I and IGFBP-3 in the monitoring of GH recipients, both adult and pediatric, has been recommended and endorsed by international bodies such as the Growth Hormone Research Society. Until the issue of cancer risk in GH therapy is fully resolved, the prudent approach appears to be regular monitoring of both IGF-I and IGFBP-3 and altering the GH dose to ensure that the theoretical risk profile induced by GH therapy is favorable. This can be done by avoiding the unlikely situation of a GH-treated patient having an IGF-I level at the upper end and IGFBP-3 at the lower end of the normal ranges.

In the 21st century, many GH-deficient patients will be receiving lifelong GH replacement, emphasizing importance of long-term, regular monitoring of IGF-I and IGFBP-3. Although many issues still remain in fully discerning the relationship, if any, between the GH/IGF axis and the risk of cancer, current data strongly suggest the safety of present indications for use of GH in children and adults.
Treatment of Growth Hormone Insensitivity Syndrome

The production of IGF peptides by recombinant DNA technology has permitted clinical trials of IGF therapy. IGF-I administration to normal adult male volunteers as a single intravenous injection of 100 µg/kg caused hypoglycemia within 15 minutes. On a molar basis, IGF-I has approximately 6% of the hypoglycemic potency of insulin, presumably reflecting increases in "free" IGF-I. In contrast, intravenous infusions of IGF-I to normal men at a rate of 20 µg/kg/hour resulted in serum IGF-I levels within the normal range and did not produce hypoglycemia but did suppress GH levels, increase creatinine clearance, and decrease plasma urea nitrogen. 

The most obvious clinical use of IGF-I therapy is in patients with GHI. In patients with GHR deficiency, intravenous bolus administration of IGF-I caused acute symptomatic hypoglycemia, presumably because of low serum levels of IGFBP-3. The subcutaneous administration of IGF-I to eight GHR deficiency patients, at a dosage of 150 µg/kg per day for 7 days, did not cause symptomatic hypoglycemia. Vaccarello and co-workers treated six adults with GHR deficiency for 7 days with subcutaneous IGF-I at a dosage of 40 µg/kg every 12 hours. Normal serum IGF-I levels were maintained for 2 to 6 hours after injection, followed by a rapid decline, because of low serum levels of IGFBP-3. Hypoglycemia did not occur, mean 24-hour GH levels were suppressed, and urinary calcium was increased.

A number of short-term growth-related studies with IGF-I treatment at varied doses have been reported. Laron and associates reported growth acceleration to rates of 8.8 to 13.6 cm/year in five children treated with a single daily dose of 150 µg/kg of IGF-I for 3 to 10 months. Similarly, Walker and colleagues found an increase in growth rate from 6.5 to 11.4 cm/year in a GHR deficiency patient treated with twice-daily subcutaneous injections of 120 µg/kg of IGF-I.

Wilton reported collaborative data on the treatment of 30 children, aged 3 to 23 years, with GHI from GHR deficiency or GHD-IA with anti-GH antibodies; the dosage of IGF-I varied from 40 to 120 µg/kg given twice daily. Except for the two oldest individuals, growth rates increased in all subjects by at least 2 cm/year.

In another study, a mean increment of more than 4 cm/year in growth velocity was found in 11 prepubertal children treated with 80 µg/kg twice a day. This study also demonstrated a significant inverse relationship between growth response to exogenous IGF-I and the severity of the GHI phenotype.

Longer-term studies of IGF-I treatment of GHI have demonstrated a persistent but progressively waning effect. Data from a European collaborative trial of 17 patients treated for at least 4 years showed an increase in mean height SDS from -6.5 to -4.9, with two adolescents reaching the 3rd percentile, and emphasize the importance of early diagnosis and initiation of therapy. Side effects included hypoglycemia, headache, convulsions, urolithiasis, and papilledema; the latter, which suggests the possibility of pseudotumor cerebri, resolved spontaneously while IGF-I was being administered. In the longest treatment study, Backeljauw and Underwood showed data similar to the European study, with an initial burst of growth followed by slowing to just above baseline by the 6th year of therapy. Height SDS improved from -5.6 to -4.2 by the end of the 6th year.

A randomized, double-blind, placebo-controlled trial of IGF-I therapy in GHR deficiency has been performed in Ecuador, probably the only place where the patient population is sufficiently large and homogeneous to permit such investigation. The placebo group grew 4.4 cm/year during the same time, and then their growth rate increased to 8.4 cm/year during IGF-I treatment. The incidence of hypoglycemia was equal in the two groups. One recipient of IGF-I developed papilledema, which resolved spontaneously while on treatment.

Although these early studies are promising, little is known about the long-term effects of IGF-I or about the optimal dose or frequency of administration. Taken together, the IGF-I treatment studies show that the growth response is neither as successful nor as long-lived as that of GHD children treated with exogenous GH. The failure of serum levels of IGFBP-3 to increase with IGF-I administration underscores the relevance of the IGFBPs to IGF pharmacokinetics. Nevertheless, these data indicate that the IGF peptides, long considered to function as autocrine or paracrine growth factors, can act as classic endocrine hormones.
EXCESS GROWTH AND TALL STATURE

Although as many children have a height greater than 2 SD above the mean as have a height greater than 2 SD below the mean, referral for tall stature is less common than referral for short stature. This pattern speaks eloquently to the psychosocial pressures to which children with growth disorders are subjected. Nevertheless, it is essential to identify situations in which tall stature or an accelerated growth rate provides a clue for the diagnosis of an underlying disorder (Table 23-14).
STATURAL OVERGROWTH IN THE FETUS

Maternal Diabetes Mellitus

Maternal diabetes mellitus is the most common cause of large-for-gestational-age (LGA) infants (height or weight

<table>
<thead>
<tr>
<th>TABLE 23-14 – Differential Diagnosis of Statural Overgrowth</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fetal Overgrowth</strong></td>
</tr>
<tr>
<td>Maternal diabetes mellitus</td>
</tr>
<tr>
<td>Cerebral gigantism (Sotos’ syndrome)</td>
</tr>
<tr>
<td>Weaver’s syndrome</td>
</tr>
<tr>
<td>Beckwith-Wiedemann syndrome</td>
</tr>
<tr>
<td>Other IGF-II excess syndromes</td>
</tr>
<tr>
<td><strong>Postnatal Overgrowth Leading to Childhood Tall Stature</strong></td>
</tr>
<tr>
<td>Familial (constitutional) tall stature</td>
</tr>
<tr>
<td>Cerebral gigantism</td>
</tr>
<tr>
<td>Beckwith-Wiedemann syndrome</td>
</tr>
<tr>
<td>Exogenous obesity</td>
</tr>
<tr>
<td>Excess GH secretion (pituitary gigantism)</td>
</tr>
<tr>
<td>McCune-Albright syndrome or MEN associated with excess GH secretion</td>
</tr>
<tr>
<td>Precocious puberty</td>
</tr>
<tr>
<td>Marfan’s syndrome</td>
</tr>
<tr>
<td>Klenefelter’s syndrome (XXY)</td>
</tr>
<tr>
<td>Weaver’s syndrome</td>
</tr>
<tr>
<td>Fragile X syndrome</td>
</tr>
<tr>
<td>Homocystinuria</td>
</tr>
<tr>
<td>XYY</td>
</tr>
<tr>
<td>Hyperthyroidism</td>
</tr>
<tr>
<td><strong>Postnatal Overgrowth Leading to Adult Tall Stature</strong></td>
</tr>
<tr>
<td>Familial (constitutional) tall stature</td>
</tr>
<tr>
<td>Androgen or estrogen deficiency/estrogen resistance (in males)</td>
</tr>
<tr>
<td>Testicular feminization</td>
</tr>
<tr>
<td>Excess GH secretion</td>
</tr>
<tr>
<td>Marfan’s syndrome</td>
</tr>
<tr>
<td>Klenefelter’s syndrome (XXY)</td>
</tr>
<tr>
<td>XYY</td>
</tr>
</tbody>
</table>

GH, growth hormone; IGF-II, insulin-like growth factor II; MEN, multiple endocrine neoplasia.

> the 90th percentile for gestational age). Even in the absence of clinical symptoms or a family history, the birth of an excessively large infant should lead to evaluation for maternal (or gestational) diabetes mellitus.
Sotos’ Syndrome and Beckwith-Wiedemann Syndrome

Two relatively rare syndromes can also cause LGA infants. Children with cerebral gigantism (Sotos’ syndrome) are typically above the 90th percentile for both length and weight at birth. Additional clinical features include a prominent forehead; dolichocephaly; macrocephaly; high-arched palate; hypertelorism with an unusual slant to the eyes; prominent ears, jaw and chin; large hands and feet with thickened subcutaneous tissue; mental retardation; and motor incoordination. Although such children continue to grow rapidly during the early years of childhood, puberty is usually early and causes premature epiphyseal fusion. Most patients have a final height within the normal population range. GH secretion and serum IGF levels are normal, and no cause of the overgrowth in cerebral gigantism has been identified.

Beckwith-Wiedemann syndrome (BWS) is the most common (1/13,700) of a group of disorders. It is associated with excessive somatic and specific organ growth, collectively referred to as overgrowth syndromes, apparently caused by excess availability of the growth factor IGF-II encoded by the gene Igf2. BWS is characterized by fetal macrosomia with omphalocele, with its clinical features due to selective organomegaly, including macroglossia, renal medullary hyperplasia, and neonatal hypoglycemia due to islet cell hyperplasia. As with cerebral gigantism, excessive fetal, neonatal, and childhood growth ultimately leads to early epiphyseal fusion, without an increase in adult height.

Although an association between BWS and disordered regulation of IGF-II gene transcription exists, to date no consistent postnatal abnormality of the GH-IGF axis has been identified. The paternally derived gene for IGF-II is overexpressed, and the maternally transmitted gene is not active. Four children with somatic overgrowth but not the diagnostic features of BWS had Igf2 gene overexpression. Various lines of investigation have localized “imprinted” genes involved in BWS and associated childhood tumors to chromosome 11p15.5. These include, in addition to Igf2, the gene H19, which is involved in Igf2 suppression as well as the gene WT-1 (the Wilms’ tumor gene). Mutations in GPC3, a glypican gene, which codes for an IGF-II neutralizing membrane receptor, cause the related Simpson-Golabi-Behmel overgrowth syndrome.
POSTNATAL STATURAL OVERGROWTH

As stated earlier, both cerebral gigantism and BWS are associated with rapid perinatal growth, but rapid growth usually ends by early to middle childhood. Nevertheless, these conditions should be considered when tall stature in childhood is accompanied by the characteristic phenotypic features or with a history of unexplained fetal overgrowth. As in the case of the child with growth failure, crossing height percentiles between infancy and the onset of puberty is an indication for further evaluation. Although such growth patterns are frequently not of concern to parents, that overly rapid statural growth can indicate serious underlying pathology. Furthermore, as with short stature, children with tall stature must be evaluated in the context of familial growth and pubertal patterns.

Familial (Constitutional) Tall Stature

GH secretion and levels of IGF-I and IGFBP-3 in familial tall stature (FTS) are often in the upper normal range.[1255] Tauber and co-workers[1561] divided 65 children with FTS into a subset with high GH secretion rates (5.4 ± 2.3 mg/L/mm) and frequent secretory bursts (5.1 ± 1.6/day) and another subset with lower GH secretion (2.1 ± 0.5 mg/L/mm) and fewer episodic spikes (3.3 ± 1.3/day). IGF-I levels were higher in the group producing more GH and were normal in the low GH group. The investigators postulated that both enhanced secretion of GH and greater efficiency of GH-mediated IGF-I production might be potential causes of FTS.

Like children with short stature, children with tall stature must be evaluated relative to familial growth patterns and parental target height.[1562] When a family history of tall stature is available, support and reassurance may be all that are required. A careful assessment of pubertal status and bone age facilitates prediction of adult height and usually obviates the need for hormonal therapy. Standard height prediction using Bayley-Pinneau tables, especially for children younger than 12 years of age, tend to overestimate final height with large confidence limits.[1563][1564][1565]

We discourage therapy for boys with predicted adult heights less than 198 cm (6 feet, 6 inches) and girls with predicted adult heights less than 183 cm (6 feet). Indeed, societal changes in attitudes toward tall individuals appear to discourage treatment except in extreme circumstances. The number of patients treated in the United States has fallen markedly since 1970. Therapy, when necessary, is aimed at the acceleration of puberty to cause premature epiphyseal fusion.[1566][1567]

Accordingly, the optimal time for treatment is before the onset of puberty. The earlier the intervention, the more likely that adult height can be decreased, although patients are not usually referred until late childhood or early puberty.

Although some success with lower dosages had been reported, administration of ethinyl estradiol at a dose of 0.15 to 0.3 mg/day is a reasonable starting level in girls and can be increased, if necessary and well-tolerated, to 0.5 mg/day. Conjugated estrogens, 7.5 to 10 mg/day, have also been successful. If breakthrough bleeding occurs, cyclic progestagens may be added to the estrogen therapy. Treatment should be continued until epipyses fuse, because post-treatment growth may be substantial if treatment is stopped early.[1565]

The mechanism of estrogen action is probably complex, because estrogen can affect both GH secretion and serum IGF levels and, more importantly, acts directly on the epiphysis. Estrogen mediates epiphyseal fusion in both girls and boys.[32][33][34] In prepubertal girls, estrogen therapy reduces adult height by as much as 5 to 6 cm, relative to predictions. When therapy is initiated after the onset of puberty, the decrement in adult height is not likely to be as large.

The use of high-dose estrogen in otherwise normal children must be weighed against the known (and unknown) toxicity of such therapy, including nausea, weight gain, edema, and hypertension. During the initial phases of therapy, growth is paradoxically accelerated as the child rapidly progresses through puberty. Other potential problems, such as thromboembolism, cystic hyperplasia of the breast, endometrial hyperplasia, and cancer, have not been definitively related to estrogen therapy in children but should be discussed with the patient and family.

Therapy in boys with tall stature is even more problematic. For the reasons discussed earlier, estrogen is likely to be most efficacious in accelerating epiphyseal fusion but is obviously undesirable in males. Androgens also accelerate skeletal maturation, presumably via aromatization to estrogen but at the price of rapid virilization.
Obesity

Obesity is frequently associated with rapid skeletal growth and early onset of puberty. Patients with obesity tend to have diminished overall GH production but normal high GHBP and IGF-I levels maintaining adequate or enhanced linear growth velocity. Early activation of adrenal androgenesis and premature pubarche are common. Bone age is usually modestly accelerated so that both puberty and epiphyseal fusion occur early and adult height is normal. This association between obesity and growth is so characteristic that the child with obesity and short stature should always be evaluated for underlying pathology, such as hypothyroidism, GHD, Cushing’s syndrome, or Prader-Willi syndrome.
Excess Growth Hormone Secretion

Pituitary gigantism is a rare condition analogous to acromegaly in adults (see Chapter 8). Typically, GH-secreting tumors of the pituitary are eosinophilic or chromophobe adenomas. Their etiology is uncertain, although many result from somatic mutations that generate constitutively activated G proteins with reduced guanosine triphosphatase activity. The resulting increase in intracellular cAMP in the pituitary leads to increased GH secretion.

McCune-Albright syndrome, which is also caused by mutations resulting in constitutive activation of G proteins, may also be characterized by somatotropic tumors and excess GH secretion. GH-secreting tumors have also been reported in multiple endocrine neoplasia and in association with neurofibromatosis and tuberous sclerosis (see Chapter 8). GH excess that occurs prior to epiphyseal fusion results in rapid growth and attainment of adult heights above the expected genetic potential. When GH hypersecretion is accompanied by gonadotropin deficiency, accelerated linear growth may persist for decades, as in the case of the Alton giant, who reached a height of 280 cm by the time of his death in his 20s. Manifestations typical of acromegaly may also appear, such as soft tissue swelling, enlargement of the nose, ears and jaw with coarsening of the facial features, pronounced increases in hand and foot size, diaphoresis, galactorrhea, and menstrual irregularity.

Serum IGF-I levels are elevated, although high IGF-I levels may also be a normal manifestation of puberty. Basal serum GH levels may be normal or increased, but serum GH is not suppressed by administration of glucose (1.75 g/kg of body weight, up to a maximum of 100 g).

Although abnormalities of the sella turcica are often evident on lateral skull films, the demonstration of increased GH-IGF secretion should lead to radiologic evaluation of the hypothalamus and pituitary by MRI or computed tomography. Definitive therapy requires surgical ablation of the tumor. Fortunately, this can usually be accomplished by a transphenoidal pituitary surgery, although macroadenomas may require a more aggressive surgical approach. As described in Chapter 8, the use of somatostatin analogues, dopamine agonists, and novel GHR antagonists is an important component of treatment programs for GH excess.
Precocious Puberty

Precocious puberty, whether mediated centrally (increased gonadotropin secretion, GnRH-dependent) or peripherally (increased androgen and/or estrogen secretion GnRH-dependent), results in accelerated linear growth in childhood, mimicking the pubertal growth spurt. Because skeletal maturation is also accelerated, adult height is frequently compromised. The diagnostic evaluation and management of precocious puberty are discussed in Chapter 24.
Miscellaneous Causes of Tall Stature

Marfan’s syndrome, an autosomal dominant disorder of collagen metabolism, is characterized by hyperextensible joints, dislocation of the lens, kyphoscoliosis, and dissecting aortic aneurysm and often leads to long, thin bones that result in arachnodactyly and moderately tall stature. Homocystinuria, an autosomal recessive disorder, phenotypically resembles Marfan’s syndrome, although patients are usually mentally retarded. The rate of linear growth may increase modestly in hyperthyroidism. Tall stature has been found in patients with familial ACTH resistance due to a defective ACTH receptor. Dosage effects of the SHOX gene may result in tall stature. In females with three copies of the SHOX gene and gonadal dysgenesis, adult stature was +2 to +2.9 SDS. In women with 47,XXX karyotype, mean final heights are around 5 to 10 cm taller and in men with 47,XXY karyotype (Klinefelter’s syndrome), about 3.5 cm taller than population means. Males with an XYY karyotype may also have moderate tall stature. In addition to the SHOX effects, however, the variable degree of estrogen production in some of these syndromes must influence skeletal maturation and final height.

It is worth commenting that although delayed puberty may be associated with short stature in childhood, failure to enter puberty and complete sexual maturation may result in sustained growth during adult life with ultimate tall stature and a characteristic eunuchoid habitus. The description of tall stature with open epiphyses resulting from mutation of the estrogen receptor or from aromatase deficiency underscores the fundamental role of estrogen in promoting epiphyseal fusion and termination of normal skeletal growth.
References


1081


References


References


413. Angelliez-Nicoud P, Binoux M. Autocrine regulation of cell proliferation by the insulin-like growth factor (IGF) and IGF-binding protein-3 protease system in a human prostate carcinoma cell line (PC-3). Endocrinology 1995; 136:54855452.


436. Conover CA, Balle LK, Durham SK, Powell DR. Insulin-like growth factor (IGF)binding protein-3 potentiating IGF action is mediated through the phosphotidylinositol-3-kinase pathway and is associated with alteration in protein kinase B/AKT sensitivity. Endocrinology 2000; 141:30963103.


446. Conover CA, Balle LK, Durham SK, Powell DR. Insulin-like growth factor (IGF)binding protein-3 potentiating IGF action is mediated through the phosphotidylinositol-3-kinase pathway and is associated with alteration in protein kinase B/AKT sensitivity. Endocrinology 2000; 141:30963103.


References


1092


References


646. Gluckman PD, Butler JH, Elliott TB. The ontogeny of somatotropic binding sites in ovine hepatic membranes. Endocrinology 1983; 112:16071612.


Counts DR, Grwirtsman H, Carlson LMS, et al. The effects of anorexia nervosa and refeeding on growth hormone-binding protein, the insulin-like growth factors (IGFs), and the IGF-binding proteins. J Clin Endocrinol Metab 1992; 75:7622.


1096
References


References


References


References


1108


References


References


1535. Renihan AG, D'Oyrier ST, Shahet SM. Colorectal neoplasia in acromegaly: the reported increase in prevalence is overestimated. Gut 2000; 46:440441.


Chapter 24 - Puberty: Ontogeny, Neuroendocrinology, Physiology, and Disorders

Melvin M. Grumbach
Dennis M. Styne

Puberty should not be considered as a de novo event but rather as a phase in the continuum of the development of gonadal function and the ontogeny of the hypothalamic-pituitary-gonadal system in the fetus, through puberty, to the attainment of full sexual maturation and fertility. By puberty, secondary sexual characteristics appear and the adolescent growth spurt occurs, which result in the striking sexual dimorphism of mature individuals, fertility is achieved, and profound psychological effects ensue. The changes are a consequence of stimulation of the gonads by pituitary gonadotropins and increases in gonadal steroid output. Adolescence, a term usually considered to relate to the psychosocial aspects of the teenage years, is accompanied by the onset of adult patterns of socioeconomic and economic behavior.

The human being is the most reproductively successful of mammals, and many anthropologists have attributed this success to the prolonged pattern of human growth and development and the delay in attaining full sexual maturity. The evolution of the human scheme of growth involves the development of two stages: a childhood stage and an adolescent stage that includes an adolescent or pubertal growth spurt. Fig. 24-1. Not even our closest biologic relative, the chimpanzee, which matures twice as rapidly as the human, unequivocally exhibits these two stages including the unique human adolescent growth spurt. (The estimated date for divergence of the chimpanzee and human lineages is 4 million to 5 million years ago.)

Evolution theorists proposed that a critical part of human success and of many biosocial characteristics emanates from the learning and practice of adult behaviors related to sex and child rearing, particularly provisioning (not just infants) with food, which is unique to humans. These include learning skills related to production of food, cooperative hunting, division of labor according to sex, sharing food, tool making, and adjusting to the social organization and cultural environment. Given, on the other hand, noting that tool making preceded the evolutionary development of adolescence, suggested that, in addition, the evolution and value of human childhood and adolescence and this unique pattern of growth and development have had a significant role in the comparatively striking reproductive advantage and success of the human being. May has called this process of selection "selection for reproductive success."

Historical evidence suggests that puberty occurs at an earlier age today than in the past, usually reflected by age of menarche, which is removed by several years from the first sign of secondary development in girls. The average age of menarche in industrialized European countries has decreased 2 to 3 months per decade over the past 150 years, and in the United States the decrease has been approximately 2 to 3 months per decade in the last century (Fig. 24-2). However, this secular trend has slowed or ceased in "developed" countries such as the United States, Australia, and Western Europe (e.g., Britain and Holland) since approximately 1940, presumably because of improved socioeconomic status and health and the benefits of urbanization. The social class difference in menarcheal age has narrowed or disappeared in most countries, whereas in Denmark, Spain, and Brazil there remains evidence for a continued decline, at least in certain local regions. In the nomadic Lapp culture, in which the standard of living changed little between 1870 and 1930, no trend toward earlier menarche was found.

According to a 1973 survey by the U.S. National Center for Health Statistics, the age of menarche in the United States is 12.8 years, and data published in 1997 indicate that this age remains true for white but not for black girls, in whom the mean age of menarche is 6 months earlier. At present there remains a difference in the age of attainment of stages of puberty in different countries even if stability was reached; for example, Japanese boys undergo changes in testicular size about 1 year earlier than Swiss boys reach the same stage. Remarkably, there is a reverse secular trend in Northern Italy (in women born between 1950 and 1959) and other areas of Europe leading to a later age of menarche; this has been attributed putatively to a resurgence of physical and psychological stress.

The method of ascertainment of the age of menarche is of importance. Contemporaneous recordings are performed with the probit method of asking, "yes" or "no," are you menstruating? These may be incorrect because of social pressures of the culture and socioeconomic group considered. Recalled ages of menarche are used in other studies and considered to be accurate within 1 year (in 90% of cases) during the teenage years and in older women, too.

Earlier menarche within the normal range may have health consequences. International studies show that earlier age at menarche is associated with a greater risk of development of breast cancer; indeed, the risk for women with menarche at age younger than 12 years is increased by about 50% compared with those with menarche at 16 years. Further, there is indirect evidence relating earlier menarche to increasing likelihood of hepatocellular carcinoma.
If the age of puberty was later in past centuries than at present, the age of attaining adult height was also later. Surveys of army recruits, schoolchildren, and workers in Europe and America, as well as records of slaves in the United States, show that large portions of the population in the past two centuries continued to grow a considerable amount into their early 20s, whereas modern adolescents cease to grow and reach stable heights by about 17 years of age. The adult heights attained during the 18th and 19th centuries were often at modern 25th percentiles or less. The secular trend toward increased height is more marked than the secular trend toward earlier puberty.

Dietary modification may affect the age of menarche and other aspects of puberty. The Harvard Longitudinal Studies of Childhood Health and Development related dietary intake to menarche and found that girls had earlier menarche if they were taller and consumed more animal protein and less vegetable protein as early as 3 to 5 years of age; further, girls had earlier peak growth if they had a history of higher dietary fat intakes at 1 to 2 years of age and higher animal protein intakes at 6 to 8 years. Girls had higher peak velocity if, controlling for body size, they consumed more calories and animal protein 2 years before peak growth. Moderate obesity (up to 30% above normal weight for age) is associated with earlier menarche and conversely adult women with obesity had a history of a tendency toward an earlier age of menarche, which is the cause and which is the effect is not clear. Pathologic obesity is associated with delayed menarche. Black American girls are advanced in secondary sexual development compared with white American girls of the same age during the first three stages of puberty; this may be related, in part, to the higher prevalence of obesity in black girls, although, as stated later, there appears to be a genetic influence in ethnic differences as well.

The interaction of socioeconomic conditions, nutrition, energy expenditure, states of health, and puberty is of particular importance in areas of the world where nutrition is suboptimal. South America and Africa have a pattern in which rural children fare better and have earlier puberty and taller stature than urban children, the opposite of the pattern found in previous eras, demonstrating a disturbing trend of adverse nutritional conditions in crowded urban centers. The onset of secondary sexual development in girls of the Kikuyu in Kenya is 13.0 years with menarche at 15.9 years; in contrast to the age of onset of puberty in black American girls of 8.9 years with menarche at 12.2 years: the Kikuyu start later and have a shorter time of transition to menarche than the American girls. Kikuyu boys enter puberty before or at the same age as Kikuyu girls. Further, boys of the Hadza of Tanzania enter puberty 2 years earlier than girls. In contrast, in the United States and wherever else puberty has been studied in the Western world, girls enter puberty before boys by an average of at least 6 months and often more.

Other studies demonstrate the effect of chronic disease on the age of menarche. Delay can occur in any serious chronic condition that is not adequately treated; for example, celiac disease can delay menarche as well as decrease growth in childhood, as does infection with Helicobacter pylori. The earlier suggestion that blindness may advance the age of menarche is not supported by more recent studies. Puberty starts at a later age and the period of pubertal development lasts longer at high altitudes than at low altitudes even when nutritional status is similar.

Strenuous physical activity in girls, especially, but not necessarily, when associated with low body weight, can delay or arrest puberty. On the contrary, inactive, bedridden children with mental retardation reach menarche at an earlier age and at a lower proportion of body fat value than do similarly retarded children who are more active.

There is evidence that even the living environment might influence menarche. A convergence of the onset of menses in women or girls living together was noted. Several publications followed that studied the timing of the spontaneous onset of menses in women living together or when auxiliary odor scent (presumably containing pheromones) was present in women; some studies noted synchrony and others did not. Subsequently, the methodology of the studies was criticized and the positive findings deemed incorrect. Nonetheless, one group of investigators found and studied menstrual synchrony in a variety of conditions in college women and older subjects and found that closeness of sleeping conditions was not a prerequisite to synchronizing menstrual cycles but that being close friends was of significance in the phenomenon.

Genetic factors play an important role in the onset of puberty, as illustrated by the similar age of menarche in members of an ethnic population and in mother-daughter and sibling pairs.

Further support for the influence of genetics on the age of menarche is found in the concordance of the ages of pubertal developmental stages and menarche, which are closer between monozygotic than dizygotic twins. Secondary sexual development occurs earlier in black girls than in white girls in the United States, and although we have alluded to the influence of body mass index (BMI) upon menarche, genetics appears of importance although there is no apparent effect of social or economic factors on this relationship (see later). Thus, when socioeconomic and environmental factors lead to good nutrition, general health, and infant care, the age of onset of puberty in normal children is determined largely by genetic factors.

The influence of genetics on mother-daughter comparisons of age of menarche may be subject to complicating factors. In one study of white girls, there was a trend for maternal age at menarche to predict adolescent’s age at menarche, but breast development, weight, family relations (including the absence of a father), and depressive affect were predictive of age at menarche in this group with family relations more strongly predicting the age at menarche than the influence of breast development or weight. This study also raised the question of whether psychological stress can decrease the age of menarche or whether stress is likely to occur because of the earlier menarche. On the other hand, another study of age of menarche confirmed the influence of the age of menarche in mothers on the age of menarche in daughters but found no influence of stress related to family problems upon early menarche. Thus, environmental influences should be considered in the study of the age of menarche, although only some of these influences appear to have greater significance than genetic patterns.
PHYSICAL CHANGES OF PUBERTY

Secondary Sexual Characteristics

Comparative description of the physical changes between individuals and populations requires an objective and reproductible method of describing the maturation of secondary sexual characteristics. Tanner published standards of the most useful signs of sexual maturation that have been widely used throughout the world (Fig. 24-3). Self-assessment scales of adolescent maturation are available and are used in some studies to avoid the embarrassment of a secondary sexual examination in normal children and adolescents. However, the answers to self-assessments may be influenced by the subjects’ views of what is considered normal or by wishes to conform with normal development and may be less accurate in some ethnic groups than others.

Female

Two distinct phenomena occur in the female. The development of the breast and its modified apocrine glands is primarily under the control of estrogens secreted by the ovaries (see Fig. 24-3). The growth of pubic and axillary hair (see Fig. 24-4) is mainly under the influence of androgens secreted by the adrenal cortex and the ovary. The glandular and connective tissue of the mammary gland begins to develop at the onset of pubertal maturation. Thus, lobules composed of small ductules and cellular connective tissue develop to a more pronounced degree in the female at puberty. Proliferation of fatty and connective tissue accounts for 80% of the volume of the adult, nonlactating female breast.

The classification of the stages of breast development depends on specific characteristics common to the female breast but does not include size or inherent shape of the breasts, which are determined by genetic and nutritional factors (see Fig. 24-3). Four stages were described by Stratz, and modifications were made to the schema by Tanner, who produced the most widely utilized staging. The initial breast development may be unilateral for several months and may be cause for unfounded concern by girls or parents. Indeed, surgical biopsies have been performed inappropriately in girls in whom it was not appreciated that asymmetrical development is normal. There are unusual cases of agenesis of the breast in which no glandular or fat enlargement occurs regardless of the level of estrogen stimulation. On the other hand, virginal breast hypertrophy, extreme and rapid increase in breast size at the onset of puberty, is rare. It has been attributed, at least in part, to increased sensitivity to estrogen action or to increased local estrogen synthesis and growth factors.

Changes in the diameter of the papilla of the nipple are sequential and linked to stages of pubertal development. Nipple papilla diameter does not increase much during pubic hair stage 1 to 3 or breast stage 1 to 3 (diameter is 3 to 4 mm) but does increase after breast stage 3, providing an objective method of differentiating stage 4 from stage 5 (final diameter is approximately 9 mm) (Table 24-1).

The stage of breast development is usually comparable to the stage of pubic hair development in normal girls, but as different endocrine organs control these two processes, these features mature at different ages, and discordance can occur in disease states, the stages should be classified separately (Table 24-2 and Table 24-3).

Areolar diameter also increases in boys at puberty, and most boys have palpable glandular enlargement of the breast, transient gynecomastia (see later in this chapter).

Although rarely evident clinically in individual girls because of its subtle nature, increase in height velocity (rather than
breast development) is actually the first sign of puberty in girls; however, breast budding is what is first noted by most lay or medical observers. There are changes in the appearance of the vaginal opening at puberty. Dullling and thickening of the vaginal mucosa from the prepubertal reddish glistening appearance are due.

<table>
<thead>
<tr>
<th>TABLE 24-1</th>
<th>Nipple Diameter Compared with Breast and Pubic Hair Stages: Comparison of Longitudinal and Cross-Sectional Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nipple Size (mm)</strong></td>
<td><strong>Cross-Sectional Data</strong></td>
</tr>
<tr>
<td>Breast</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2.89 (0.81)</td>
</tr>
<tr>
<td>2</td>
<td>3.28 (0.89)</td>
</tr>
<tr>
<td>3</td>
<td>4.07 (1.32)</td>
</tr>
<tr>
<td>4</td>
<td>7.74 (1.64)</td>
</tr>
<tr>
<td>5</td>
<td>9.94 (1.38)</td>
</tr>
<tr>
<td>Public hair</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2.95 (1.02)</td>
</tr>
<tr>
<td>2</td>
<td>3.32 (0.91)</td>
</tr>
<tr>
<td>3</td>
<td>4.11 (1.54)</td>
</tr>
<tr>
<td>4</td>
<td>7.15 (1.81)</td>
</tr>
<tr>
<td>5</td>
<td>9.66 (1.59)</td>
</tr>
</tbody>
</table>

*Results are means ± standard deviation (SD; in parentheses). Significantly different from previous stage, P <.05.*


The growth and maturation of the penis usually correlate closely with pubic hair development because both features are under androgen control. However, for the most accurate assessment the stages of pubic hair development and genital development should be determined independently and recorded separately because discordant stages are a clue to potential disease states of the adrenal gland or testes. Partially to the development of the thigh but not up the linea alba or elsewhere above the base of the inverse triangle. Most men will have further spread of the pubic hair. (Photographs from Van Wieringen JD, Wafelbakker F, Verbrugge HP, et al. Growth Diagrams 1965 Netherlands: Second National Survey on 024 Year Olds. Netherlands Institute for Preventative Medicine TNO. Groningen, Wolters-Noordhoff, 1971. © Wolters-Noordhoff, Groningen.)

TABLE 24-2 -- Reported Mean Ages (Years) at Onset of Sexual Maturity Stages in Females

<table>
<thead>
<tr>
<th>Study</th>
<th>Longitudinal or Cross-sectional</th>
<th>No. of Subjects</th>
<th>Age Range</th>
<th>Breast Stages</th>
<th>Public Hair Stages</th>
<th>PHV</th>
<th>Menarche</th>
</tr>
</thead>
<tbody>
<tr>
<td>Billwicz et al. (1981), United Kingdom</td>
<td>L</td>
<td>753</td>
<td>917</td>
<td>10.8</td>
<td>12.0</td>
<td>13.1</td>
<td>14.0</td>
</tr>
<tr>
<td>Roy et al. (1972), France</td>
<td>L</td>
<td>80</td>
<td>715</td>
<td>11.4</td>
<td>12.5</td>
<td>13.4</td>
<td>14.0</td>
</tr>
<tr>
<td>Taranger (1976), Sweden</td>
<td>L</td>
<td>90</td>
<td>817</td>
<td>11.0</td>
<td>11.8</td>
<td>13.1</td>
<td>15.6</td>
</tr>
<tr>
<td>Largo and Prader (1983), Switzerland</td>
<td>L</td>
<td>142</td>
<td>818</td>
<td>10.9</td>
<td>12.2</td>
<td>13.2</td>
<td>14.0</td>
</tr>
<tr>
<td>Van Wieringen et al. (1971), Holland</td>
<td>L</td>
<td>110</td>
<td>12.1</td>
<td>13.4</td>
<td>15.2</td>
<td>11.3</td>
<td>12.2</td>
</tr>
<tr>
<td>Neyzi et al. (1975), Turkey</td>
<td>C</td>
<td>1468</td>
<td>917</td>
<td>10.0</td>
<td>11.6</td>
<td>12.8</td>
<td>15.2</td>
</tr>
<tr>
<td>Villareal et al. (1989), Mexican-American</td>
<td>C</td>
<td>699</td>
<td>1017</td>
<td>10.9</td>
<td>12.2</td>
<td>13.9</td>
<td>15.1</td>
</tr>
<tr>
<td>Roche et al. (1995), United States (Ohio)</td>
<td>L</td>
<td>67</td>
<td>9.516</td>
<td>11.2</td>
<td>12.0</td>
<td>12.4</td>
<td>11</td>
</tr>
<tr>
<td>Herman-Giddens (1997), United States</td>
<td>C</td>
<td>17,077</td>
<td>312</td>
<td>1.638</td>
<td>8.9</td>
<td>10.2</td>
<td>8.8</td>
</tr>
<tr>
<td>African-American</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>15,439</td>
<td>10.0</td>
<td>11.3</td>
<td>10.5</td>
<td>11.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
testis is normally larger than the left, and the left testis is located lower in the scrotum than the right testis.

The phallus is most accurately measured stretched while in the flaccid state as there is much variation between individuals in the length of the unstretched penis. The length of the erectile tissue (excluding the foreskin) increases from an average of 6.2 cm in the prepubertal state to 12.4 ± 2.7 cm in the white adult. Ethnic differences have been noted; the mean value in black men is 14.6 cm and in Asians 10.6 cm.  

As in girls, the areolar diameter increases in boys during puberty; a distinct separation between the sexes occurs in stage 4, when female areolar diameter increases much more than male values. In addition, in gynecomastia the areolar diameter increases above normal.
Limits of Normal Pubertal Development

Data on the normal variation in pubertal development in the United States are becoming more plentiful, but guidelines remain controversial. A prior survey of the age of attainment of various stages of puberty in the United States began with subjects 12 years of age and, although useful in defining the upper limits of normal pubertal development, the survey is uninformative about the lower limits of the age of onset of puberty.\(^9\)\(^8\)\(^7\)\(^6\) A later longitudinal study (see Table 24-3) of white boys and girls started at 9.5 years of age and adds much to the determination of the mean age of attainment of stages of puberty\(^8\)\(^9\)\(^7\); nonetheless, it starts too late to include normal children who enter puberty at an earlier age (see Table 24-3). A large cross-sectional study sponsored by the American Academy of Pediatrics in which 17,070 girls visited the offices of 225 specially trained pediatricians across the United States started at 3 years of age but ended at 12 years of age and so excludes a proportion of normal children who enter puberty at a later age.\(^8\) The standard deviation (SD) of the longitudinal study\(^9\) was low, 1.0 years or less in most cases, whereas the cross-sectional study had a larger SD of approximately 2 years.\(^9\) This difference may be due to the difference in the study design, but it may reflect a limit in the spread of the upper end of the normal pubertal curve (rarely does even the subject with the most severe constitutional-delay spontaneously enter puberty after 18 years of age), where there may be a skewing of the normal age of onset of puberty to an earlier spread of ages.

The latest and largest multiracial and ethnic study of the age of onset of puberty in 2114 American boys aged 8 through 19 years suggests that a decrease has occurred over the last decades, but the observations are controversial.\(^8\)\(^9\)\(^7\)\(^6\) as is the report of the data for girls presented before. Using the National Health and Nutrition Examination Survey (NHANES) III database, the authors determined that the onset of pubic hair occurs earlier in black boys than in white boys, with a tendency to later onset in Mexican American boys. Between the ages of 8 and 9 no white boys had pubic hair, whereas 5.3% of black boys were at least at Tanner stage 2.\(^7\) The age of genital development was based on the visual change in scrotal skin and enlargement of the testes (an unsatisfactory method of assessing the size of the testes compared with direct examination); by the beginning of age 8, the earliest age studied, 29.35% of white boys, 37.8% of black boys, and 27.3% of Mexican American boys purportedly had genital development. These observations suggest an earlier onset of puberty in a substantial number of boys before the previously accepted lower age limit of 9 years. Further, the height and weight of these boys reaching the earlier stages of puberty were greater than reported in the past, although the heights and weights of boys reaching adult ages were equivalent to the past data. (The

\*African-American girls enter puberty approximately 1 to 1½ years earlier than white girls and begin menses 8½ months earlier.

<table>
<thead>
<tr>
<th>TABLE 24-3 – Descriptive Statistics for the Timing of Sexual Maturity Stages in Females</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Onset of Stage</strong></td>
</tr>
<tr>
<td>Stage</td>
</tr>
<tr>
<td><strong>Breast Stages</strong></td>
</tr>
<tr>
<td>Stage 2</td>
</tr>
<tr>
<td>Roche et al. (Ohio)(^28)</td>
</tr>
<tr>
<td>Herman-Giddens et al. (USA)(^27)</td>
</tr>
<tr>
<td>African-American</td>
</tr>
<tr>
<td>White</td>
</tr>
<tr>
<td>Stage 3</td>
</tr>
<tr>
<td>Roche et al. (Ohio)(^28)</td>
</tr>
<tr>
<td>Herman-Giddens et al. (USA)(^27)</td>
</tr>
<tr>
<td>African-American</td>
</tr>
<tr>
<td>White</td>
</tr>
<tr>
<td>Stage 4</td>
</tr>
<tr>
<td>Roche et al. (Ohio)(^28)</td>
</tr>
<tr>
<td><strong>Tanner Pubic Hair</strong></td>
</tr>
<tr>
<td><strong>Stage</strong></td>
</tr>
<tr>
<td><strong>Tanner Stage 2</strong></td>
</tr>
<tr>
<td>Roche et al. (Ohio)(^28)</td>
</tr>
<tr>
<td>Herman-Giddens et al. (USA)(^27)</td>
</tr>
<tr>
<td>African-American</td>
</tr>
<tr>
<td>White</td>
</tr>
<tr>
<td><strong>Tanner Stage 3</strong></td>
</tr>
<tr>
<td>Roche et al. (Ohio)(^28)</td>
</tr>
<tr>
<td>Herman-Giddens et al. (USA)(^27)</td>
</tr>
<tr>
<td>African-American</td>
</tr>
<tr>
<td>White</td>
</tr>
<tr>
<td><strong>Tanner Stage 4</strong></td>
</tr>
<tr>
<td>Roche et al. (Ohio)(^28)</td>
</tr>
<tr>
<td><strong>Menarche</strong></td>
</tr>
<tr>
<td>Herman-Giddens et al. (USA)(^27)</td>
</tr>
<tr>
<td>African-American</td>
</tr>
<tr>
<td>White</td>
</tr>
<tr>
<td><strong>Percent Menstruating</strong></td>
</tr>
<tr>
<td>African-American</td>
</tr>
<tr>
<td>White</td>
</tr>
<tr>
<td><strong>Onset Axillary Hair (Stage 2)</strong></td>
</tr>
<tr>
<td>African-American</td>
</tr>
<tr>
<td>White</td>
</tr>
</tbody>
</table>
new growth charts from the National Center for Health Statistics, found at www.dcd.gov, are virtually superimposable on the charts developed in the 1970s, indicating no change in growth rate in the overall population. The earlier onset of puberty occurred in only a subset of boys.)

A number of aspects of this study have been questioned. For example, the observers for NHANES III did not measure testicular volume, an important factor in the assessment of pubertal development, and a one-stage variance was allowed between the observers’ finding and the quality control, a variance that is quite significant between stage 1 and stage 2 of pubertal development in boys. Thus, just as there is difficulty in determining the difference between stages 1 and 2 of breast development, there is substantial difficulty in determining the difference between stages 1 and 2 of genital development without a direct assessment of testicular size. Clearly, the United States is lacking a comprehensive large, preferably longitudinal, study that would start early enough to include the youngest normal pubertal subjects and continue long enough to include the oldest. Failing such a study, we define the normal onset of puberty as 2.5 SD from the mean (percentile 99.4) and use a combination of the studies just discussed to establish reasonable guidelines for such boundaries.

The mean age for a white boy to reach genital stage 2 or pubic hair stage 2 in the longitudinal study is 11.2 years with an SD of 0.7 years; a 2.5 SD range from 8.9 to 13.3, and for simplicity we say the mean age of onset of puberty in boys is 11 years with the limits of 2.5 SD at 9 to 13.5 years of age (these are similar to the limits invoked in the past), although it is quite possible that some normal boys, especially black boys, enter puberty between 8 and 9 years of age.

In the evaluation of girls, it may be difficult for the physician to detect the onset of stage 2 breast development during an office visit, whereas assessment of stage 3 is generally obvious. Using the longitudinal data of a mean age of 11.2 and an SD of 0.7 years, the normal range is defined as 8.9 to 13.3. This correlates well with the U.S. Health Examination Survey of 1977. Using the large number of white and black girls in the cross-sectional study and the average ages of stage 2 and stage 3 from that study, the mean age of breast development for white girls is 10.6 years with 2.5 SD encompassing 8.7 to 14.5 years. However, in the cross-sectional study 3.0% of white girls had stage 2 breast development by 6 years and 5.0% by 7 years, whereas 6.4% of blacks had stage 2 breast development by 6 years and 15.4% by 7 years. It seems best to combine these findings, deviate from the ± 2.5 SD limits in this case, and set the mean at 10.6 years and the range between 6.7 and 13 years for whites and accept 8.9 years as the mean age for black girls with -2.5 SD at 6 years and the upper age range at 13 years. Black girls have an earlier onset of pubertal development of about 1 year even though their average age of menarche in the cross-sectional study was only 8½ months different (12.2 years for blacks and 12.9 for whites).

These data provide guidelines for choosing which candidates with early onset of puberty need expensive diagnostic tests and long-term therapy; many of the children who previously appeared to have mild sexual precocity may now be considered to represent normal variation. It would be inappropriate to include the youngest normal pubertal subjects and continue long enough to include the oldest. Failing such a study, we define the normal onset of puberty as 2.5 SD from the mean (percentile 99.4) and use a combination of the studies just discussed to establish reasonable guidelines for such boundaries.

The sequence of events at puberty in females. The design of the figure is described in the legend of Figure 24-6. (From Marshall WA, Tanner JM. Variations in pattern of pubertal changes in girls. Arch Dis Child 1970; 45:1323.)

The sequence of events at puberty in males. An average is represented in relation to the scale of ages; the range of ages within which some of the changes occur is indicated by the figures below. (From Marshall WA, Tanner JM. Variations in the pattern of pubertal changes in boys. Arch Dis Child 1970; 45:1323.)
study extensively all normal girls found in these lower limits of normal. We emphasize that family history, the rapidity of development of secondary sexual characteristics, the rate of growth, and the presence or absence of central nervous system (CNS) or other types of disease must enter into the decision of whether to evaluate a child. As we recommended in the previous edition, the Drug and Therapeutics and Executive Committees of the Lawson Wilkins Pediatric Endocrine Society support such a revision of the lower limits of the normal age of onset of puberty. Thus, the limits of normal puberty might be set at 9 to 14 years for boys (rounding 13.5 years to 14 years for simplicity), at 7 to 13 years for white girls and 6 to 13 years for black girls.

On the basis of the Herman-Giddens study, it seems likely that the onset of puberty is earlier today than before 1950 to 1960, and that the age of menarche in white girls has changed little if at all over the past four to five decades. Marti-Henneberg

### TABLE 24-6 -- Correlation of Testicular Volume with Stage of Pubertal Development

<table>
<thead>
<tr>
<th>Parameter</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight (kg)</td>
<td>1.8</td>
<td>4.5</td>
<td>8.2</td>
<td>10.5</td>
<td>14.8</td>
</tr>
<tr>
<td>Volume (cm³)</td>
<td>2.5</td>
<td>3.4</td>
<td>9.1</td>
<td>11.8</td>
<td>14.6</td>
</tr>
</tbody>
</table>

*Volume estimated by comparison with ellipsoid of known volume (orchidometer) that is equal to or smaller than the testes. Data from Zachmann et al.
Volume by comparison with orchidometer. Data from Waaler et al.
*Measurement with calipers and average volume of both testes calculated by 0.52 × longitudinal axis × transverse axis. Data from Waaler et al.

### TABLE 24-7 -- Mean Values of Age, Height, Weight, Body Mass Index, and Serum Hormone Levels by Pubertal Stage, in 515 (Ohio) Boys (237 African-American, 278 White; Age 1015 yr at Intake) Followed Every 6 Months for 3 Years

<table>
<thead>
<tr>
<th>Pubertal Stage</th>
<th>Variable</th>
<th>PS1</th>
<th>PS2a</th>
<th>PS2b</th>
<th>PS3</th>
<th>PS4</th>
<th>PS5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (yr)</td>
<td>11.44</td>
<td>12.18</td>
<td>12.79</td>
<td>13.74</td>
<td>14.63</td>
<td>15.19</td>
<td></td>
</tr>
<tr>
<td>Height (cm)</td>
<td>144.2</td>
<td>149.8</td>
<td>154.6</td>
<td>162.3</td>
<td>169.9</td>
<td>173.3</td>
<td></td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>38.18</td>
<td>41.65</td>
<td>47.27</td>
<td>54.67</td>
<td>61.1</td>
<td>66.88</td>
<td></td>
</tr>
<tr>
<td>Body mass index (kg/m²)</td>
<td>18.1</td>
<td>18.4</td>
<td>19.5</td>
<td>20.6</td>
<td>21.0</td>
<td>22.2</td>
<td></td>
</tr>
<tr>
<td>Testosterone: nmol/L (ng/dL)</td>
<td>0.6 (23)</td>
<td>3.0 (86)</td>
<td>4.9 (141)</td>
<td>11.5 (331)</td>
<td>13.4 (338)</td>
<td>15.5 (449)</td>
<td></td>
</tr>
<tr>
<td>White subjects</td>
<td>0.6 (16)</td>
<td>2.9 (83)</td>
<td>4.6 (132)</td>
<td>9.7 (281)</td>
<td>13.3 (383)</td>
<td>14.6 (422)</td>
<td></td>
</tr>
<tr>
<td>Free testosterone: pmol/L (ng/dL)</td>
<td>11 (0.33)</td>
<td>60 (1.74)</td>
<td>114 (3.28)</td>
<td>284 (8.49)</td>
<td>413 (11.9)</td>
<td>504 (14.5)</td>
<td></td>
</tr>
<tr>
<td>DHEAS: μmol/L (μg/dL)</td>
<td>2.71 (99.7)</td>
<td>3.31 (121.6)</td>
<td>4.04 (148.7)</td>
<td>4.75 (175.0)</td>
<td>5.08 (187.0)</td>
<td>5.89 (217.0)</td>
<td></td>
</tr>
<tr>
<td>TeBG (nmol/L)</td>
<td>34.6</td>
<td>33.3</td>
<td>28.4</td>
<td>21.5</td>
<td>14.4</td>
<td>10.7</td>
<td></td>
</tr>
</tbody>
</table>

DHEAS, dehydroepiandrosterone sulfate; TeBG, testosterone-binding globulin.

PS1, absence of public hair, testicular volume < 3 ml; PS2a, absence of public hair, testicular volume 3 ml; PS2b, Tanner stage 2 public hair; PS3, Tanner pubic hair stages.
*Duncan post-hoc analysis significant at P < .01.

and Vizmanos showed that the earlier normal girls entered puberty, the longer the duration of puberty before menarche. Thus, girls who started puberty at 9, 10, 11, and 12 years of age experienced menarche 2.77, 2.27, 1.78, 1.44, and 0.65 years later, respectively. Interpretations of the age of onset of puberty in white American girls are tentative as there are no previous comparably large data sets and none that analyzes white and black girls separately.

There is a disturbing trend in the United States and elsewhere to an increasing prevalence of overweight (often defined as BMI > 85th percentile for age) and obesity (BMI > 95th percentile for age) in childhood. An analysis of the same pediatric office study data used for the updated statistics on the age of puberty for girls in the United States noted earlier yielded a relationship between BMI corrected for age and an earlier onset of puberty and of reaching various stages of puberty for black girls but less so for white girls. Further, the study of the age of onset for boys of various ethnic groups in the United States, although controversial, indicates a direct relationship of BMI to age of onset of puberty in boys. If the trend to increased BMI in U.S. children continues, we may anticipate a noncontroversial decrease in the age of attainment of puberty in the future.

International data are available for the age of pubertal stages

1124

and Vizmanos showed that the earlier normal girls entered puberty, the longer the duration of puberty before menarche. Thus, girls who started puberty at 9, 10, 11, and 13 years of age experienced menarche 2.77, 2.27, 1.78, 1.44, and 0.65 years later, respectively. Interpretations of the age of onset of puberty in white American girls are tentative as there are no previous comparably large data sets and none that analyzes white and black girls separately.

There is a disturbing trend in the United States and elsewhere to an increasing prevalence of overweight (often defined as BMI > 85th percentile for age) and obesity (BMI > 95th percentile for age) in childhood. An analysis of the same pediatric office study data used for the updated statistics on the age of puberty for girls in the United States noted earlier yielded a relationship between BMI corrected for age and an earlier onset of puberty and of reaching various stages of puberty for black girls but less so for white girls. Further, the study of the age of onset for boys of various ethnic groups in the United States, although controversial, indicates a direct relationship of BMI to age of onset of puberty in boys. If the trend to increased BMI in U.S. children continues, we may anticipate a noncontroversial decrease in the age of attainment of puberty in the future.

International data are available for the age of pubertal stages

collected in cross-sectional and longitudinal studies (see Table 24-2 and Table 24-4). The results show similarities to United States data in boys (see Table 24-4), but the data differ for girls (see Table 24-2). The effect of increased BMI on age of onset has been studied in Sweden, where a greater gain in BMI between 2 and 8 years was related to an earlier onset of puberty and each BMI unit gained decreased the age of onset of puberty 0.6 years in boys and 0.7 years in girls. Further, each increased unit of BMI gain in childhood also reduced the height gain in adolescence, 0.88 cm for boys and 0.51 cm for girls, and there was no effect of BMI on final height.

Usually, growth patterns for black and white children are plotted on the standard North American growth charts, but ethnic group-specific growth velocity charts are available for some groups.
Other Dimorphic Physical Changes

Other physical changes show sexual dimorphism. In boys, both the membranous and cartilaginous components of the vocal cords lengthen during puberty. In the peripubertal period the length of the vocal cords in both boys and girls is about 12 to 15 mm, of which the membranous portion is 7 to 8 mm. In adult men the vocal cords attain a length of 18 to 23 mm (membranous portion 12 to 16 mm), whereas in women the cords enlarge only slightly (13 to 18 mm). In the castrati, whose vocal brilliance had a great influence on the lyrical Italian operas, the membranous component of the vocal cords was the same length as in prepubertal boys and even shorter than in women.\[108\] During puberty, the male larynx, cricothyroid cartilage, and laryngeal muscles enlarge. The relatively largest change in singing and speaking frequencies occurs between Tanner genital stages 3 and 4,\[109\]\[110\] breaking of the voice occurs at approximately 13 years, and the adult voice is achieved by about 15 years.\[111\] The singing voice changes after the deepening of the speaking voice by one octave.\[112\]

Facial hair in boys is first apparent on the corners of the upper lip and the upper cheeks; it then spreads to the midline of the lower lip and finally to the sides and the lower border of the chin. The first stage of facial hair development usually occurs during public hair stage 3 (average age of 14.9 in the United States), and the last stage occurs after public hair stage 5 and genital stage 5.

Axillary hair appears at approximately 14 years in boys. Ninety-three percent of black American girls have axillary hair by age 12, in contrast to 68% of white girls.\[113\] Axillary sweat glands begin to function as the hair appears. The appearance of circumanal hair slightly precedes that of axillary hair in boys.

Comedones, acne, and seborrhea of the scalp appear as a result of the increased secretion of gonadal and adrenal sex steroids.\[114\] Early-onset acne correlates with the development of severe acne later in puberty. The most serious variety, acne fulminans, occurs mainly in pubertal males.\[115\] Acne vulgaris, the most prevalent skin disorder in adolescence, occurs at a mean age of 12.2 years ± 1.4 years (SD; range 9 to 15 years) in boys and progresses with advancement through puberty. However, acne vulgaris can be the first notable sign of puberty in a girl, preceding public hair and breast development.\[116\] At late prepuberty comedones are present in many boys, and 100% of boys have comedones by genital stage 5.

Dental hygiene is often worse in boys than girls after the onset of puberty; unfortunately, this is also the age when gingivitis begins to appear along with the formation of pockets at the gum line.\[117\] There is a positive correlation between sex steroid concentrations and the presence of the bacteria most responsible for the inflammation of gingivitis in boys and girls.\[118\]
Gonadal Development and Function

Female

Ovarian Development in Puberty

The peak number of germ cells in the fetal ovary is attained at 16 to 20 weeks of gestation (also see Chapter 2). Primordial follicles appear beginning at 20 weeks of fetal life, soon followed by the appearance of primary follicles; they constitute the lifelong store of follicles for the individual because no more arise. Follicle-stimulating hormone (FSH) receptors have not been detected in midtrimester human fetal ovaries; fetal pituitary FSH is not required for proliferation of oogonia, oocyte differentiation, or formation of primordial follicles. During fetal life and childhood, follicular growth to the large antral stage occurs, but before menarche all developing follicles are destined to undergo atresia (Fig. 24-8). Thus, large preovulatory follicles are rarely present before puberty. During follicular growth, the oocyte enlarges and granulosa cells are transformed from spindle-shaped to cuboidal cells that, along with the ovum, secrete the zona pellucida.

Our understanding of the mechanism of FSH and luteinizing hormone (LH) action through plasma membrane-associated receptors has advanced considerably. The granulosa cells multiply with a substantial increase in volume of the follicle, and ultimately a plasma transudate, the follicular fluid, forms in response to FSH and fills the antrum. The theca develops during the time of antrum formation. During reproductive life the follicle progresses further to luteinization or terminal differentiation into a corpus luteum, which is the major source of gonadal steroids after ovulation. After 8 days of gonadal steroid secretion in the absence of fertilization, the follicle undergoes programmed cell death or apoptosis and cytolysis results in an avascular scar. If fertilization occurs, fetal tissue, through the secretion of human chorionic gonadotropin (hCG), supports the maternal corpus luteum throughout pregnancy.

Ultrasoundographic studies (Fig. 24-9) show that the corpus of the uterus increases during pubertal progression from an initial tubular shape to a bulbous structure and that the length of the uterus increases from 2 to 3 cm to 5 to 8 cm and the volume from 0.4 to 1.6 mL to 3 to 15 mL. During prepuberty the ovarian volume is 0.2 to 1.6 mL on ultrasound scans, and after the onset of puberty the volume increases to 2.8 to 15 mL. Tall girls have a greater ovarian volume than girls of average size. An increase in the multicystic appearance of the ovaries on ultrasonography occurs with the progression through puberty and should not be considered a sign of disease.

Menarche

Menarche (see Chapter 16) usually occurs in the 6-month period preceding or following the fusion of the second and first distal phalanges and the appearance of the sesamoid bone; this corresponds to Tanner stage 4 in most cases. The 95th percentile for menarche is 14.5 years, although many textbooks define primary amenorrhea as absence of menses at 16 years of age; the reconsideration of the age of onset of female puberty should lead to a reconsideration of the age of definition of primary amenorrhea. Anovulatory cycles are common in the first years after menarche. There is a reported prevalence of 85% anovulation in the first 2 years after menarche that decreases to 20% anovulatory cycles by the fifth year; others have observed a lower number of ovulatory events shortly after menarche as well as 5 years after the event. Whereas the majority of pubertal females are infertile in terms of risk of pregnancy, a substantial number are fertile as evidenced by the prevalence of teenage pregnancy. Although the number of early teenage pregnancies is decreasing, in 2000 there were 8519 births to mothers 10 to 14 years of age.

Male

Testicular Development in Puberty

The testes are active during the prepubertal period, albeit at a lower level than during pubertal development. During pubertal development the testes increase in size, principally because of the growth of the seminiferous tubules associated with the onset of spermatogenetic activity, and mitosis of Sertoli cells and testosterone production increase (Fig. 24-10 and Table 24-8). The Sertoli cells are the major cell type in the seminiferous cords in prepuberty and early puberty, but in the adult, germ cells predominate. During progression through puberty, the Sertoli cells cease to undergo mitosis, differentiate into adult-type Sertoli cells, and form clumps replete with the testis blood vessel barrier.

Although Leydig cells can be detected in early gestation and again during the neonatal period of increased testosterone secretion, during childhood the interstitial tissue is composed principally of undifferentiated mesenchyme-type cells. With pubertal development and rising serum LH levels, adult-type Leydig cells appear. It is suggested that three phases of Leydig cell maturation associated with ages of increased testosterone production be recognized: 14 to 18 weeks of fetal life, 2 to 3 months after birth, and from puberty through adulthood. The seminal vesicle enlarges through childhood to puberty to hold 3.4 to 4.5 mL or 70% of the seminal fluid. The mean blood flow in the testes increases to adult values, as measured by Doppler sonography, in boys with a testicular volume greater than 4 cm.

Spermatogenesis

The first histologic evidence of spermatogenesis appears between ages 11 and 15 years; sperm can be detected in the first morning urine specimen at a mean chronicologic age of 13.3 years, although other studies extend this milestone up to 16 years depending on the population studied. There is a higher incidence of spermatogonia in early puberty than in late puberty, suggesting that there may be a continuous flow of sperm through the urethra in early puberty but that ejaculation is necessary for sperm to appear in the urine in late puberty. Spermatogonia probably reflects the maturation of spermatogenesis, but normospermia (sperm concentration, morphology, motility) is not present until a bone age of 17 years. Nonetheless, potential for fertility is reached before an adult phenotype is attained; spermatogonia was detected in 2 of 28 normal boys with bilateral testicular volumes of 3 mL and no other signs of puberty. Hence, spermarche (the onset of spermatogenesis) is a relatively early pubertal event that occurs.
at a mean pubic hair stage of 2.5, before the attainment of adult plasma testosterone concentrations and before peak height velocity is reached. The first conscious ejaculation occurs at a mean chronologic age of 13 ½ in normal boys and at a mean bone age of 13 ½ in boys with delayed puberty.
Adolescent Growth

Pubertal Growth Spurt

Prepubertal height and growth velocity are similar in boys and girls. The greatest growth occurs in infancy, and growth decreases to the nadir known as the minimal prespurt velocity just before the pubertal growth spurt. During puberty boys and girls experience a growth velocity greater than at any postnatal age since infancy. The pubertal growth spurt may be divided for purposes of comparison into three stages: the time of minimal growth velocity in peripuberty just before the spurt (takeoff velocity), the time of most rapid growth or peak height velocity, and the stage of decreased velocity and cessation of growth at epiphyseal fusion. Boys reach peak height velocity approximately 2 years later than girls and are taller at takeoff (Fig. 24-11 and Table 24-9); peak height velocity occurs during stage 3 to 4 of puberty in most boys (see Table 24-4, Table 24-5, and Table 24-7 and Fig. 24-6) and is completed by stage 5 in more than 95% of boys. 90,91 The pubertal growth spurt in girls occurs between stages 2 and 3 (see Table 24-2 and Fig. 24-8). 92 Boys grow a mean of 28 cm and girls grew 25 cm between takeoff and cessation of growth in a study in the United Kingdom. 93 The mean height difference of 12.5 cm between adult men and women in the Zurich growth study resulted partly from the greater prespurt growth of boys (+1.5 cm); partly from the height difference at age of takeoff, with boys being taller at their later age of takeoff than girls (+6.5 cm); partly from the greater gain in height of boys during the pubertal growth spurt (+6 cm); and partly from the greater postspurt growth in girls (-1.5 cm). 94

A mathematical model of growth, based on longitudinal data, separates the infancy, childhood, and pubertal phases of growth and allows evaluation of growth in spite of the variation in the age of the onset of puberty. A slowly decelerating childhood component is the base, with a sigmoidal pubertal component added during secondary sexual development (Fig. 24-12). This model provides a new means of predicting adult height, the height adjusted for pubertal onset (HAPO). 95 Variations on this technique of height prediction use the infancy-childhood-pubertry growth curve either (1) without bone age or (2) without bone age but with the use of parental height information when available. 96,97 Tanner and Davies 98,99 have constructed growth curves for American children using longitudinal data from the National Center for Health Statistics and calculated data from theoretical growth curves thereafter (see Chapter 23); these curves can be adjusted for time of peak height velocity and appear to have greater validity for evaluation of growth of individual children during adolescence than the standard cross-sectional charts.

A host of other physiologic and biochemical measurements change with the onset of puberty and must be interpreted in terms of the stage of pubertal development. The mean heart rate and maximal oxygen uptake do not change in boys with the passage of peak height velocity, but respiratory quotient increases at the time of peak height velocity; 100 Serum inorganic phosphate, alkaline phosphatase, serum osteocalcin (Gla protein level) and urinary pyridinoline, deoxypyridinoline, and galactosyl-hydroxylysine excretion reflect the increased osteoblastic activity and growth rate at that time in both sexes. 101,102,103,104,105,106 Serum urate rises at the end of pubertal development in average boys but obese boys experience a rise earlier in puberty. 107

Because girls reach peak height velocity about 1.3 years before menarche, there is limited growth potential after menarche; most girls grow only about 2.5 cm in height after menarche. 108

although there is a variation from 1 to as much as 7 cm. Boys have no event comparable to menarche during pubertal development to mark the amount of remaining growth; all that can be deduced by physical examination is that a boy in early puberty is likely to have significant growth left, whereas in late puberty limited growth is likely. The ages at menarche, takeoff, and peak height velocity are not good predictors of adult height because the duration of pubertal growth is the more important determinant of final height; later onset of pubertal and consequent increase in height at takeoff of the pubertal growth spurt can be balanced by a decrease in actual height achieved during peak height velocity and result in no net change in adult height. Nonetheless, extremely early onset of puberty can diminish ultimate adult stature, 109 and prolonged delay of puberty 110 can increase stature.

The age at peak height velocity and the age at initiation of puberty correlate well with the rate of passage through the stages of pubertal development in normal children. 111 Both stature and the upper/lower segment ratio, defined as the length from the top of the pubic ramus to the sole of the foot, change markedly during the peripubertal and early pubertal periods because of the elongation of the extremities. 112 As a rule, at birth the upper/lower segment ratio is 1.7, at 1 year the ratio is 1.4, and at 10 years the ratio is 1.0 in a normal healthy individual. The legs begin to grow before the trunk, although late in puberty, during the growth spurt, growth of the legs is similar to growth of the upper torso. 113 The mean upper/lower segment ratio of

<table>
<thead>
<tr>
<th>Stage</th>
<th>Germ Cells</th>
<th>Sertoli Cells</th>
<th>Leydig Cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prepubertal</td>
<td>Prespermatogenic cells present</td>
<td>Predominant cells in seminiferous cords</td>
<td>Scattered, partially differentiated cells present</td>
</tr>
<tr>
<td>Pubertal</td>
<td>Initiation of spermatogenesis</td>
<td>Increased complexity, formation of occlusive junctions</td>
<td>Fully differentiated cells appear</td>
</tr>
<tr>
<td>Adult</td>
<td>Active spermatogenesis, predominant cells</td>
<td>Individual cells associated with groups of germ cells</td>
<td>Groups of fully differentiated cells present</td>
</tr>
</tbody>
</table>

Table 24-6 – Cellular Activity in Human Testis at Different Stages of Development

white adults is 0.92, and that of black adults is 0.85. There are no differences in upper/lower segment ratio between the sexes; however, the

| TABLE 24-9 — Difference in the Relationship of Onset of Pubertal Growth to Sexual Maturation in Boys and Girls |
|-------------|-------------|-------------|
| **Girls**   | The onset of the pubertal growth spurt precedes or is associated with the earliest signs of female secondary sexual maturation (e.g., sexual hair, breast development). |
| **Boys**    | The onset of sexual maturation, including testicular enlargement and male secondary sexual characteristics, precedes the onset of the pubertal growth spurt. |
| **Peak height velocity is not achieved until a late stage of sexual maturation (stage III-IV).** |


ratio of sitting height to standing height is higher in pubertal and adult females than in males.

In general, hypogonadal patients have delayed epiphyseal fusion and lack a pubertal growth spurt; therefore, their extremities grow for a prolonged period, leading to a decreased upper/lower segment ratio and an increased span for height, a condition known as eunuchoid proportions. Eunuchoid proportions are found in subjects with defects in estrogen synthesis and estrogen receptor deficiency, but normal proportions occur in complete androgen insensitivity syndrome, demonstrating the primary role of estrogen in mitigating or establishing these proportions.

Further, the distal parts of the extremities, the hands and feet, grow before the proximal parts; a rapid increase in shoe size is a harbinger of the pubertal growth spurt. Note that boys with Klinefelter’s syndrome have long legs, but not long arms, as a physical feature found even before the onset of puberty.

The shoulders become wider in boys, whereas the hips enlarge more in girls; the ratio of bicipital (shoulder) breadth to bicipital (hip) breadth remains constant in boys at about 1.37 but decreases in girls from 1.35 to 1.27.

The female pelvic inlet widens, mainly because of the growth of the os acetabuli. The size of the brain reaches 95% of adult size by the onset of puberty (see more about brain development later).

The size of the head approaches the adult size by age 10, but changes in relationships of the parts of the face are apparent during puberty. Thus, the mandible and nose enlarge more in boys, but they and the maxilla, brow, frontal sinuses, and middle and posterior fossae enlarge in both sexes, mainly during the pubertal growth spurt. Children with isosexual precocity have the facial appearance of older children, and individuals with delayed puberty have faces of younger children. The pituitary gland enlarges more in the female, in a magnetic resonance imaging (MRI) study the height of the pituitary gland was no greater than 6 mm before puberty, was 8 to 10 mm in teen-age females and had a spherical appearance in some, was no more than 7 mm in teen-age males, and decreased in young adults of both sexes.

Hormonal Control of the Pubertal Growth Spurt

Hormonal control of the pubertal growth spurt is complex (Fig. 24-13 and Fig. 24-14). Growth hormone (GH) is clearly involved in increasing growth at puberty through the stimulation of insulin-like growth factor I (IGF-I, previously called somatomedin-C) production. Gonadal steroids have two effects on pubertal growth: (1) induction of an increase in GH secretion and thus the consequent increase in IGF-I production, thereby indirectly stimulating pubertal growth, and (2) a direct effect on cartilage and bone by stimulating local production of IGF-I, among other local factors.

In the developing human skeleton, gonadal steroids have growth-promoting and maturational actions on chondrocytes and osteoblasts among other bone constituents, the latter action, which eventually leads to epiphyseal fusion and the cessation of longitudinal growth in both boys and girls, is mediated mainly by estrogen either directly secreted (in girls) or arising from the conversion of testosterone and androstenedione to estrogen in peripheral tissues by aromatase. Individuals with a mutation in the estrogen receptor gene or the CYP19 gene encoding aromatase continue to grow, lack a pubertal growth spurt, and have open epiphyses and osteopenia. Further, estrogen treatment of men with aromatase deficiency leads to epiphyseal closure, cessation of growth, and a striking increase in bone mass. On the other hand, patients with aromatase excess, who produce excess estrogen, have advanced skeletal maturation and rapid growth and may ultimately reach short adult stature (see later).

Although estradiol secreted by the ovary has been recognized for over two decades as the major sex steroid responsible for the pubertal growth spurt, skeletal maturation, and bone mineral accrual in the female, until the detection of the rare human genetic defects in estrogen synthesis or action, conventional wisdom dictated that in the male testosterone mediated these maturational changes during puberty. In the male as well as the female, estrogen (not androgen) is the critical sex hormone in the pubertal growth spurt, skeletal maturation, the accrual of peak bone mass, and the maintenance of bone mass in the adult. Estrogen stimulates chondrogenesis in the epiphyseal growth plate, increasing pubertal linear growth. At puberty, estrogen promotes skeletal maturation and the gradual progressive closure of the epiphyseal growth plate.

Surprisingly little is known about the mechanism of action of estrogen on the skeletal growth plates (Table 24-10). During puberty and into the third decade, estrogen has an anabolic effect on the osteoblast and an apoptotic effect on the osteoclast, increasing bone mineral acquisition in the axial and appendicular skeleton. Further, in the adult, estrogen is important in maintaining the constancy of bone mass through its effect on bone remodeling and bone turnover (see Table 24-10). The evidence establishing a primary role for estrogen in the developing skeleton by no means excludes a direct action of testosterone on bone in the human male, but this action is less well characterized and quantified than thought in the past. The greater increase in periosteal bone deposition, with resultant thickening of cortical bone and greater bone density in boys, is probably related to a direct effect of testosterone. In a carefully studied group of individuals with complete androgen insensitivity (and reportedly appropriate estrogen replacement therapy), a modest decrease in bone mineral density (BMD) Z scores was noted at the spine but not hip using age-specific female standard values, but the reductions were greater against male standards and when apparent BMD (a measure of volumetric BMD) was used. These findings support the concept that lack of a direct effect of testosterone on the skeleton, especially the spine, has a part in the defects in bone mineralization in women with complete androgen insensitivity (Table 24-11)
Growth Hormone

GH and IGF-I are major hormonal and growth factors in the pubertal growth spurt. The secretion of GH approximately doubles during puberty in both boys and girls.

Increased GH pulse amplitude and the amount of GH secreted per pulse (not frequency or metabolic clearance rate) in the basal state are mainly responsible for the augmented GH levels. Pulsatile GH secretion rose through pubertal development, although the frequency of the GH pulses did not change with pubertal development, nor did the intersecretory burst interval and half-life of GH.

The increase in GH secretion occurs earlier in girls, coincident with the onset of breast development (Tanner stage 2), and is maximal at Tanner stage 3 to 4 breast development; in boys, in contrast, the increase occurs later and peaks at stage 4 genital development. GH secretion and IGF-I levels decrease from the higher pubertal values to adult values after late puberty in both sexes. Stimulated GH secretion also increases at puberty; indeed, a study of 88 subjects of normal height found that 61% of prepubertal children did not have a GH peak above 7 ng/mL after exercise, arginine, or -dopa, but by stage two of puberty 44% met the 7 ng/mL threshold.


### TABLE 24-10 -- Earlier Clinical Clues to the Effect of Estrogen on Growth and Skeletal Maturation in the Male

<table>
<thead>
<tr>
<th>Clinical Clues to the Effect of Estrogen on Growth and Skeletal Maturation in the Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aromatase inhibitor decreased rapid growth and skeletal maturation in testotoxicosis whereas antiandrogen had no effect on skeletal maturation (Laue L, Jones J, Barnes K, Cutler GB Jr. Treatment of familial male precocious puberty with spironolactone and deslorelin. J Clin Endocrinol Metab 1993; 76:151155)</td>
</tr>
<tr>
<td>Aromatase excess syndrome in boys associated with increased rate of growth and skeletal maturation, elevated plasma estrogen concentrations, but prepubertal testosterone values (Stratakis CA, Voltero A, Brodie A, et al. The aromatase excess syndrome is associated with feminization of both sexes and autosomal dominant transmission of aberrant P450 aromatase gene transcription. J Clin Endocrinol Metab 1998; 83:13481357)</td>
</tr>
</tbody>
</table>

---

**Growth Hormone**

GH and IGF-I are major hormonal and growth factors in the pubertal growth spurt. The secretion of GH approximately doubles during puberty in both boys and girls. The increase in GH pulse amplitude and the amount of GH secreted per pulse (not frequency or metabolic clearance rate) in the basal state is mainly responsible for the augmented GH levels. Pulsatile GH secretion rose through pubertal development, although the frequency of the GH pulses did not change with pubertal development, nor did the intersecretory burst interval and half-life of GH. The increase in GH secretion occurs earlier in girls, coincident with the onset of breast development (Tanner stage 2), and is maximal at Tanner stage 3 to 4 breast development; in boys, in contrast, the increase occurs later and peaks at stage 4 genital development. GH secretion and IGF-I levels decrease from the higher pubertal values to adult values after late puberty in both sexes. Stimulated GH secretion also increases at puberty; indeed, a study of 88 subjects of normal height found that 61% of prepubertal children did not have a GH peak above 7 ng/mL after exercise, arginine, or -dopa, but by stage two of puberty 44% met the 7 ng/mL threshold.

**Interactions of the major growth-promoting hormones during puberty.** Plus (+) indicates stimulatory action, minus (-) inhibitory action. Growth hormone and gonadal steroids have a direct stimulatory effect on the generation of IGF-I (paracrine action) locally in bone and cartilage cells. For simplification, the feedback loops for IGF-I and gonadal steroids on the hypothalamic-pituitary unit are omitted.


**Figure 24-14 Interactions of the major growth-promoting hormones during puberty.** Plus (+) indicates stimulatory action, minus (-) inhibitory action. Growth hormone and gonadal steroids have a direct stimulatory effect on the generation of IGF-I (paracrine action) locally in bone and cartilage cells. For simplification, the feedback loops for IGF-I and gonadal steroids on the hypothalamic-pituitary unit are omitted.

**Table 24-11 -- Some Sites of Estrogen Action on Bone**

<table>
<thead>
<tr>
<th>Sites of Estrogen Action on Bone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear growth: chondrogenesisproliferation and differentiation of growth plate chondrocytesand its link to osteogenesis</td>
</tr>
<tr>
<td>Skeletal maturation: the gradual, progressive ossification of the epiphyseal growth plate during puberty, possibly as a consequence of both estrogen-induced vascular and osteoblastic invasion and the termination of chondrogenesis</td>
</tr>
<tr>
<td>Accrual of bone mass during puberty and into the third decade</td>
</tr>
<tr>
<td>Estrogen and the constancy of bone mass in the adult: remodeling, maintenance, and repair of the osteoclast and the osteoblast and osteocyte</td>
</tr>
</tbody>
</table>


ng/mL level, by stage three 11% met the level, and at stage four and five 100% did so. Hexarelin, a six-amino-acid GH-releasing peptide, stimulated as much GH secretion in prepuberty as it did in puberty, in contrast to the responses to other secretagogues, which changed with puberty.

The increase in estradiol at puberty, arising in boys from both testicular secretion and extraglandular synthesis from testosterone and androstenedione and in girls from secretion by the ovaries, is the principal mediator of the increase in pulse amplitude and the amount of GH secreted per pulse. The effect of testosterone, however, is mediated mainly through its conversion to estradiol as estrogen blockade with tamoxifen decreases GH secretion in boys. The administration of androgens that cannot be aromatized to estrogen (e.g., oxandrolone and dihydrotestosterone) had less effect upon GH secretion, but androgen blockade with flutamide increased GH secretion.

Dihydrotestosterone, which is not aromatized to estrogen, does not increase GH secretion or the plasma concentration of IGF-I and may even decrease the integrated GH secretion, but it still stimulates an increased rate of growth, suggesting a possible direct effect of androgen on pubertal growth independent of GH or estradiol. This priming effect of estrogen on GH secretion is used in clinical practice, as estrogen administered before a provocative test in pubertal subjects increases the GH response. Increased GH secretion also occurs in sexual precocity. GH secretion decreases with the fall in gonadal steroid levels after treatment of children with true precocious puberty with potent luteinizing hormone releasing hormone (LHRH) agonists.

Patients with isolated GH deficiency or GH resistance have an attenuated pubertal growth spurt, indicating the importance of GH and IGF-I in this time of rapid growth. In a series of 26 women with apparently isolated GH deficiency who had spontaneous puberty, 39% had menarche after 16 years of age and almost 50% continued to have menstrual disorders. Individuals with severe primary or secondary hypogonadism have a minimal growth spurt or no pubertal growth spurt, demonstrating the primary role of gonadal steroids in pubertal growth. Hypopituitary patients deficient in both GH and gonadotropins do not have an adolescent growth spurt when GH alone is replaced; gonadal steroids must also be given, substantiating the interaction of GH and gonadal steroids in the pubertal growth spurt.

However, in normal puberty, neither the magnitude of the increase in GH secretion nor the concentration of plasma IGF-I correlates with peak height velocity of the pubertal growth spurt. This suggests that although a threshold level of GH secretion is a necessary component, the extent of the growth spurt correlates with gonadal sex steroid secretion. Individuals with both true precocious puberty and GH deficiency (usually a consequence of cranial radiation for a brain tumor) can exhibit a growth spurt similar to and clinically indistinguishable from that of children with true precocious puberty and normal GH secretion. After treatment with an LHRH agonist for sexual precocity, patients with GH deficiency and true precocious puberty have a significant decrease in growth velocity along with the suppression of their pubertal progression, illustrating the direct effect of gonadal steroids, principally estradiol, on the pubertal growth spurt in both males and females.
The concentration of plasma IGF-I increases during puberty to reach an earlier peak in girls than in boys and then decreases to adult levels (Fig. 24-15). Increased GH secretion at puberty induces the rise in IGF-I level, and the increase in both hormones is associated temporally with the pubertal growth spurt. The concentration of plasma IGF-I is high for chronologic age in sexual precocity and low in delayed puberty. Although the precise role of gonadal steroids in the pubertal increase in IGF-I concentration is uncertain, the major effect appears to be mediated through increased secretion of GH and circulating IGF-I with an additional effect through the gonadal steroid-induced local generation of IGF-I in cartilage and bone. Bone contains both estrogen and androgen receptors as well as the enzyme aromatase. Studies of patients with true precocious puberty before and after therapy with an LHRH agonist show elevated serum GH concentrations in the untreated state and suppressed GH concentrations for age and a decrease of plasma IGF-I concentration after therapy but values do not decrease to prepubertal values, further supporting the concept that GH is the major factor that raises circulating IGF-I levels in puberty.

A confounding factor is the relative role of hepatic-generated circulating IGF-I (endocrine role) and of locally produced IGF-I (paracrine-autocrine role) in linear growth. For example, mice with a selectively and totally deleted hepatic IGF-I gene had strikingly reduced circulating levels of IGF-I but normal postnatal body and bone growth.

**Gonadal Sex Steroids**

The adolescent growth spurt in normal girls and boys depends on both estradiol and GH. A pubertal growth spurt, which leads to a final height close to that of normal men, occurs in individuals with the complete form of androgen resistance (androgen insensitivity syndrome), a finding that supports the critical role of estrogen rather than androgen in the adolescent growth spurt in boys. In a study of 18 adult XY women with the complete androgen insensitivity syndrome, 7 (38.8%) exceeded 180.00 cm in height. The detection of estrogen resistance related to a null mutation in the gene encoding the estrogen receptor and of derangements in the CYP19 gene leading to severe cytochrome P450 aromatase deficiency has highlighted the cardinal role of estradiol but not testosterone in both boys and girls in the pubertal growth spurt, complete epiphyseal maturation, and normal skeletal proportions and mineralization. Further, a supersensitive assay for plasma estradiol in prepubertal and pubertal boys showed a high positive correlation for estradiol and growth in prepubertal boys and a positive relationship between serum estrogen concentrations and peak growth velocity, which occurred about 3 years after the onset of puberty. The concentration of serum estradiol did not correlate with that of serum GH, further implicating estrogen in the pubertal growth spurt and skeletal maturation of boys as well as girls.

Children with chronic adrenal insufficiency who are given appropriate replacement therapy have a normal pubertal growth spurt despite deficient adrenal androgen secretion, indicating a minimal impact of these adrenal androgens on normal growth at puberty. Thyroid hormone has a permissive role in the pubertal growth spurt but is a requisite for normal growth. Patients with primary hypothyroidism may not have a growth spurt even when the disorder is accompanied by sexual precocity (see "Juvenile Hypothyroidism").
Bone Age

Skeletal maturation is assessed by comparing radiographs of the hand, the knee, or the elbow with standards of maturation in a normal population. Ossification centers appear in early life, the bones mature in shape and size and develop articulation of surfaces, and ultimately the epiphyses or growth plates fuse with their shafts. Bone age, an index of physiologic maturation, does not have a well-defined relationship in normal children to the onset of puberty; it is just as variable as chronologic age. However, bone age is useful for predicting the age of menarche and in delayed puberty correlates better with the onset of secondary sexual development than does chronologic age.

There is controversy concerning the relationship between the initiation of the pubertal growth spurt and the rate of maturation of bone age. Some studies suggest that there is a strong relationship between the timing of the pubertal growth spurt and rate of skeletal maturation and that the timing and rate of growth in stature and rate of skeletal maturation are highly integrated genetic processes. Another view is that whereas peak height velocity occurs at a limited range of bone ages, the takeoff of the pubertal spurt occurs at a wider range of bone ages. In addition, bone age, height, and chronologic age can be used for the prediction of final adult height from the Bayley-Pinneau tables or by use of the Roche-Wainer-Thissen, Tanner-Whitehouse, or Walker technique.

Separate standards are used for boys and girls; skeletal maturation is more advanced in girls than in boys of the same chronologic age. For example, the bone ages of 11 years in girls and 13 years in boys (bone ages of early puberty in each sex) are equivalent stages of bone maturation by the hand-wrist method. Although there are no separate standards of bone age, black children have slightly more advanced bone ages than white children of the same chronologic age. Differences between bone age and chronologic age must exceed 2 SD (according to tables available in the respective bone age atlases) to be of biologic significance.

As commonly estimated, bone age is imprecise and a qualitative rather than a quantitative measure. The development of techniques for scanning radiographs coupled with computer analysis should increase the precision of the procedure. The estimation of skeletal age may be confounded by asymmetrical conditions. For example, patients with hemiplegia related to cerebral palsy have a less advanced bone age on the affected side than the normal side. This difference (mean of 7.3 months) is greater than the difference noted between left and right sides in normal children (less than 6 months).

Skeletal Density

Areal BMD (two-dimensional image), a function of the size of bone, increases throughout childhood but volumetric bone density (the amount of bone within the periosteal envelope) does not. During growth, the increase in bone mass is attributable to the increase in both length and diameter. The increase in BMD during the prepubertal and pubertal years is due to the increase in the size of the long bones.

Bone density increases at the spine in both sexes. The BMD at the beginning of puberty in 40 white children studied longitudinally predicted the peak bone mass at sexual maturity and appeared to predict the likelihood of osteoporosis as an adult; such data should allow identification of the girls most in need of intervention.

BMD of the total body, lumbar spine, and femoral neck measured by dual-energy x-ray absorptiometry (DXA) increased at a mean annualized rate of 0.047 g/cm² for boys and 0.039 g/cm² for girls. Data from a longitudinal total-body DXA study suggest that boys accumulate mineral at 407 g/year and girls at 322 g/year or 359 mg/day for boys and 284 mg/day for girls; thus, 26% of adult calcium is laid down during the two adolescent years of peak calcium accretion of 14 years (mean) for boys and 12.5 years for girls.

BMD approaches a peak in girls by the age of 16 years and in boys by about 17 years; the rate then decreases, reaching a plateau in the third decade of life. The difference in the timing of the peak of BMD is related to the difference in the time of peak height velocity. There is a different tempo of growth of the axial and appendicular skeleton. Before puberty the legs grow more rapidly than the trunk, but during puberty there is more trunical growth in girls. Thus, different times of exposure to disease make different portions of the skeleton subject to different pathologic conditions; limb dimensions may suffer in prepuberty, spine dimensions may be affected in early puberty, and volumetric bone mineral content may change in late pubertal conditions.

There are specific differences in BMD acquisition related to ethnicity, and standards for black, white, Asian, and Hispanic youth are being developed (Fig. 24-16). Other DXA manufacturers have provided standards for BMD by young adults but not for children and adolescents (see standard values in clinical situations. Patients may be referred for osteoporosis if their DXA results are compared with young adult values when they have not yet reached maximal bone density. In fact, bone size reaches its adult value and peak height velocity occurs before maximal bone mineral content is reached; these factors may result in a period of increased fragility and susceptibility to trauma. Although quantitative computed tomographic (CT) scanning demonstrates an increase in the cortical bone density of the lumbar spine with age, less relationship is noted between cancellous bone density and age until the later stages of puberty. The increase in BMD correlates well with height, weight, age, pubertal development, and BMI but has less relationship with serum IGF-I. Weight is a main determinant of bone density in postpubertal females.

The strength of the femoral head increases markedly during puberty, and the femoral neck increases in density even more with participation in impact loading sports such as running (compared with active load sports such as swimming) (Table 24-12). The latter table contains bone volumetric density BMD (gm/cm³), which corrects, in part, for variations in bone size. Increased weight-bearing sports or gymnastic activities increase BMD before menarche in boys and girls. Bone hip density at 18 years was increased in women who participated in daily sports between 12 and 18 years. However, in one study, bone density and geometry were not related to physical activity and calcium intake in European girls.

The expense and complexity of DXA and CT scanning led to the investigation of other methods of assessment of bone growth. Quantitative ultrasound standards are being developed that may simplify the process. In sum, bone densitometry is useful in assessing the attainment of bone mass and the risk of osteoporosis and fracture. Calcium is a critical nutrient for skeletal health throughout the life cycle, with the intake of calcium during the years of puberty having a major effect on bone density later in life in most but not all studies. and is probably dependent upon other aspects of nutrition, although there are many gaps in our knowledge in this area. Extra calcium administration may increase bone accretion and may be accomplished safely by increased dairy product intake, but the effect of increased ingestion of calcium may last only as long as the calcium is administered. Girls who drink more milk during adolescence are likely to have greater BMD as adults. Increased sodium intake at the expense of calcium intake adversely affects bone accretion, although increased calcium intake does not disturb the magnesium balance of the individual.

Black children retain more calcium than whites and the bone structure is thicker in black children: the difference in vertebral bone density appears to develop by late puberty because in prepuberty there is no difference between the groups.

Boys tend to add bone mostly on the periosteal surface, which increases bone strength, whereas girls add bone on the endocortical surface, which, rather than
increasing strength, is postulated to serve as a reservoir for calcium perhaps for later lactation and pregnancy. Testosterone administration to normal prepubertal boys increased calcium retention and bone growth, an effect mediated mainly by peripheral aromatization to estradiol, including skeletal tissue. Although bone mineral content may normally be higher in boys than girls, when corrected for the increased volumetric bone density, boys and girls have identical values for bone mineral content; bone mineral content increases with pubertal development in boys and girls.

Abnormalities of puberty impair bone accretion in both sexes, mainly as a consequence of estrogen deficiency related to either decreased secretion or peripheral aromatization of androgens. Boys with constitutional delay in adolescence were reported to have decreased areal bone density as adults; however, normal volumetric bone density was found in adults previously affected by constitutional delay or primary hypogonadism. Treatment of hypogonadotropic hypogonadism in males with testosterone raised serum osteocalcin and increased bone density, and, at least in a 6-month period, testosterone increased bone density in adolescents with constitutional delay. Individuals with Klinefelter's syndrome and decreased testosterone concentrations have decreased bone density; for selected patients, testosterone replacement has been recommended in puberty to avoid this outcome. Further, loss of bone density occurs in girls with anorexia nervosa, hypothalamic amenorrhea, or ovarian failure. Children with true precocious puberty have increased bone density, but successful treatment with an LHRH agonist decreases bone density.

There is a correlation between the bone density of children and parents with osteoporosis, demonstrating the importance of genetics in the pubertal period of bone accretion; indeed, there is a relationship of bone density between generations if the effects of age and puberty are eliminated. In fact 60% to 80% of variance in peak bone mass has been attributed to genetic factors.

Bone turnover is reflected by changes in biochemical markers. The most significant indicators of bone turnover are bone-specific alkaline phosphatase, osteocalcin, and urinary deoxypyridinoline, which reach a peak at midpuberty and decrease thereafter. Lesser changes are reflected in concentrations of carboxy-terminal pyridinoline cross-linked telopeptide, immunoreactive urinary pyridinolines, and urinary galactosyl hydroxylysine. Estrogen was negatively correlated with bone turnover, suggesting that estrogen is responsible for the decrease in bone turnover in late puberty.
Body Composition

Striking changes in body composition and energy requirement occur during puberty along with the increase in gonadal hormone levels and maturation of secondary sexual characteristics. Lean body mass, skeletal mass, and body fat are equal in prepubertal boys and girls, but by maturity men have 1.5 times the lean body mass and almost 1.5 times the skeletal mass of women, whereas women have twice as much body fat as men. The increase in lean body mass starts at 6 years in girls and 9.5 years in boys and is the earliest change in body composition at puberty. A "strength spurt" occurs during puberty after the pubertal growth spurt. The discrepancy in adult strength between men and women is due partly to the fact that men have more muscle cells and partly to the greater size of individual muscle cells; the muscle mass is 54% of body weight in adolescent boys and 42% of body weight in adolescent girls.

Weight is not necessarily a reflection of body fat; the BMI (calculated as weight in kilograms/height in square meters) is often invoked in describing the shape of the body in age-adjusted terms BMI changes with age, and there is no specific number indicating normal or abnormal BMI at all stages of development. Charts of BMI versus age between the 3rd and 97th percentiles are available at www.CDC.gov. Indeed, BMI is not as useful a reflection of body fat in growing children as in adults, but for consistency BMI is used at all ages. The use of weight/height or BMI was proposed as an index of obesity unrelated to age or height, which simplifies the assessment of adiposity; values of 12.74 in boys and 12.45 in girls before stage 4 pubertal development are said to be the mean values for normal individuals but this method is not in general use. Another quantitative measure of body fat is triceps skinfold.

Table 24-12 -- Summary of Clinical Characteristics by Age for Figure 24-16

<table>
<thead>
<tr>
<th>Age</th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td>911 yr (n = 48)</td>
<td>1113 yr (n = 75)</td>
<td>1315 yr (n = 78)</td>
</tr>
<tr>
<td>Tanner stage</td>
<td>N %</td>
<td>N %</td>
</tr>
<tr>
<td>Pre-early puberty</td>
<td>39 81</td>
<td>42 56</td>
</tr>
<tr>
<td>Midpuberty</td>
<td>9 19</td>
<td>33 44</td>
</tr>
<tr>
<td>Maturity</td>
<td>0 0</td>
<td>0 0</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>139.9 ± 8.5</td>
<td>149.4 ± 8.7</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>36.4 ± 8.7</td>
<td>43.9 ± 9.5</td>
</tr>
<tr>
<td>L24 BMD (g/cm²)</td>
<td>0.672 ± 0.066</td>
<td>0.720 ± 0.082</td>
</tr>
<tr>
<td>L24 BMAD (g/cm²)</td>
<td>0.125 ± 0.011</td>
<td>0.124 ± 0.012</td>
</tr>
<tr>
<td>TH BMD (g/cm²)</td>
<td>0.778 ± 0.080</td>
<td>0.838 ± 0.087</td>
</tr>
<tr>
<td>FN BMD (g/cm²)</td>
<td>0.737 ± 0.070</td>
<td>0.784 ± 0.083</td>
</tr>
<tr>
<td>FN BMAD (g/cm²)</td>
<td>0.172 ± 0.023</td>
<td>0.169 ± 0.019</td>
</tr>
<tr>
<td>WB BMD (g/cm²)</td>
<td>0.812 ± 0.053</td>
<td>0.869 ± 0.059</td>
</tr>
<tr>
<td>WB BMC (g)</td>
<td>1058 ± 264</td>
<td>1350 ± 296</td>
</tr>
<tr>
<td>WB BMC/h (g/cm)</td>
<td>7.50 ± 1.51</td>
<td>9.00 ± 1.62</td>
</tr>
</tbody>
</table>

Figure 24-16b. 5. Spine BMD for males by age. The curve for black males was significantly greater than the mean levels of all nonblacks; Asian and white males had greater mean spine BMD than Hispanics. Mean and SD curves are shown. (A and B, From Bachrach UK, Hastle T, Wang R, et al. Bone mineral acquisition in healthy Asian, Hispanic, black and caucasian youth: A longitudinal study. J Clin Endocrinol Metab 1999; 84:4707.)
which requires specialized training to achieve accuracy. DXA is used to determine the percentage of body fat, water, and bone with great accuracy but cannot differentiate visceral from subcutaneous or other adipose tissue. CT scanning was necessary to determine the distribution of fat in different body locations until the validation of MRI as a method to determine the subcutaneous and subcutaneous fat distribution without the use of radiation. Equations were developed to utilize anthropomorphic techniques to determine intra-abdominal adipose tissue (IAAT) and subcutaneous abdominal adipose tissue (SAAT). Studies of populations aiming to relate body fat to adverse outcomes may use BMI, but the results may not be comparable with the outcome if a direct measurement of IAAT or calculations directly reflecting IAAT are employed. Thus, a clear statement of methods is needed in any study before conclusions are validated.

It is the visceral adipose tissue that predisposes to metabolic complications of obesity, and the subcutaneous adipose tissue that leads to these different body forms is only an imperfect reflection of this internal distribution of fat cells. The IAAT is the most metabolically active fat, which can quickly undergo lipolysis. IAAT comprises visceral adipose tissue (metabolic products drain immediately into the portal system and into the liver) and retroperitoneal adipose tissue, whereas subcutaneous fat, the visible fat on physical examination, is spread throughout the body. Increased visceral fat is associated with hypertension, insulin resistance, increased high-density lipoprotein (HDL) cholesterol, and small very-low-density lipoprotein (VLDL) particles that are cholesterol laden and most likely to be converted to low-density lipoprotein (LDL); subcutaneous fat is associated with large, lipid-laden VLDL particles that are removed directly from the circulation and pose less risk. Thus, it is the visceral fat that has been implicated in metabolic derangements in adults, for example, dyslipidemia, hyperinsulinemia, and cardiovascular risk factors. Studies support the role of increased intra-abdominal fat in children as a cause of insulin resistance and dyslipidemia. Whereas waist/hip ratios are used as a substitute for the direct measurement of visceral adipose tissue in adults, waist/hip ratios do not reflect intra-abdominal adipose tissue in children and adolescents. Further, increased IAAT may cause these metabolic derangements without increased total body fat.

Hips enlarge with pubertal development in girls, but there is no change in waist circumference and the waist/hip ratio normally drops; the waist/hip ratios of various ethnic groups develop differently. Further, girls at breast stage 2 with predominant fat on the hips had higher gonadal steroid and gonadotropin concentrations, and girls with predominant fat in the abdomen had lower androgen/testosterone ratios, probably because of increased aromatization of androgens to estrogens in adipose tissue.

The generalized distribution of fat in males (central fat or apple shaped; android) is different from that in females (lower body fat predominance or pear shaped; gynecoid). Some evidence suggests a relationship between fat distribution and testosterone-binding globulin (TeBG; sex hormone-binding globulin).

With the change in fat composition there is also a change in body water; body water increases 5% in men and decreases 5% in women. Whereas extracellular water is about 25% of body weight in boys and girls, intracellular water increases at puberty in boys from 36% to 39% and decreases in girls from 36% to 29%.

Frisch and Revelle pointed out that late-maturing girls gain fat more slowly; they related menarche and the maintenance of menstrual function to the percentage and absolute amount of body fat. By 9 to 10 years of age, black females tend to have a higher caloric intake with more fat ingested and less physical activity than white girls; by 9 to 10 years there is a higher percentage of obesity in black girls than in white girls. Black and white girls have an equal prevalence of dieting, but black girls more frequently attempt to gain weight, usually under parental influence, possibly related to the average 20 pounds greater weight of their mothers. Black girls have lower resting energy expenditure and total energy expenditure than white girls. This relationship does not change with pubertal development if the data are normalized for body composition. These data imply that lower caloric intake or increased activity is necessary for age- and pubertal stagedmatched black girls compared with white girls if weight is to remain stable.

The surgeon general has stated that there is an epidemic of obesity in children and adults in the United States. Between the last two NHANES studies (II to III), roughly in the last 20 years, there was a doubling of the prevalence of individuals over the 95th percentile in childhood and adolescence and a 50% increase in those above the 85th percentile. Excessive body fat has significant medical effects during childhood as well as later in life as the individual matures. However, medical complications occur with excessive body weight as well as can occur with inadequate dietary intake.

Dieting, defined as limiting caloric intake in order to lose weight, is not limited to adults or teenagers. In the third to the fifth grade, girls are already expressing dissatisfaction with their weight. At 10 years of age there is already a demonstrable difference in concern about eating and weight gain in girls compared with lack of such concern in boys. By fifth grade, 31% of girls are dieting, and by sixth grade, 62% are dieting. There is either no difference in the prevalence of dieting between white and black girls or an increased prevalence in whites, depending on the study, although there is evidence that black girls more often try to gain weight, apparently because of parental suggestion that they are too thin.

In Holland, a 40-year follow-up demonstrated increased mortality from ischemic heart disease of 2.3-fold increase in cardiovascular mortality. A study extending up to 52 years of children between 5 and 15 years in Maryland demonstrated a positive relationship between increasing relative weight and adult mortality for prepubertal boys and girls and postpubertal girls. In Sweden, a 40-year follow-up of hospitalized adolescents revealed a higher than expected mortality.

Countries that obesity in childhood is associated with increased risk factors for cardiac disease. Many workers recommend prevention of obesity as a method of improving cardiac risk factors or decrease in body fat if prevention has failed. A BMI above the 85th percentile was associated with one cardiac risk factor (dyslipidemia, hypertension, or elevated insulin) in 58% of the 813 overweight 5- to 17-year-olds surveyed in the Bogalusa Heart Study. Odds ratios for other associated with 2.4 (diastolic blood pressure), 3.0 (LDL cholesterol), 3.4 (HDL cholesterol), 4.5 (systolic blood pressure), 7.1 (triglycerides), and 12.6 (fasting insulin) for those with elevated BMI.

Countries that obesity in childhood is associated with increased risk factors for cardiac disease. Many workers recommend prevention of obesity as a method of improving cardiac risk factors or decrease in body fat if prevention has failed. A BMI above the 85th percentile was associated with one cardiac risk factor (dyslipidemia, hypertension, or elevated insulin) in 58% of the 813 overweight 5- to 17-year-olds surveyed in the Bogalusa Heart Study. Odds ratios for other associated with 2.4 (diastolic blood pressure), 3.0 (LDL cholesterol), 3.4 (HDL cholesterol), 4.5 (systolic blood pressure), 7.1 (triglycerides), and 12.6 (fasting insulin) for those with elevated BMI.
other studies in Finland, serum insulin correlated positively with BMI, concentrations of serum triglycerides, and blood pressure and inversely with the concentration of HDL cholesterol in 3- to 18-year-olds. High triglycerides, high systolic blood pressure, and low level of HDL cholesterol clustered among subjects within the highest insulin values. The hemostatic risk factors for adults, factor VIIc, von Willebrand factor antigen, and tissue-type plasminogen activator antigen, are related to plasma insulin and triglyceride concentrations in obese teenagers; plasminogen activator inhibitor-1 antigen correlated with fat mass in boys and girls; and factor VIIc correlated with fat mass in girls in an initial study.

Macro or micro evidence of arteriosclerosis is already found in youth and the tendency is increased by obesity. By 15 to 19 years of age, 2% of autopsied males in one study had advanced (American Heart Association grade 4 or 5) atherosclerotic coronary artery lesions associated with increased serum cholesterol, obesity, and hypertension. Thus, lesions with a vulnerability to rupture are present at a young age and related to coronary risk factors.

In several longitudinal studies, unfavorable lipid levels were tracked from childhood to adulthood; elevated cholesterol in adolescents was related to obesity, and adult values followed the values found in teenagers. Individuals tested during childhood or adolescence for the Muscatine Heart Study were retested later in adulthood in a longitudinal manner. On average, of children found to have cholesterol levels greater than or equal to the 90th percentile for their age and gender, on a single measurement 43% remained above the 90th percentile, 62% remained above the 75th percentile, and 81% remained in later adult life. Those between 5 and 23 years from the Bogalusa Heart Study who demonstrated elevated insulin values also had higher BMI, serum triglycerides, LDL cholesterol, VLDL cholesterol, and glucose and lower HDL cholesterol but higher systolic and diastolic blood pressure. There was a clustering of various components of syndrome X (dyslipidemia and hyperinsulinemia) that intensified as BMI increased in a subset of children studied longitudinally in the population. Of obese children, 39% had no risk factors, 34% had one, 16% had two, and 11% had three or more. A longitudinal study of black young adults found that BMI after birth and at age 7 years was correlated with later increased BMI and insulin resistance.
Blood Pressure

Although increased blood pressure may be demonstrated with pubertal maturation, this effect seems most related to the increase in BMI with development. Blood pressure is dependent upon BMI as well as height, factors that are interrelated. As teenagers are becoming more obese, there is an increasing prevalence of elevated blood pressure in adolescence. Blood pressure is higher at certain chronologic ages in black adolescents than in white subjects, and this difference has been attributed to the earlier pubertal maturation in black children. However, blood pressure rises in black children at lower BMI values than in white children, making the problem of hypertension worse in the black population. In sexual precocity, blood pressure rises above prepubertal levels to values commensurate with body size and BMI.

Blood pressure is elevated more frequently in obese than in lean children, and the effect is more severe if there is an abdominal distribution of fat. Of children with elevated blood pressure, 60% had 120% of the ideal body weight. Conversely, 20% to 30% of obese children have elevated blood pressure, and obese children have an 8.5- to 10-fold increased risk of hypertension as adults.

Blood pressure in adolescence is quite responsive to stress, and this lability may make it difficult to determine a true resting blood pressure in a brief clinic visit. Thus, multiple determinations are needed while the individual is quietly resting for optimal accuracy. With the increase in obesity in the population, determinations of blood pressure must be made with appropriately large cuffs; many obese teenagers must have blood pressure measured with a thigh cuff rather than the standard adult cuff. The bladder of the blood pressure cuff must be equal to or greater than the circumference of the arm midway between the olecranon and the acromial process. Blood pressure must be related to the height of the child using appropriate standards (found, for example, in the Harriet Lane Handbook of Pediatrics and Voors and colleagues). There may be a change in the blood pressure values throughout the day with puberty. A circadian rhythm of blood pressure with values lowest at 11 PM to 5:30 AM was present more often in mid-pubertal individuals (80% to 90%) than in early or prepubertal subjects (50%), suggesting the development of a circadian pacemaker during puberty.

Blood pressure in childhood and adolescence is predictive of adult blood pressure. Indeed, blood pressure tends to track with age so that a moderate elevation of blood pressure in puberty may lead to significant elevations as the individual grows older.

Other cardiovascular complications of obesity in adolescence include increased heart rate and cardiac output, possible risk for sudden death related to ventricular arrhythmias, sequelae of severe obstructive sleep apnea or prolonged QTc interval, slipped capital femoral epiphysis, Blount’s disease and flat feet, obstructive sleep apnea, CO₂ retention, hypoxia, right ventricular hypertrophy and failure associated with prolongation of the QT interval, and the obesity hypoventilation syndrome or pickwickian syndrome (characterized by hypoventilation, somnolence, CO₂ retention, hypoxia, polycythemia, right ventricular hypertrophy and failure, pulmonary embolism, and even sudden death). Although the heart enlarges in lean children during puberty, it does so to a lesser extent; during adolescence in the absence of obesity, fat-free mass is an important determinant of heart size, including left ventricular mass, and heart growth.
Brain anatomy and function change substantially during late childhood and adolescence. A reduction in cortical synaptic density and neuronal density, analogous to programmed cell death, occurs between 2 and 16 years. There is an increase in cortical metabolic rate that is higher in infancy but in late childhood declines to adult levels; this decline ceases by the end of the second decade. Examples of the reduction in brain plasticity during puberty are the inability to learn to speak a foreign language without an accent after puberty and the ability of a child to recover completely from a CNS injury that in the adult might lead to aphasia. Puberty is also the time of appearance of the ability to solve complex problems in a mature manner. Mania, depression, and schizophrenia are more common after puberty, possibly because of defects in these normal changes in brain architecture and function of puberty.

An increasing complexity of brain function during puberty is reflected in changes in electroencephalographic patterns. An increase in delta waves per minute and an increase in the amplitude of delta waves (0 to 3 Hz electroencephalogram waves) are found during deep sleep. The function of deep sleep (slow wave or nonrapid eye movement sleep) is thought to be restorative of the functions during the awake stage, such as learning, and the most restorative portion is high-amplitude delta wave sleep. During adolescence the time spent in deep (stage 4) sleep declines by 50%. Further, there is a 50% to 75% decrease in the amplitude of delta waves during sleep across childhood and adolescence.

Primary reading epilepsy, juvenile absence epilepsy, juvenile myoclonic epilepsy, and epilepsy with grand mal on awakening increase in prevalence during puberty. On the other hand, benign epilepsy with centrotemporal (rolandic) spikes, the most common idiopathic epilepsy, often goes into remission at puberty. Aminobutyric acid (GABA) is an inhibitory neurotransmitter thought to withhold the onset of puberty by decreasing gonadotropin-releasing hormone (GnRH) secretion. However, the use of GABAergic drugs in clinical treatment of epilepsy in children does not appear to cause a delay of the onset or progression of puberty; this may simply be a result of inadequate data or may be an indication that disruption of GABA by the oral route cannot interfere with the onset or progression of puberty. In spite of the lack of association of a change in pubertal development with exogenous GABAergic agents, a shift in endogenous GABA at this stage may be part of the explanation for the change in incidence of epilepsy at puberty.

A longitudinal study of the biologic aspects of childhood-onset schizophrenia reported changes in affected children's MRI and positron emission tomographic scans of the CNS. The increased onset of schizophrenia at the age of puberty may have an anatomic explanation related to the prefrontal cortex, an area implicated in the etiology of schizophrenia. Studies of changes in the dorsolateral prefrontal cortex in monkeys during puberty suggest a (speculative) relationship with the histologic study of the brains of human beings with schizophrenia. It is postulated that the remodeling of this area is particularly vulnerable to adverse influence at the time of puberty and that these influences may be implicated in the pathophysiology of schizophrenia.
NORMAL PUBERTAL BEHAVIOR AND PATHOLOGY IN PUBERTY

Although the attainment of an adult role in society occurs within a few years of achievement of reproductive maturity in non-westernized societies, the more technologically advanced society, the more protracted the time society allows for adolescent psychosocial development. The prolonged current period of the adolescent role in society, ranging from the age of 11 years to 20 years in America, arose recently in human history, dating to no more than the last 100 years in Western society.

The most important psychological and psychosocial changes in adolescence are the emergence of abstract thinking, the growing ability of absorbing the perspectives or viewpoints of others, an increased ability of introspection, the development of personal and sexual identity, the establishment of a system of values increasing autonomy from family and personal independence, greater importance of peer relationships of sometimes sub cultural quality, and the emergence of skills and coping strategies to overcome problems and crises. Adolescence may normally be divided into three periods by chronologic age: early, middle, and late adolescence. However, these periods may be reached at different chronological ages because rates of physiologic maturation differ in individuals within these age groups. Early adolescence, ages 11 to 15 years, is the period encompassing most of the biologic changes of puberty outlined earlier. Early adolescence includes a profound social change from the sheltered, single-classroom environment of elementary school to the multiple classrooms and multiple teachers of junior high school. There is exposure to new peers, often with different life experiences and behavior patterns. In contrast to the concrete reasoning of childhood, the individual develops maturing, but not mature, abstract thought and decision-making processes.

Middle adolescence, ages 15 to 17 years, the period of the high school years, is a calmer period than early adolescence; the school experience does not undergo a striking change, and many of the most prominent biologic and physical changes of puberty are past. This is a period of partial independence. There is acceptance of some increased autonomy (as reflected in society's acceptance of drivers' permits and licenses at these ages), but the individual still lives at home. The individual moves away from the family emotionally and is less influenced by his or her peer group than are early adolescent individuals; friendships play an increasingly important role.

Late adolescence starts at the senior year of high school and is the age of acceptance of adult roles in work, family, and community. If the individual attends college, this stage is prolonged.

Behavior and Normal Puberty

Almost 100 years ago, Hall, without using what would be considered contemporary research techniques, characterized the maturing child as experiencing "Sturm und Drang" (storm and stress), which is normally restrained by cultural influences. Many depictions of adolescent turmoil followed, continuing to this day in the popular media as well as in clinical treatises. Contrary to this view, most later empirical studies describe adolescent development as a continuous, adaptive phase of emotional growth characterized more by stability than disorder and by harmonious relationships between generations rather than conflict. Although mood changes are normally more rapid and marked in the teenage years than in adults (occurring over hours or days), these shifts must be differentiated from long-standing mood and behavioral changes of serious psychopathology. Thus, turmoil or truly tumultuous behavior in adolescence is not a normal phase but may reflect actual psychopathology that requires diagnosis and treatment.

In a longitudinal study of 320 normal first-year U.S. high school students observed for 4 years and of 64 observed for 8 years, 25% experienced "continuous growth" characterized by smooth, well-adjusted functioning in spite of stressful situations; 34% experienced "surgent growth" demonstrating good adaptation in general and short periods of difficulty and distress after some stressful situations; and 21% were judged to be in "turmoil" characterized by mood swings, anxiety, and depression. Thus, 79% had successful adaptive development and 21% did not, and the latter mainly came from homes characterized by conflict, familial mental illness, and socioeconomic distress. Another study of adolescent psychopathology demonstrated that many with adolescent turmoil "did not grow out of it" when studied 5 years later as they had eventual diagnoses of unipolar and bipolar depressive disorders. It may be concluded that 80% to 90% of adolescents do well psychologically during puberty and are happy individuals, but 10% to 20% have significant difficulties.
Mood and Self-Image in Puberty

Young girls at the beginning of puberty frequently exhibit a negative self-image. Body Image and Adolescent Adjustment questionnaire scores rise and positive body image, positive peer relationships, and superior adjustment improvement are noted with breast development. Mood in adolescence is not closely related to stage of puberty; for example, one study found no significant mood or behavioral changes as a function of pubertal stages in girls aged 10.6 to 13.3 years, controlling for age effects, except for a decrease in interest in sports with progression of stage. However, a significant curvilinear trend for depressive affect (increase, then decrease; \( P < .01 \)), impulse control (decrease, then increase; \( P < .04 \)), and psychopathology (increase, then decrease; \( P < .03 \)) scales emerged after grouping the questionnaire results by four levels of serum estradiol, indicating that these measures of psychological status undergo significant changes during times of rapid increases in hormone levels. Such data suggest that hormonal changes may be more important than physical changes as determinants of certain mood and behavior patterns at adolescence.

Change in mood during the menstrual cycle is frequently described but rarely exhaustively studied. Anecdotal evidence noted during a study of urinary hormone changes during the menstrual period indicated that mothers felt that their daughters changed from being "irrational and difficult" to "more reasonable and responsible adults" with the change from anovulatory cycles to ovulatory cycles.

Depression in Puberty

Reports of attempted suicide increase sharply during puberty, and suicide now ranks fourth as a cause of death among 15- to 19-year-olds. A retrospective analysis showed that adolescents who actually committed suicide during puberty had the onset of their depression in childhood or early puberty, even though the act of suicide occurred later in puberty.

Prepubertal boys and girls demonstrate an equal frequency of depression, although there is a more frequent occurrence in girls by midpuberty with a sharp demarcation at stage 3. This change in the prevalence of depression appears more related to serum sex steroid concentrations than to LH or FSH values or the physical changes of puberty.

Changes in the stress response with pubertal development (characterized by increased corticotropin-releasing factor leading to increased corticotropin, causing elevated cortisol secretion) are implicated in the development of various psychopathologic conditions in childhood depression. It is postulated that periods of biologic transition, such as puberty, are times of increased psychological vulnerability in which depression is manifest because of a malfunction of the stress response.

An estimated 1.7% to 5.5% of adolescents have seasonal affective disorders. Children with seasonal affective disorders displayed dysregulated circadian activity rhythms comparable to those reported in depressed children not related to seasonal depression but different from those observed in adults.

Schizophrenia in Puberty

Childhood-onset schizophrenia is rare, but the prevalence rises with the onset of puberty. More boys than girls are diagnosed with isolated (but not familial) schizophrenia that begins in the pubertal years; in contrast, adult-onset schizophrenia is more common in women than men. It was suggested that a protective effect of estrogen raises the threshold for schizophrenia in women until the age of menopause, at which time more severe cases occur in women than in men of the same age.

There are other psychological conditions that usually appear first during puberty, such as panic attack and migraine headaches.

Risk-Taking Behavior

Adolescents who function at lower levels of cognitive complexity or concrete thinking and have an early onset of puberty demonstrate an increase in risk-taking behavior. The age of onset of cigarette smoking and alcohol use is proportional to the age of onset of puberty in girls; earlier maturing girls partake earlier, and boys may follow the same pattern.
Sleep Patterns in Puberty

Without the pressure of work, school, and so forth, adolescents would stay up later and awaken hours later than a normal weekday schedule would dictate or than occurred in the schedule they followed at a younger age. Daytime sleepiness is prevalent in puberty. Older people have earlier waking times and rate themselves as more morning-like than young adults. This change to eveningness from morningness appears related to biologic in contrast to social factors, which were previously thought to be more important. However, there are confounding effects in the study of circadian rhythms in teenagers.

Many factors affect sleep schedules, such as parental involvement, obligations of peers and work, and the frequent occurrence of insufficient sleep. Thus, conclusions of studies in sleep laboratories are difficult to interpret. An initial attempt to study this phenomenon demonstrated a significant relationship between the morning rise in melatonin (offset phase of melatonin secretion) and advancing sexual maturation stage and a tendency to a later midpoint phase of melatonin secretion (between the nocturnal rise and morning fall in melatonin values) with advancing sexual maturation stage. These data support a biologic process to explain the shift in sleep patterns rather than a social process.
Sexuality in Puberty

Early and middle adolescence is the period of introduction to sexuality for many but not for a majority of teenagers. There was an increase in sexual intercourse in urban teenage girls between 1971 and 1981 of over 50% in white girls, with 17.3% of white 15-year-olds and 23.2% of black girls reporting sexual intercourse at 15 years and 39.5% of white and 46.7% of black girls reporting sexual intercourse by 17 years of age. The mean age of first intercourse was 17.5 years for white males, 15.5 years for black males, 17 years for Hispanic males, 18.5 years for white females, 17.5 years for black females, and 18.5 years for Hispanic females in the last decade. Twenty-four percent of U.S. boys and 27% of U.S. girls reported sexual activity by 15 years of age in 1996 according to the Alan Guttmacher Institute (http://www.agiusa.org/), with a rise to 39% of 16-year-old girls and 45% of 16-year-old boys. By 17 years, the totals rise above 50% with 52% of girls and 59% of boys reporting intercourse.

Fertility is reached before the adult phenotype is acquired. One million adolescents become pregnant in the United States each year, a rate of 110 per 100,000. However, there has been a decrease in pregnancies during the last 20 years in adolescents: a 27% decrease in pregnancies in black 15- to 17-year-olds (99.5 to 72.9 per 1000) and a 6% decrease in whites (25.7 to 24.1 per 1000). There is a continuing decline in teenage pregnancies, although the problem is still substantial. In 2000 there were 8519 births to mothers aged 10 to 14 years, the lowest value since 1966, a rate of 0.6 per 1000. There has been an increase in abortions among teenagers during the last decade.

Testosterone is commonly thought to be a stimulator of sexuality, and in laboratory animals this may be proved experimentally. Results of the study of pubertal human subjects, however, are more complex. Sexuality appears to be correlated with testosterone production in boys in some studies, but in others it appears to be modified by the social effects of pubertal maturation. Correlation of salivary testosterone determinations every month over 3 years with sexual activity revealed that rising testosterone is associated with sexual activity (coital and noncoital). In addition, rising testosterone was related to sexuality and falling salivary testosterone to decreased sexual activity when pubertal status was controlled. The hormonal influences on female sexual behavior remain elusive; the striking increase in testosterone at puberty in boys stimulates their sexual interest more than the modest increase in testosterone in females. Also, social pressures are more mixed in their messages to girls, both encouraging sexuality and restricting it in a way more disparate than encountered by boys. Surely, the earlier onset of puberty today compared with previous centuries has had a profound effect on societal norms of sexual behavior.

Longitudinal and cross-sectional studies demonstrate an effect of religious activity on the onset of coitus regardless of pubertal development or endocrine status; boys who attended religious services regularly had a lower likelihood of progression to coitus or even to more substantial sexual ideation than those who did not attend regularly. Boys with higher free testosterone levels at study entry who never or infrequently attended religious services were the most sexually active and had the most permissive attitudes. The conclusion of the study was that endogenous testosterone production is directly related to sexual behavior in boys but this relationship may be modified by religious exposure.

A study of normal puberty in girls in the 8th to 10th grades demonstrated that follicular phase levels of testosterone were associated with increased frequency of thinking about sex and masturbation. Although this study demonstrated a relationship between hormones and female sexuality, peer influences exerted a strong effect on masturbation and the progression of sexuality and transition to first coitus. Seventh and eighth grade girls were observed over 2 years to determine the relationship between pubertal hormone levels and the first intercourse in a longitudinal study. The expected relationship between testosterone and pubertal development was found, and there was a relationship between a rise in testosterone and transition to coitus and between a fall in testosterone and less likelihood of transition to coitus (similar to the results described in boys). Further, when white girls who attended religious services regularly were considered, the relationship with testosterone became nonsignificant; the relationship remained significant among girls who didn’t regularly attend religious services were considered. Thus, it appeared that rising testosterone increased the likelihood of transition to first coitus but this tendency could be opposed by religious exposure as a reflection of social pressure in white girls. Pubertal developmental stage itself did not predict coitus, suggesting that physical maturity was not the motivating feature.

Administration of exogenous sex steroids was related to sexual behavior and mood. A randomized, double-blind, placebo-controlled, crossover clinical trial involving 39 boys and 16 girls with delayed puberty evaluated the effects of administration of oral conjugated estrogen to girls and testosterone enanthate to boys at three dose levels that were intended to simulate early, middle, and late pubertal levels. A significant effect of the administration of testosterone to boys was reflected in increased nocturnal emission and touching behaviors at the middle and high doses. However, no other treatment effects on sexual behaviors or responses were seen in boys. Girls demonstrated a significant increase in “necking” related to the administration of estrogen only at the late pubertal dose. No other treatment effects on sexual behaviors or responses were seen in girls. Thus, the administration of physiologic (rather than higher) doses of sex steroids to boys or girls with delayed puberty had few effects on sexual behaviors and responses (it is likely to be more difficult to find progression to sexual activity in delayed puberty). Because the subjects included patients with Turner’s syndrome, gonadotropin deficiency, and constitutional delay in puberty, the results cannot be directed to a single clinical disorder. Further, a 3-month period may be considered too short to allow the study of the evolution of sexuality. Nonetheless, this model is a promising method of determining the effects of hormones on behavior in a prospective, intervention, and ethical manner and may be expected to yield further interesting results in the future. It may be concluded that exogenous testosterone administered in the short term to reflect physiologic levels has no significant effects on boys and girls.

A relationship between intelligence and sexuality emerged from the analysis of data from the National Longitudinal Study of Adolescent Health (Add Health), which includes approximately 12,000 adolescents enrolled in the 7th to 12th grades. The Biosocial Factors in Adolescent Development projects observed approximately 100 white males and 200 black and white females from this larger data bank over 3- and 2-year periods, respectively. On the basis of the Peabody Picture Vocabulary Test as a measure of intelligence and confidential self-reports of sexual activity, logistic regression models were used to determine relationships between the two measures and proportional hazard models were used to examine the timing of initiation of noncoital and coital activities as a function of intelligence. After controlling for age, physical maturity, and mother’s education, a significant curvilinear relationship was found between intelligence and coital status; adolescents at the upper and lower ends of the intelligence distribution were less likely to have sex. Higher intelligence was also associated with postponement of the initiation of the full range of partnered sexual activities.
Behavior in Variations of the Normal Age of Onset of Puberty

The timing of the onset of puberty may have an effect on psychosocial development and function in puberty. With regard to children who progress through puberty at the normal limits of pubertal development described earlier, early maturing girls and late maturing boys have the greatest prevalence of adjustment reactions in puberty and thereafter.\[454\]

In general, early developing boys are perceived to be more mature, are given more leadership roles, and are accepted as more attractive and smart by their peers. Late developing boys are more insecure, more susceptible to lower levels of self-esteem and body image, \[455\] and more vulnerable to peer pressure, especially in working class and minority groups. Much of the problem of late maturation is said to focus on the decreased height of the individual rather than the lack of sexual development.\[456\] Social maturation is said to lag even after androgen treatment in severe constitutional delay in puberty according to an older study.\[457\]

In contrast to early maturing boys, early maturing girls tend to experience more difficulty, especially in the junior high school setting, where they may attract the attention of older, more mature boys. Girls with early puberty had a higher prevalence of internalizing symptoms and even internalizing disorders.\[458\] Early puberty may lead to a negative body image in girls, whereas in boys the effect is positive. Early pubertal maturation in girls may be related to a small IQ advantage over late maturing girls, but there is no support for differences in specific areas of cognitive abilities.\[459\] Late maturing girls are often more comfortable, remain with the support of their families longer, and are less often brought to medical attention than late maturing boys.\[460\]

Delay in menarche related to stress or decreased weight may be further analyzed with respect to dating behavior. Menarche was delayed in lean ballet dancers compared with a control group of adolescent girls of the same age and compared with their own mothers.\[461\] The dancers studied were less likely to date than age-matched peers, and whereas menarcheal status had no effect on dating behavior in nondance control subjects, the dancers who did achieve menarche were more likely to date.\[462\]
HORMONAL AND METABOLIC CHANGES IN PUBERTY

Puberty is a stage in a continuum extending from sexual differentiation and the ontogeny of the hypothalamic-pituitary-gonadal apparatus in the fetus to the completion of sexual maturation. This process involves changes in the CNS and increased frequency and amplitude of LHRH secretion at puberty, which initiates and regulates the sequential increases in the secretion of pituitary gonadotropins and gonadal steroids that culminate in sexual maturity and fertility.

Gonadotropins

Because of the pulsatile secretion of LHRH, gonadotropin secretion is also episodic. Plasma levels of LH and FSH in the fetus rise after the establishment of the hypothalamic-pituitary portal system, with midmorning LH peaks being the most consistent. In the neonate, episodic LH release as a consequence of pulsatile LHRH secretion is the basis for the pulsatile 24-hour pattern of LH release in the neonate. Plasma levels of LH tend to remain high throughout childhood until puberty, when there is a gradual rise in LH levels, with an additional increase during sleep. The frequency and amplitude of LH pulses increase during sleep through puberty. The gonadal steroid values, however, are useful in determining the stage of pubertal development. To that end, the mean plasma estradiol, FSH, and LH concentrations in prepubertal and pubertal females by pubertal stage of maturation (1 = prepubertal; 5 = menstruating adolescents) and the mean bone age for each stage are shown in Table 24-13-

TABLE 24-13 -- Cardinal Hormonal Characteristics of Puberty

<table>
<thead>
<tr>
<th>Increased amplitude and frequency of LH pulses (initially at night)</th>
<th>Increased LH response to intravenous LHRH</th>
<th>Increased serum IGF-I concentration</th>
<th>Increased prolactin secretion in girls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increased GH secretion</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Remarkably, in the normal or the normal male monkey, there is relatively small change in LHRH messenger ribonucleic acid (mRNA) in the hypothalamus with the onset of pubertal development, suggesting that other factors repress LHRH secretion during the juvenile pause.

The peripubertal period is the time that immediately precedes signs of sexual maturation. In the prepubertal period, ultrasensitive LH assays provide indirect evidence of a diurnal variation of LHRH secretion with a preponderance at night; the frequency is similar to that of the early pubertal period, although the amplitude of LH does not increase until the peripubertal period. During the prepubertal period, enhanced release of LH can first be shown in response to intravenous LHRH, and the augmented release of pulsatile LH during sleep is prominent. During puberty, the episodic secretion of FSH and LH becomes more clear-cut as the amplitudes of the gonadotropin pulses increase. An increased amplitude of LH and FSH secretion occurs at night in prepubertal boys and girls; there is a temporal concordance of LH pulses with FSH pulses of 43%. Primary testicular failure is associated with enhanced amplitude and frequency of both FSH and LH pulses.

In the past, single daytime serum samples did not reliably indicate the stage of puberty because of the insensitivity of the assays. Nonetheless, studies of a large number of individuals using single daytime samples documented changes in the mean serum gonadotropin levels between prepuberty and puberty. In girls, FSH levels rise during the early stages of puberty and LH levels tend to rise in the later stages, from beginning to late puberty, the LH concentration rises over 100-fold. In boys, FSH levels rise progressively through puberty and LH levels rise and reach an early plateau.

Ultrasensitive LH and FSH assays now allow the accurate determination of basal levels of serum LH and FSH, and the results are lower than previous assays could detect. The basal values of serum LH and FSH are reported to predict the onset of pubertal development as well as can LHRH testing, a value of serum LH greater than 4 mIU/mL measured by immunochromatimometric assay is consistent with the onset of puberty. There is a more striking rise in serum LH amplitude by at least 1 year before the onset of puberty (the latter indicated by a testicular volume of at least 3 mL), whereas FSH rises more consistently through male pubertal increase with increased pulse amplitude. Moreover, the use of these ultrasensitive assays to determine concentrations of LH and FSH in urine reveals a pattern of a 5-fold rise in urinary FSH in boys and girls and a 52-fold rise in urinary LH in boys and a 100-fold rise in girls during puberty.

Doses of exogenous LHRH that are relatively ineffective in stimulating gonadotropin or gonadal steroid secretion before puberty become effective with the onset of puberty; thus, an amplification occurs in the hypothalamic-pituitary-gonadal axis with progression of puberty. Whereas the LHRH test usually requires multiple sampling after the administration of LHRH, a single determination at 30, 45, or 60 minutes may suffice with the new, sensitive immunoradiometric assay, but only if a positive result is obtained. Further, the use of an LHRH agonist (e.g., goserelin) in a single dose with determination of serum gonadotropins and sex steroids can help to differentiate the pubertal from the prepubertal state.

The pattern of release of gonadotropins and testosterone was determined in boys at every stage of pubertal development by sampling serum every 20 minutes for 24 hours. Disorderly patterns of secretion of LH but not FSH were noted just before the onset of puberty, followed by increased orderliness in early puberty and then...
increased disorderliness again in later puberty. This suggests that a more integrated feedback system operates in early puberty, which is followed by less stability.

Levels of biologically active LH as determined by rat or mouse interstitial cell assays and of biologically active FSH as assessed by rat Sertoli or granulosa cell assays have been compared with immunoreactive LH and FSH levels estimated by radioimmunoassay. Although discrepancies between serum bioactivity and immunoactivity of LH were reported earlier, more recent data indicate that a change in the bioactive/immunoactive ratio does not occur during puberty in boys or girls.

Qualitative as well as the well-defined quantitative changes occur in the pattern of FSH and LH in the pituitary gland, serum, and urine during development. The pattern of glycosylation of the a and b subunits of the gonadotropins is influenced by maturation, LHRH secretion, and the action of gonadal steroids on the pituitary gonadotrophs. Variation in glycosylation that affects the size and charge of the hormone is the principal cause of the heterogeneity of FSH and LH and the large number of isoforms, which vary according to the more acidic or more basic charge. This pleomorphism has an important effect on biologic half-life and biologic activity and provides an additional mechanism of regulating the biologic activity of the gonadotropins.

Although it has been difficult to characterize a diurnal variation of immunoreactive FSH secretion, secretion of bioactive FSH increases at night during sleep and is more resistant to testosterone-induced suppression than is immunoreactive LH.
Gonadal Steroids

Estrogen exerts different effects than testosterone, but only lately has it been appreciated that many actions on linear skeletal growth, skeletal maturation, and accretion of bone mass thought to be due to testosterone in the male are mainly attributable to its peripheral aromatization to estrogen.
Testosterone

The Leydig cells of the testes produce testosterone and, in lesser amounts, androstenedione, 5-androstenediol, dihydrotestosterone, and estradiol. In addition to direct secretion, a small amount of testosterone is derived from extraglandular conversion of androstenedione secreted by the testes and the adrenal. Although testosterone induces development of a male body habitus and voice change, dihydrotestosterone derived by 5-reduction in the target cell is the major mediator of the development of the phallus and the prostate, temporal hair recession, and beard growth. In the female, extraglandular conversion of ovarian and adrenal androstenedione accounts for almost all of the circulating testosterone.

Prepubertal boys and girls have plasma testosterone concentrations less than 0.3 nmol/L (0.1 ng/mL) except during the first 3 to 5 months of infancy in the male, when pubertal levels are found. Nighttime elevations of serum testosterone levels are detectable in the male by 5 years of age, before the onset of physical signs of puberty, and increase during early puberty after the appearance of sleep-entrained secretion of LH and increased pituitary sensitivity to LHRH. There is a lag of about 60 minutes between the peak of LH and the increase in testosterone, presumably related to synthesis and secretion of the steroid. In the daytime, increases in testosterone levels are detectable at approximately 11 years in boys when the testis volume is at least 4 mL, with a consistent increase throughout puberty. The steepest increment in testosterone occurs between pubertal stages 2 and 3 in males; testosterone concentrations can rise from 0.7 to 8 nmol/L (0.2 to 2.4 ng/mL) within 10 months.

Normal values for testosterone and other androgen metabolites are described; measurements of the ratio of testosterone to its metabolites can be used to identify athletes using illicit androgen preparations. Unfortunately, the ratio of testosterone to epitestosterone in the urine, which is used in this manner to evaluate "doping," may be elevated normally during the progression through puberty, casting doubts upon this testing procedure during puberty.

Free testosterone values are low or nondetectable until the age of normal pubertal development, at which time they rise in boys and girls. A sensitive mammalian cell recombinant bioassay for androgen bioactivity strongly correlated with serum immunoreactive testosterone concentration but not with 5-dihydrotestosterone, dehydroepiandrosterone, or androstenedione.

Sex steroids can be measured in saliva (as can many other types of steroids) for screening purposes or for monitoring.
Estrogens

In the female, the major estrogen, estradiol, is principally secreted (90%) by the ovary; a small fraction of circulating estradiol arises from the extraglandular conversion of testosterone and androstenedione. In the male, approximately 75% of estradiol is derived from extraglandular aromatization of testosterone and (indirectly) androstenedione and 25% from testicular secretion. Aromatase is absent or present in barely detectable amounts in prepubertal testes but maximal amounts appear in late puberty. In normal testes aromatase is predominantly present in the Leydig cells, but in testicular tumors of either Sertoli or Leydig cells, for example, associated with the Peutz-Jeghers syndrome, the Sertoli cells of the tumor express aromatase.

In the fetus and at term, levels of estrogens are high because of the conversion of fetal and maternal adrenal C19 steroids to estrogen by the placenta. Plasma levels of estrogen drop precipitously in the first few days of life. Estrogen levels are so low in prepuberty that detection has been difficult with standard techniques, but a highly sensitive bioassay demonstrated measurable serum concentrations of estradiol in both boys and girls before puberty with higher estradiol concentrations in girls than boys.

The mean concentration of serum estradiol equivalents in 21 prepubertal girls (7.7 ± 1.9 years) was 0.6 ± 0.6 (SD) pg/mL, significantly greater than the concentration (0.08 ± 0.2 pg/mL) found in 23 prepubertal boys (9.4 ± 2.0 years). The higher estrogen levels in girls may be an important factor in the more advanced levels of skeletal maturation in girls and play a part in their earlier onset of sexual maturation. Subsequently, the plasma estradiol level rises steadily through the stages of puberty until maturity and exhibits a diurnal rhythm (see Fig. 24-17), when concentrations of about 500 pg/mL are reached in the follicular stage and about 200 pg/mL in the luteal phase; estrone levels rise early and reach a plateau by midpuberty. The daily peak of estradiol in early pubertal girls occurs about 6 to 9 hours after the peak of serum LH detected during the night, apparently related to the time necessary for ovarian synthesis of estradiol.

### TABLE 24-14 -- Differences in the Timing of the Onset of Estrogen Synthesis in Girls and Boys

<table>
<thead>
<tr>
<th>Girls</th>
<th>Follicle-stimulating hormone (FSH) from late fetal life through puberty stimulates aromatase and estrogen synthesis by the ovary.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boys</td>
<td>FSH leads to enlargement of the testes, the earliest sign of puberty in the male; spermarche occurs early in puberty and spermaturia at a mean age of 13.3 years before the sharp rise in testosterone levels and peak height velocity. Estradiol synthesis is not detectable in fetal or prepubertal Leydig cells and is at a very low level, until luteinizing hormone stimulates Leydig cell aromatase at late stage II to stage III of male secondary sexual maturation. Estradiol does not reach the level found in girls in early puberty who exhibit a pubertal growth spurt until at least midpuberty.</td>
</tr>
</tbody>
</table>

In all stages of puberty, boys have higher concentrations of estrone than estradiol, and levels of both estrogens are lower than those in girls at comparable stages. Boys have higher levels of estrone and estradiol in pubertal stage 5 than in stage 1. A new ultrasensitive estradiol assay demonstrated measurable serum estradiol equivalents in prepubertal boys with a rise through puberty until the pubertal growth spurt and a decrease thereafter. Klein and co-workers found a high correlation with peak growth velocity and the rise in estradiol concentration, which in boys occurred about 3 years after the onset of puberty. The mean estradiol level at peak growth velocity was similar in boys and girls at about 3 to 4 pg/mL. (Table 24-14)
Adrenal Androgens

There is a progressive increase in plasma levels of 5-steroids, dehydroepiandrosterone (DHEA) and dehydroepiandrosterone sulfate (DHEAS), in both boys and girls beginning before age 8 (skeletal age of 6 to 8) and continuing through early adulthood (Table 24-15). The increase in the secretion of adrenal androgen and its precursors is known as adrenarche. Plasma DHEA has a diurnal rhythm similar to that of cortisol, but plasma DHEAS shows less variation and is a useful biochemical marker of adrenarche. The role of adrenarche in puberty is discussed later (see "Adrenal Androgens and Adrenarche").
Testosterone-Binding Globulin (Sex Steroid-Binding Globulin)

Between 97% and 99% of circulating testosterone and estradiol is reversibly bound to TeBG; only the free steroid is physiologically active. TeBG is a glycoprotein of 90 to 100 kd, consists of heterogeneous monomers, and has one steroid binding site per dimeric molecule. Prepubertal levels of TeBG are approximately equal in boys and girls, and a decrease in TeBG level occurs with advancing prepubertal age and the concomitant increase in the plasma gonadal steroid levels; at puberty there is a small decrease in TeBG levels in girls and, as a consequence of testosterone, a greater decrease in boys. The rise in adrenal androgen levels at adrenarche may explain the early drop in TeBG levels, which allows more circulating free hormone at a given concentration of testosterone. Although the plasma concentration of testosterone is 20 times greater in men than in women, the concentration of free testosterone is 40 times greater. Boys with hypogonadotropic hypogonadism and patients with the androgen resistance syndrome show the same characteristic fall in TeBG levels at puberty, but values are intermediate between normal adult males and females. TeBG production is regulated by a number of factors. For example, it is down-regulated by GH administration in prepubertal children, perhaps by the action of IGF-I. TeBG is decreased in prepubertal children with diabetes mellitus.

| TABLE 24-15 -- Mean Serum Concentrations of Dehydroepiandrosterone Sulfate During Childhood |
|-----------------------------------------------|---------------|---------------|---------------|---------------|---------------|
| Concentration, µmol/L (ng/mL), at Chronologic Age | 68 yr | 810 yr | 1012 yr | 1214 yr | 1416 yr |
| Boys | 0.5 (188) | 1.6 (586) | 3.4 (1260) | 3.6 (1330) | 7.2 (2640) | 7.2 (2640) |
| Girls | 0.8 (306) | 3.2 (1170) | 3.1 (1130) | 4.6 (1690) | 6.9 (2540) | 6.3 (2320) |

| Concentration, µmol/L (ng/mL), at Bone Age |
|-----------------------------------------------|---------------|---------------|---------------|---------------|
| Boys | 0.98 (360) | 1.6 (574) | 3.4 (1250) | 5.8 (2150) | 10.9 (4030) |
| Girls | 0.73 (276) | 3.1 (1130) | 4.33 (1560) | 7.1 (2610) | 3.9 (1450) |

Prolactin levels rise in girls during puberty. Prepubertal mean (± standard error) plasma prolactin concentrations are 4.0 ± 0.5 µg/L in boys and 4.5 ± 0.6 µg/L in girls. Late pubertal girls and adult women have higher concentrations of prolactin (7.5 ± 0.7 and 8.3 ± 0.7 µg/L), whereas the mean concentration in adult men is 5.2 ± 0.4 µg/L. This sex difference is likely to be a consequence of the higher estradiol levels during puberty in girls and in women.
Inhibin, Activin, Follistatin

Inhibin, activin, and follistatin were discovered by their effect on FSH secretion. Inhibin and follistatin inhibit and activin stimulates FSH subunit expression and hence FSH biosynthesis and secretion. It is now recognized that they are synthesized in a variety of tissues in addition to the gonads and have diverse activities apart from those on the reproductive apparatus. Two distinct binding proteins for inhibin and activin have been described that are present in the circulation, the gonads, and other tissues: a 2-macroalbumin, a high-capacity, low-affinity binding protein, and follistatin, a glycosylated single peptide chain that functions not only as a high-affinity binding protein but also as a regulator of activin bioactivity (e.g., in the pituitary gland, a site of synthesis of both activin and follistatin). 1414

Inhibin, a heterodimeric glycoprotein product of the Sertoli cell of the testes and the ovarian granulosa cell (as well as the placenta and other tissues), exerts a negative feedback action on the secretion of FSH from the pituitary. Inhibin is composed of a subunit and one of two subunits, A or B, which form inhibin A or inhibin B, respectively, dimers with apparently identical function. Inhibin is a member of the transforming growth factor (TGF-) superfamily, which includes antimüllerian hormone (AMH, also called müllerian-inhibiting factor) and the dimers of two inhibin subunits, activin A and activin B, that stimulate the release of FSH from pituitary cells. 5 FSH induces synthesis and secretion of gonadal inhibin. Inhibin plays a role in the feedback regulation of FSH secretion during puberty in males and females. 1415 In men, inhibin B is a major feedback regulator of FSH release. 1416 (Inhibin B and inhibin A exhibit specific patterns of secretion during the menstrual cycle. 1417)

The assay for plasma inhibin is confounded by the fact that the dimeric inhibins occur in a wide range of molecular weights, including combinations of precursor forms of each of the subunits as well as a precursor of the subunit of inhibin, pro-C. The Monash radioimmunoassay detects the dimeric inhibins, inhibins A and B, as well as pro-C and related peptides and is essentially a nondiscriminatory total inhibin assay. 1418 Since the development of highly specific enzyme-linked immunosorbent assays for the mature 31-kd inhibin A and inhibin B dimers and for inhibin pro-C, 1419 1420 the detection of sex-specific differences in the pattern of inhibin secretion has advanced our knowledge of the biologic action of inhibin and the clinical usefulness of inhibin determinations. 1419 1420

During pregnancy, the placenta secretes inhibin A and the fetal membranes secrete both inhibin A and inhibin B, whereas, at least for the first 20 weeks of gestation, only inhibin A was detected in maternal serum. 1421 In umbilical cord serum from term female newborn infants, no inhibin was detected, whereas cord serum from male newborns contained inhibin B, the only inhibin detected in adult males; the median value was 167 pg/mL. 1422 In the human fetal testis -and B (but not A) subunits are present in both Sertoli and Leydig cells at 16 weeks of gestation; by 24 weeks of gestation immunoperoxidase of both subunits was greater in the Sertoli cells. Postnatally, the expression of both subunits was decreased by 4 months of age. Inhibin subunits were not detected in the fetal ovary, nor was immunoreactive follistatin present in fetal or neonatal gonads. 1423 These findings are consistent with inhibin A and inhibin B values in midgestation and term fetuses. 1424

Immunoreactive inhibin-like activity measured by the Monash assay increases in both boys and girls during puberty, with mean plasma levels increasing in boys from 161 to 442 U/L and in girls from 97 to 231 U/L between stage 1 and stage 5. 1425 In boys, the rise in serum immunoreactive inhibin was relatively constant during puberty, increasing 1.5-fold between a testis volume of 1 and 10 mL. 1426 However, as discussed before, the early inhibin immunosays cross-react with inactive monomeric inhibin precursors. There is a striking sex dimorphism in the pattern of circulating inhibin B and A from fetal life through full sexual maturation. 1427 In large cross-sectional studies 1428 1429 using highly specific inhibin B and inhibin A immunosays that correlate with the bioactivity of inhibin and distinguish inhibin B from inhibin A, the mean concentration of serum inhibin B increased between prepuberty (a stage when it is higher than the undetectable levels in castrate men) 1428 and the first stage of puberty; when the strong correlation with chronologic age was taken into account, a correlation with LH and testosterone values remained. From genital stage 2 puberty on, inhibin B levels were relatively constant despite a rise in the mean concentration of serum FSH between stages 2 and 3, after which the FSH value was relatively unchanged. By genital stage 3 a negative partial correlation between inhibin B and FSH was found that persisted as puberty advanced. 1428 1429 and by genital stage 4 there was a clear negative correlation of inhibin with serum FSH. 1428 Crofton and colleagues found that dimeric inhibin B rose twice in development, reflecting the two periods of Sertoli cell proliferation during infancy and early puberty, whereas an inverse relationship between inhibin and FSH was seen at midpuberty and thereafter, indicating the development of the negative feedback inhibition. 1428

Serum inhibin A and inhibin B increase early in puberty in girls, although there are individual increases in the prepubertal period directly related to FSH levels demonstrating sporadic follicular development in the infant and child related to FSH stimulation. 1428 Inhibin B is predominant in the follicular phase and inhibin A during the luteal phase. 1428 More specifically, inhibin A and inhibin B peaked in midpuberty and inhibin B decreased thereafter. 1428 Although there is no significant change in activin during female puberty, follistatin decreases from a midpuberty peak to later values that fall below prepubertal values.

Serum inhibins and FSH are markedly elevated in chronic renal failure but are reduced after renal transplantation. 1430 A low concentration of inhibin B in men and pubertal boys is an indicator of impaired seminiferous tubule function. 1431

Early pubertal boys with testicular defects have higher FSH concentrations and low inhibin levels. 1432 Inhibin B is the form most closely related to testicular function and is absent in orchidectomized men. 1432 Inhibin B is related to Sertoli cell function in prepuberty, but a developmental change occurs during puberty so that later in life inhibin B is related to spermatogenesis. Prepubertal boys with the Sertoli cellonly syndrome had normal inhibin B levels; postpubertal boys and men with Sertoli cell-only syndrome and early stage spermatogenic arrest had undetectable or low levels of inhibin B, whereas those with late stage spermatogenic arrest or obstructive azoospermia had normal or near-normal levels of serum inhibin B. 1433 1434 It is suggested that in prepuberty both the inhibin B and subunits are expressed in Sertoli cells, but during puberty and in men fully differentiated Sertoli cells express only the subunit and the subunit is expressed in germ cells; inhibin B in the adult appears to be a product of both germ and Sertoli cells. In prepubertal boys, basal plasma inhibin B concentrations have a high correlation with the incremental testosterone response to the administration of hCG and provide a useful assessment of both the presence of testes and its function. 1435
Antimüllerian Hormone

AMH (or müllerian inhibitory substance), a 14-kd glycoprotein dimer structurally related to the subunit of inhibin and TGF-, is produced by the Sertoli cell of the fetal testis and later in gestation by granulosa cells of the fetal ovary. Immunoassayable concentrations of AMH rise from birth to relatively high levels in the first year in newborn males, decrease by age 10, and decrease further during puberty. Newborn females have low or nondetectable serum levels of AMH, which rise only slightly thereafter; serum AMH concentrations are virtually nondetectable in most girls just before puberty.

There is an inverse relationship between serum AMH and androgen concentrations in pubertal boys and in boys with true precocious puberty, in whom values were appropriate for pubertal stage rather than chronologic age; in addition, patients with androgen resistance have elevated serum AMH concentrations in the newborn period and again at puberty and thereafter. Values are elevated in males with primitive Sertoli-like tumors and in girls and women with granulosa cell tumors; AMH is a useful gonadal tumor marker. Concentrations of AMH are slightly higher in individuals with delayed puberty than in pubertal age-matched control subjects and lower in those with testicular dysgenesis associated with impaired virilization than in normal boys. However, boys with isolated cryptorchidism have normal values of AMH. The serum concentration of AMH is useful in determining the presence of testicular tissue in differentiating anorchia from bilateral cryptorchidism in prepubertal boys; in the former AMH is absent, and in the latter AMH is in the normal range. Patients with dysgenetic testes often have low serum levels of AMH, and measuring the testosterone response to hCG is indicated to assess the presence of testicular tissue.
Prostate Specific Antigen

Prostate specific antigen (PSA) is detectable in both male and female cord blood and in the serum of infants, but PSA concentrations decrease to undetectable levels during childhood. PSA concentrations rise to the measurable range with the onset of puberty in the male and correlate with the progression of pubertal stage, the size of the testes and presumably the prostate, and serum LH and testosterone concentrations. PSA values were increased into the pubertal range in boys with idiopathic true precocious puberty and showed a significant decrease with GnRH agonist treatment.
Growth Hormone

As discussed earlier, serum GH concentrations rise during pubertal development in relation to increased gonadal sex steroid secretion. The secretion of GH increases twofold to threefold during puberty. In individuals with delayed puberty, increased GH secretion can be induced by the administration of exogenous androgens. GH rises in both boys and girls of normal stature during puberty but secretion decreases after the end of pubertal development; there is a relatively greater rise in girls, which starts at an earlier chronologic age than in boys because of the earlier onset of puberty in girls and occurs at an earlier pubertal stage in girls than in boys. During puberty in adolescents of normal height, there is an inverse relationship between weight and GH levels. Even though the frequency of GH secretory episodes is not altered by puberty or androgen administration, the amplitude and mass of the pulses increase. After pubertal development, GH secretion decreases with advancing age. Although it would appear that the rise in testosterone in boys mediates the rise in GH, it is mainly the aromatization of testosterone to estradiol that induces the effect on GH secretion (see earlier).

Treatment of late pubertal boys with the estrogen receptor blocker tamoxifen led to smaller GH secretory peaks and, to a lesser degree, fewer GH secretory episodes, supporting the critical role of estrogen. Further, the administration of exogenous estrogen increased the peak GH reached after insulin-induced hypoglycemia, exercise, and arginine. GH release is stimulated by hypothalamic GH-releasing hormone (GHRH), but another class of six- and seven-amino-acid peptides (growth hormonereleasing peptide or GHRP) also stimulate GH release independently and in a manner additive to GHRH. GHRP stimulates the release of GH in the absence of GHRH as well as in its presence. GH secretion is increased in puberty compared with prepuberty both in the basal state and after GHRH or GHRP stimulation. Urinary GH excretion reflects serum levels and changes with pubertal development as a peak was reached at pubertal stage 3 to 4 in studies with sensitive enough assays; higher values of urinary GH content were reached in boys than in girls.

Urinary GH determinations after intravenous administration of arginine correlate inversely with BMI and GH excretion in pubertal but not in prepubertal short normal children or in patients with Turner's syndrome.

Growth hormonebinding protein (GH-BP) has the same amino acid sequence as the extracellular component of the GH receptor and is directly related to the amount of cellular GH receptors; in normal children, plasma GH-BP is inversely related to 24-hour GH secretion. There is disagreement over the changes in GH-BP with maturation. Serum GH-BP rises early in childhood and rises through puberty in some cross-sectional studies but not in others or in longitudinal studies. In the longitudinal study in which plasma GH-BP did not change appreciably with the onset of puberty, it was suggested that at the time of the pubertal growth spurt there is a relative increase in unbound (free) GH in relation to GH bound to GH-BP. GH-BP is related to adiposity, and it may be this factor that accounts for the increased levels of GH-BP in girls compared with boys and for the rise in GH-BP in girls with precocious puberty as well as the negative influence of testosterone on GH-BP levels.
Insulin-Like Growth Factor I

A cross-sectional study of free IGF-I concentrations in 1030 subjects confirmed the pattern of a slow rise in serum free IGF-I in prepuberty followed by a steeper rise during puberty. The concentration of IGF-I rises during puberty to levels higher than those of prepubertal or adult subjects, remains elevated past the time of peak height velocity with a peak attained 1 or 2 years after the pubertal growth spurt, thus later in boys than in girls, and then falls to normal adult levels. The pattern of the GH-dependent serum insulin-like growth factorbinding protein 3 (IGF-BP-3) in pubertal development is similar to that of serum IGF-I. However, serum IGF-BP-3 concentrations correlate with BMI even though IGF-I does not.

Measurement of free IGF-I shows the same pattern of change with development as the measurement of total IGF-I, but the magnitude of the change in puberty differs between the two. A decrease of free IGF-I is described with age in the later stages of puberty. The increase in the serum ratio of IGF-I to IGF-BP-3 at the time of the pubertal growth spurt appears to be due to production because proteolysis of IGF-BP-3 does not change in puberty in normal children. The testosterone level in boys and the estradiol level in girls correlate with the rise in IGF-I concentration, but gonadal steroids are not the direct cause of the increase in circulating IGF-I levels. Secretion of GH approximately doubles during puberty; the major effect of estrogen and testosterone on IGF-I generation is mediated indirectly through augmented release of GH (see earlier). Children with true precocious puberty have plasma IGF-I values characteristic of children in the same stage of normal puberty rather than children of the same chronologic age. After treatment of children with true precocious puberty with LHRH agonists to lower plasma gonadotropin and sex steroid levels, IGF-I values slowly decrease along with the secretion of GH. In sexual precocity, as in normal puberty, gonadal steroid secretion appears to stimulate GH secretion, which in turn increases IGF-I generation.

Serum IGF-II shows no pubertal peak and falls during adulthood in boys.
Insulin

Insulin sensitivity decreases normally during puberty. Serum fasting insulin concentration increases twofold to threefold with peak height velocity, and insulin secretion after a glucose load increases over prepubertal levels, suggesting a degree of insulin resistance during normal puberty. *This impairment of insulin-stimulated insulin metabolism can be demonstrated by the euglycemic insulin clamp technique* in normal subjects and is more striking in adolescents with diabetes mellitus. Further, insulin-mediated glucose disposal in oxidative and nonoxidative pathways of cellular glucose utilization in peripheral tissues is about 30% lower during Tanner stages 2 and 4 than in prepuberty and the young adult stage as determined by using the hyperinsulinemic euglycemic clamp technique or the minimal model, frequently sampled intravenous glucose tolerance test. *Hyperglycemic clamp studies also indicate that pubertal individuals compensate for this defect by increasing insulin secretion and suggest that the insulin resistance does not involve the effect of insulin on amino acid metabolism* in normal pubertal individuals. *Thus the enhanced insulin response to glucose in puberty related to the relative insulin resistance may increase the anabolic effects of insulin with regard to protein metabolism.*

This normal phase of insulin resistance appears temporarily and possibly causally related to the rise in GH, a hormone that opposes the action of insulin. A normal individual adapts to these changes, but an individual at risk for type 2 diabetes may not adapt to the insulin resistance and with the accompanying defect in pancreatic action characteristic of type 2 diabetes often develops clinical type 2 diabetes during the pubertal years. Indeed, with the increasing prevalence of obesity in the young, there is a secondary peak in the onset of type 2 diabetes at 13.5, although it is encountered in younger children as well.

Insulin resistance is related to many cardiovascular risk factors. Further, insulin resistance is characteristic of the state of functional ovarian hyperandrogenism seen after a history of premature puberty; this constellation is more frequent in children with a history of low birth weight. Insulin resistance in childhood appears to decrease the likelihood of tracking of obesity into adult age. The response of insulin to an oral glucose tolerance test is greater in black subjects than in white subjects at all stages of pubertal development; this ethnic difference in insulin resistance is suggested as a cause for the increased incidence of type 2 or non-insulin-dependent diabetes mellitus in black adults compared to white adults.

Patients with type 1 (insulin-dependent) diabetes mellitus usually require an increase in the dose of insulin for euglycemic control at puberty. The insulin resistance has been attributed, at least in part, to increased fat oxidation at puberty, which correlates with rising serum IGF-I and may be linked to increased GH secretion. Insulin sensitivity is related to pubertal stage and BMI; in a longitudinal study, insulin sensitivity inversely correlated with BMI, decreased with progression from pubertal stage 2 to stage 3, and was lower in girls than boys at either stage 2 or 3. Insulin resistance is present early in the course of Turner’s syndrome and thalassemia major. In children with well-managed type 1 or insulin-dependent diabetes mellitus, the growth rate may decrease mildly and transiently in the 10 years after diagnosis; bone age advancement slows during this period, leading to a transient slowing of development. However, the effect of diabetes on growth is smaller than the genetic influence of parental height. Although growth velocity increases again, mainly during puberty, weight gain increases even more in affected children during puberty, leading to a higher incidence of obesity in children with type 1 diabetes mellitus than expected from family patterns.

Unfortunately, some adolescents with type 1 diabetes mellitus, predominantly girls, reduce their insulin use in order to lose weight, with dire consequences. A retrospective study found a decrease in final height if the diagnosis of type 1 diabetes mellitus was made prior to 5 years of age but not if the diagnosis of type 1 diabetes mellitus was made later; the pubertal growth rate was reduced in all patients, but girls were more affected than boys. Adolescents with type 1 diabetes mellitus have a lower serum concentration of IGF-I than control subjects. The elevated GH levels are associated with low serum IGF-I and low serum GH-BP concentrations, and there is no longer the usual reciprocal relationship between serum GH and GH-BP.

Obese teenage girls with predominant abdominal adiposity have insulin resistance and are at higher risk for the development of breast cancer; abdominal adiposity may be recognized in the prepubertal state and is associated with early puberty, early menarche, and, later exposure to an endocrine profile predisposing to breast cancer. Plasma insulin in obese adolescent boys correlates with fasting plasma glucose, plasma triglycerides, uric acid, and systolic blood pressure, whereas in obese adolescent girls it correlates with plasma triglycerides and systolic and diastolic blood pressure; plasma insulin correlated negatively with HDL cholesterol in both boys and girls. Retinopathy related to type 1 diabetes mellitus characteristically appears in the teenage years or later, but the duration and control of diabetes in the pubertal years are contributing factors; there is an increased appreciation of the prevalence of retinopathy in the prepubertal years.

The fluoroscopic photometry ratio increases at the time of puberty, indicating a decrease in the blood-retina barrier during this period. The American Diabetes Association recommends screening for microalbuminuria, an indicator of the development of diabetic nephropathy; microalbuminuria may develop quite early in puberty rather than at the later stages as previously suggested. With the increased prevalence in type 2 diabetes, screening criteria have been proposed but are under evaluation. At present, a child with BMI-SC more than the 85th percentile should be screened if one of the following criteria is present: (1) a family history of type 1 diabetes mellitus, (2) signs of insulin resistance (acanthosis nigricans, FOH, hypertension, or dyslipidemia), or (3) membership in certain ethnic groups (black, Native American, Hispanic American, and Asian American). If a fasting plasma glucose level is greater than 126 mg/dL, a 2-hour postprandial value is over 200 mg/dL, or there are symptoms such as weight loss, polyuria, or polydipsia and a casual plasma glucose is over 200 mg/dL, the diagnosis of diabetes mellitus is likely and determination of the type of diabetes is appropriate.

Type 2 diabetes mellitus has been recognized in children and adolescents in the past, but the incidence is increasing, probably because of the increase in obesity in these age groups. Type 2 diabetes mellitus appears most often among children and adolescents of ethnic minority populations (African American, Native American, Hispanic American, and, more recently appreciated, Asian American) and should be considered in subjects manifesting acanthosis nigricans as a marker for insulin resistance; girls with hyperandrogenism not otherwise explained are also concerned of Type 2 diabetes mellitus in the young is a heterogeneous disorder and obesity, although common, is a variable feature.

This remarkable increase in type 2 diabetes mellitus should not be confused with maturity-onset diabetes of the young (MODY) syndromes. Six variations of MODY have been defined, all with slowly progressive loss of pancreatic beta cell function and inherited as an autosomal dominant trait in individuals who need not be obese. Type 2 MODY appears to be linked to a mitochondrial defect in glucokinase. Patients with MODY may ultimately require insulin therapy.

Several syndromes of insulin resistance combine hyperglycemia and virilization. The Kahn type A syndrome features include a lean, muscular adolescent female phenotype with acanthosis nigricans, hirsutism, oligomenorrhea or amenorrhea, and ovarian hyperthecosis with stromal hyperplasia associated with abnormalities of the insulin receptor gene. Hyperandrogenism, acanthosis nigricans (HAIR-AN) syndrome and polyolcystic ovary syndrome (PCOS) are less severe than Kahn type A and are usually manifest in adolescent females. Rabson-Mendenhall is a syndrome of severe insulin resistance (possibly leading to diabetic ketoacidosis), dysmorphic facies, acanthosis nigricans, thickened nails, hirsutism, dental dysplasia, abdominal distention, and phallic or clitoral enlargement. The Rabson-Mendenhall syndrome, like the Donohue leprechaunism syndrome, which shares some features, is due to homozgyous or compound heterozygote defects in the insulin receptor gene. Kahn type B syndrome is due to inhibitory or stimulatory antibodies to the insulin receptor, sometimes with acanthosis nigricans and ovarian hyperandrogenism; this syndrome can occur in ataxia-telangiectasia syndrome or in otherwise normal adolescents. Individuals with the Seip-Berardinelli syndrome combine lipodystrophy and severe insulin resistance and complete or partial absence of subcutaneous fat with increased growth and skeletal maturation, muscle hypotrophy, acanthosis nigricans, hypertrichosis, organomegaly, and mild hypertrophy of the external genitalia. Most of these type 2 diabetes mellitus syndromes can be treated with oral hypoglycemic agents initially; progression of the disorder may require the use of insulin.
Serum Lipids

Testosterone increases serum LDL cholesterol and decreases HDL cholesterol concentrations and thereby accounts for the adverse LDL/HDL ratio in adult males compared with adult females. Postheparin hepatic lipase activity is increased by exogenous androgens (and decreased by estrogens), accounting for the decrease in HDL after androgen treatment or after a rise in endogenous androgen secretion. Serum lipoprotein A is not related to pubertal stage in normal subjects but rather appears related to genetic influences. Even though there is an increased risk of coronary heart disease in patients with type 1 diabetes mellitus and lipoprotein A is related to coronary heart disease, lipoprotein A concentrations are not different in children with type 1 diabetes mellitus compared with control subjects. However, lipids tracked over the pubertal period in a longitudinal study showed that 9-year-old girls with high LDL cholesterol retained high LDL concentrations 12 years later and those starting with lower LDL continued to have lower LDL. In an Australian study of prepubertal and early pubertal children, obesity was related to elevation of apolipoprotein B and the ratio of apolipoprotein B to apolipoprotein A-1; these measures are correlated with the development of atherosclerosis. In an American study, the waist circumference and the waist-to-hip ratio correlated inversely with serum HDL and directly with serum triglycerides and the BMI correlated positively with serum triglycerides and diastolic blood pressure; waist circumference was correlated with apolipoprotein B and the apolipoprotein B/apolipoprotein A-1 ratio.

Patients with familial hypercholesterolemia already have carotid intimal plaques by the age of puberty demonstrable by B-mode ultrasonography. Further, intra-abdominal adipose tissue indicated by MRI is positively related to serum total and LDL cholesterol and triglycerides in obese adolescents 10 to 15 years of age, an indicator of cardiovascular risk factors at these early ages. A 40-year longitudinal study of obese children and adolescence found continuing obesity into adulthood; the heaviest subjects and those who gained the most weight during the study had a greater risk of death or cardiovascular disease as well as diabetes.
Cortisol

Although no change occurs in the secretory rate of cortisol, salivary cortisol values increase slightly and correlate with pubertal stage without a sex difference.
The onset of puberty is a consequence of maturational changes that are incompletely understood. The development of secondary sexual characteristics, the adolescent growth spurt, the attainment of fertility, and the psychosocial changes entain from the maturation of the gonads and the increase in gonadal steroid secretion. The events characterizing the development of gonadal function can be viewed as a continuum extending from sexual differentiation and the ontogenesis of the hypothalamic-pituitary-gonadal system through a juvenile pause (in which the system is largely quiescent) to the attainment of full sexual maturation and fertility during puberty. These developmental and maturational events have as their end point procreation.

Two independent, but associated processes (controlled by different mechanisms but closely linked temporally) are involved in the increased secretion of gonadal steroids in the peripubertal and pubertal period. The first, adrenarche, the increase in adrenal androgen secretion, precedes by 2 years or so the second, gonadarche, the consequence of the pubertal reactivation of the hypothalamic-pituitary-gonadal system

---

**TABLE 24-16** — Hypothesis of the Control of the Onset of Human Puberty

| 1. Central Dogma: The CNS exercises the only major restraint on the onset of puberty. The neuroendocrine control of puberty is mediated by the hypothalamic LHRH-secreting neurosecretory neurons in the mediobasal hypothalamus, which act as an endogenous pulse generator (oscillator). |
| 2. The development of reproductive function is a continuum extending from sexual differentiation and the ontogenesis of the hypothalamic-pituitary-gonadal system in the fetus to the attainment of full sexual maturation and fertility. |
| 3. In the prepubertal child the LHRH pulse generator, operative in the fetus and infant, functions at a low level of activity (the juvenile pause) because of steroid-independent and steroid-dependent inhibitory mechanisms. |
| 4. Puberty represents the reactivation (disinhibition) of the CNS suppressed LHRH pulse generator characteristic of late infancy and childhood, leading to increased amplitude and frequency of LHRH pulsatile discharges, to increased stimulation of the pituitary gonadotropes, and finally to gonadal maturation. Hormonally, puberty is initiated by the recurrence of augmented pulsatile LHRH and gonadotropin secretion, mainly at night. |

---

CNS, central nervous system; LHRH, luteinizing hormone-releasing hormone.

---

intercellular as well as intracellular messenger. Further, LHRH acting as an autocrine factor may play a role in the synchronization mechanism. LHRH is synthesized in these neurons and released episodically from axon terminals at the median eminence into the primary plexus of the hypothalamic-hypophyseal portal circulation. The hormone is then transported by the portal vessels to the anterior pituitary gland. LHRH is essential for the release of both FSH and LH. In some species, notably rodents, extrahypothalamic CNS structures, including the limbic system (hippocampus and amygdala), influence gonadotropin secretion. Further, the amplitude and frequency of the pulsatile LHRH signal are modified by catecholaminergic and serotoninergic neurons, through their effect on hypothalamic norepinephrine, dopamine, and serotonin, and by opioid peptide, neuropeptide Y (NPY), leptin, galanin, corticotropin-releasing hormone, GABA, and excitatory amino acid neuronal networks. Whether the influence of extrahypothalamic factors on episodic LHRH release in humans is mediated by these pathways remains to be established; among these factors in humans and nonhuman primates, the inhibitory effects of GABA, opioid peptides, and corticotropin-releasing hormone on the LHRH pulse generator and the stimulatory effects of excitatory amino acids and adrenergic pathways are the most firmly established. Thus, the hypothalamic-pituitary gonadotropin unit is influenced by gonadal steroids, inhibin, activin, and follistatin and by complex neural influences that integrate a variety of intrinsic and extrinsic stimuli and functional factors and cues. In vitro studies suggest that the generation of the LHRH pulse is an intrinsic property of the LHRH neurosecretory neuronal network and that other factors modulate the fundamental autoregulation of the LHRH neuron. Several factors

2. The pulsatile gonadotrophs, which contain the seven-transmembrane-domain Gs protein-coupled LH/FSH receptors, are in response to the LHRH rhythmic signal, release LH and FSH in a pulsatile manner. Each LH (and FSH) pulse is induced by a pulse of LHRH.

3. The gonads, which are modulated primarily by the amplitude of the gonadotropin pulse, transmit the episodic gonadotropin signal into pulsatile secretion of gonadal steroids. This control mechanism, with its three principal components (medial basal hypothalamic LHRH neurosecretory neurons, pituitary gonadotrophs, and gonadotropin-responsive elements of the gonad), is common to all mammalian species.

Pattern of Gonadotropin Secretion

There are two pulsatile secretory patterns of gonadotropins: tonic and cyclic. Tonic, or basal, secretion is regulated by a negative, or inhibitory, feedback mechanism in which changes in the concentration of circulating gonadal steroids and inhibin result in reciprocal changes in the secretion of pituitary gonadotropins. This is the pattern of secretion in the male and one of the control mechanisms in the female. Cyclic secretion involves a positive, or stimulatory, feedback mechanism in which an increase in circulating estrogens, to a critical level and of sufficient duration, initiates the synchronous release of LH and FSH (the preovulatory LH surge) that is characteristic of the normal adult woman before menopause. The secretion of FSH and LH is probably always pulsatile or episodic, whether the pattern is tonic or cyclic and regardless of age (i.e., in the fetus, infant, or child, during puberty, or in the adult). However, it is difficult to detect small pulses when the plasma gonadotropin concentration is low (as in prepubertal individuals) because of methodologic limitations. In women, the overall pattern of LH pulse frequency and amplitude varies widely during the menstrual cycle from about one pulse per hour in the midfollicular phase to one pulse per 5 hours in the late luteal phase. Striking changes in the pattern of the pulsatile gonadotropin spikes and their circadian rhythm occur in the peripubertal period and during puberty. Even though LHRH stimulates the release of both FSH and LH, the pulsatile secretion of immunoreactive FSH in normal adult women is prominent; this discordance in FSH and LH pulses is attributed in part to the longer half-life of FSH than LH, to differences in the factors that modulate the action of LHRH on FSH and LH release by the gonadotrophs (especially gonadal steroids, inhibin, and possibly activin and follistatin), and to intrinsic differences in the secretory pattern of the two gonadotrophs. For example, a change in the frequency of LHRH pulses can modify the ratio of FSH to LH released; midfollicular phase concentrations of estradiol and adult male concentrations of plasma testosterone have a greater inhibitory effect on the response of FSH than on that of LH to pulsatile injections of LHRH.

The inherent oscillatory characteristic of gonadotropin secretion is a consequence of the pulsatile release of LHRH. However, the physiologic significance of the episodic, rhythmic pattern of gonadotropin secretion was unclear until studies of the rhesus monkey by Knobil and associates and revealed the essential nature of a periodic, oscillatory LHRH signal for the regulation of gonadotropin secretion. Inhibition of gonadotropin secretion results from the continuous infusion of LHRH because of desensitization of LHRH receptors on the gonadotroph. Intermittent, or pulsatile, administration (e.g., LHRH 1 µg/minute for 6 minutes every hour) restored pulsatile release of LH and FSH in adult monkeys in which hypophysectomy lesions obliterated the arcuate nucleus region and thus eliminated endogenous LH secretion. Further, pulsatile LHRH administration reestablished gonadotropin secretion in animals in which gonadotropin secretion had been suppressed by the continuous infusion of LHRH. These classical studies provided evidence that the LHRH signal to the pituitary gonadotrophs of the adult is frequency as well as amplitude coded. Therapeutic pulsatile administration of natural LHRH has made possible the induction of ovarian or testicular maturation, including fertility, in patients with hypothalamic hypogonadism and the suppression of gonadotropin secretion by long-acting potent LHRH analogues, for example, in boys and girls with true precocious puberty (see later).
Ontogeny

Studies in the mouse, rhesus monkey, human, and all vertebrates examined indicate that LHRH neurons do not originate in the CNS. Instead, they arise in the embryo from the epithelium of the olfactory placode and migrate by an ordered spatiotemporal course along the pathway of the nervous terminals-vernamental complex to the forebrain; the latter also originates in the olfactory placode and forms a connection between the nasal septum and the forebrain (Fig. 24-22). In the mouse embryo (studied in detail by Schwanzel-Fukuda and Pfaff) and by Wray and associates by immunocytochemistry, [H]thymidine autoradiography, and in situ hybridization histochemistry, the LHRH cells arise by embryonic day 9.5 (gestation days 18 and 19), exhibit a sharp peak of mitosis between days 10 and 11, and express LHRH mRNA and immunoreactive LHRH by day 10.5; by day 12.5 all cells that make up the postnatal population of LHRH neurosecretory neurons are present. The cells migrate in a rostralcaudal direction through the nasal septum into the forebrain from day 12.5 to 15.5 along with the terminalis nerve. By day 16.5 the LHRH neurosecretory neurons have a postnatal distribution in the hypothalamus (see Fig. 24-22). LHRH cells were limited to the nasal region in the 36-day monkey embryo.

LHRH neurons in the mouse. The route of migration of the LHRH neurosecretory neurons (black dots) in the mouse embryo is shown from their origin in the medial olfactory placode (a plate-like thickening of embryonic ectoderm) in the nasal region through the forebrain into the hypothalamus and preoptic areas. At embryonic (E) day 11 to 11.5 LHRH cells are in the anlage of the vomeronasal organ and medial wall of the olfactory placode. By E day 13 the number of LHRH neurons has increased, and most are in the nasal septum with the nerves terminals and the vomeronasal nerves; only a few cells are in the brain. By E day 14 the majority of LHRH cells are in the ganglion terminale and the central root of the nerves terminals and arch through the forebrain to the hypothalamus. By E day 16 most of the LHRH neurons are in the hypothalamus and preoptic areas, and the migration is almost complete. GT, ganglion terminalis; OB, olfactory bulb; POA, preoptic area, VNO, vomeronasal organ. (Adapted from Schwanzel-Fukuda M, Pfaff DW. Origin of luteinizing hormone-releasing hormone neurons. Nature 1989; 338:161–164. Reprinted by permission from Nature, Vol. 338, pp. 161-164. Copyright © 1989 Macmillan Magazines Ltd.)

The Human Fetus

Schwanzel-Fukuda and co-workers studied a 19-week gestational male human fetus with Kallmann’s syndrome (see later). No LHRH neurosecretory neurons were detected in the brain, including the hypothalamus. However, dense clusters of LHRH cells and fibers were present in the nose, including the nasal septum and cribriform plate, and within the dural layers of the meninges under the forebrain. The olfactory bulbs were absent. In contrast, normal male fetuses at 19 weeks of gestation had the expected distribution of LHRH neurons in the hypothalamus. In subsequent studies, LHRH immunoreactivity was observed in the epithelium of the medial aspect of the olfactory placode by 42 days of gestation but not at 28 to 32 days. These findings in the human are consistent with the migration of LHRH neurosecretory neurons from the olfactory placode to the hypothalamus in other mammals. Adults have about 1500 to 2000 hypothalamic LHRH neurons. In the mouse, the number of LHRH neurons in the adult is similar to that in late fetal life.

LHRH has been detected in human embryonic brain extracts by 4.5 weeks and in the fetal hypothalamus early in gestation. Further, the fetal pituitary gonadotrophs are responsive to LHRH. The hypotalamic-hypophyseal portal system is functional by 11.5 weeks of gestation. By 9 weeks, LHRH neurons are detectable in the fetal hypothalamus, and by 16 weeks axon fibers that contain LHRH are present in the median eminence and terminate in contact with capillaries of the portal system. In fetal sheep the hypothalamus secretes LHRH in a pulsatile manner. Thus, the available data are consistent with the development of a human fetal hypothalamic LHRH pulse generator by at least the end of the first trimester.

In the previous discussion of hormonal changes, the changing pattern of gonadotropin and gonadal steroid secretion was considered in relation to age. The human fetal gonad is affected by placental gonadotropins and by fetal pituitary FSH and LH. Early in gestation, the placental gonadotropin hCG may play an important role in the secretion of testosterone by the Leydig cells of the fetal testes during the masculinization of the Wolffian ducts and the external genitalia. However, it is uncertain whether functional hCG/LH and FSH receptors are present in the fetal testes by 12 weeks of gestation and whether the early fetal testes responds to hCG. Fetal Leydig cells are a unique population of Leydig cells limited to the fetus and infant, which regress to be followed by the differentiation of adult-type Leydig cells in the peripheral period. In comparison with the adult type, fetal Leydig cells form tightly opposed clusters joined by gap junctions and lack Reinke crystals, are resistant to hCG/LH-induced desensitization (indeed, hCG/LH produces up-regulation of LH/hCG receptors), and contain little aromatase activity and few estradiol receptors.

In contrast to the fetal testis, the fetal ovary apparently does not have to have FSH receptors early in gestation. It is only late in the second trimester, after completion of male phenotypic differentiation, that fetal FSH and LH have been documented to have an effect on the growth and maturation of the fetal testis and ovaries. There is an apparent sex difference in the stage of gestation at which fetal pituitary gonadotropins have an important effect on the development of the fetal gonad. In the anencephalic fetus (which as a consequence of the severe CNS defect is deficient in hypothalamic LHRH, resulting in deficiency of pituitary gonadotropins) the testes appear hypoplastic by early in the third trimester; however, the ovaries in this disorder are normal until at least 32 weeks of gestation.

FSH and LH are detectable in the human fetal pituitary gland by 10 weeks of gestation, and the content increases until approximately 25 to 29 weeks of gestation. The fetal pituitary gland not only can synthesize and store FSH and LH but also can secrete these hormones by 11 to 12 weeks. The fetal serum LH and FSH concentrations rise to peak levels by midgestation and then decrease; the values in umbilical venous blood at term are low (see Fig. 24-24 and Fig. 24-25). Subsequent studies, using fetal blood obtained by cordocentesis and highly sensitive gonadotropin immunoassays, while largely confirming earlier studies have provided new information on the pattern of change in both immunoreactive and bioactive gonadotropins and sex steroids between 17 weeks and term. The serum concentrations of FSH and LH and of bioactive FSH at 17 to 24 weeks of gestation were strikingly higher in female than male fetuses and in both sexes decreased remarkably between 25 and 40 weeks of gestation. Again using highly specific immunoassays of serum FSH and LH in umbilical cord blood from fetuses at 26 to 40 weeks of gestation, the mean FSH and LH concentrations were elevated at the beginning of the third trimester and decreased with
advancing gestational age to undetectable values in term fetuses. The mean FSH value was higher in female fetuses between 26 and 38 weeks, whereas the mean LH level was higher in males. These findings give additional support for the continued secretion of FSH and LH during late gestation but at gradually decreasing amounts.

In the ovine fetus, LH and FSH are secreted in a pulsatile manner in response to the episodic secretion of fetal hypophysial LHRH (Fig. 24-26); human fetal pituitary gonadotropins are probably released in the same mode. The mean FSH and LH content of fetal pituitary glands and the concentration of fetal serum FSH are greater in female than in male fetuses at midgestation. This difference has been ascribed to the higher concentration of plasma testosterone between 11 and 24 weeks in the male fetus (the only major difference in gonadal steroids between the male and female fetus and fetal testicular inhibin), and the decrease in both serum FSH and LH concentrations toward term during late gestation has been attributed to the maturation of the negative feedback mechanism.

the development of gonadal steroid receptors in the hypothalamic-pituitary unit, and the effect of inhibin.

Consistent with this sequence of events, in vitro studies indicate that the human fetal pituitary gland is responsive to LHRH as early as 10 weeks of gestation; the LHRH-stimulated release of LH is greater in second-trimester fetal pituitary cells cultured from females than males and is augmented by estradiol in both sexes. In vivo studies during middle and late gestation demonstrate the stimulating action of exogenous LHRH on fetal FSH and LH release by 16 weeks of gestation with a striking sex difference in the FSH response and a fall in responsivity to LHRH in late gestation (see Fig. 24-24 and Fig. 24-26). The anencephalic infant and some infants with neonatal hypothyroid hypopituitarism have an absent or diminished gonadotropin response to LHRH; in contrast to the brisk increase in gonadotropins elicited by LHRH in the normal infant.

The pattern of changes in FSH and LH concentration in the fetal pituitary glands and serum is consistent with a sequence of increasing synthesis and secretion in which peak serum concentrations reach castrate levels, followed by a decline after midgestation that persists to term. The high serum concentrations of FSH and LH in the female and LH in the male in early and midgestation are probably the result of relatively autonomous, unrestrained activity of the fetal hypothalamic LHRH pulse generator and subsequent stimulation of the fetal gonadotrophs by LHRH. As a consequence of the pulsatile secretion of LHRH, the release of fetal LH and FSH is episodic (see Fig. 24-24 and Fig. 24-26).

TABLE 24-17 — The Early Development of the Human Fetal Pituitary and Hypothalamus

<table>
<thead>
<tr>
<th>Gestational Age (wk)</th>
<th>Hypothalamus</th>
<th>Pituitary</th>
<th>Portal Circulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Forebrain appears</td>
<td>Rathke's pouch in contact with stomodeum</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Diencephalon differentiated</td>
<td>Rathke's pouch separated from stomodeum</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Premammillary preoptic nucleus; LHRH detected</td>
<td>Intermediate-toe primordia; cell cords penetrate mesencephal hypophyseal portal system</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Accute, supraoptic nucleus</td>
<td>Sphenoidal plate forms</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Median eminence differentiated; TRH detected;</td>
<td>Basophilis appear</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Paraventricular nucleus; dorsal medial nucleus</td>
<td>Pars tuberis formed; -endorphin detected;</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Serotonin and noradrenaline detected;</td>
<td>Acidophilis appear</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Mammillary nucleus; primary (hypothalamic) portal plexus present; -endorphin and opioidergic neurons detected;</td>
<td>Secondary (pituitary) portal plexus present catecholamines (IF)</td>
<td>Functional hypothalamic-hypophyseal portal system</td>
</tr>
<tr>
<td>11</td>
<td>Dopamine present</td>
<td>Melanocyte-stimulating hormone detected</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Corticotropic-releasing hormone detected</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Fully differentiated hypothalamus</td>
<td>Adult form of hypophysis developed</td>
<td></td>
</tr>
</tbody>
</table>

FSH, Follicle-stimulating hormone; GH, growth hormone; LHRH, luteinizing hormone-releasing hormone; TRH, thyrotropin-releasing hormone.


*Hormone detected at this gestational age but may be present earlier.

IF, detected by immunofluorescence.

advances, the negative feedback mechanism matures and the hypothalamus secretes less LH, which in turn leads to decreased secretion of FSH and LH. This inhibition of hypothalamic LHRH release and pituitary gonadotropin secretion appears to be a consequence of the increasing sensitivity of the hypothalamus and its LH pulse generator to the inhibitory effects of high concentrations of sex steroids (estrogens and progesterone from the placenta and, in the male, testosterone from the fetal testes) in the fetal circulation and in the male fetus to a contributory effect to the decrease in FSH by luteinizing inhibin in late gestation. The increasing CNS control of gonadotropin secretion seems to require the maturation of gonad steroid receptors (intracellular or on the cell surface or both) in the fetal hypothalamus and in the pituitary gonadotrophs.

The Sheep Fetus

Studies in the human fetus did not provide insight into the mechanisms of maturation or regulation of the hypothalamic LHRH-pituitary gonadotropin gonadal apparatus. The fetal sheep model in which indwelling vascular catheters are placed in the fetus and pregnant ewe affords an opportunity for mechanistic studies. The length of gestation in the sheep is about 145 days. The ontogeny of fetal gonadotropins, hypothalamic LHRH, and gonadal steroids is similar to that in the human fetus (see Fig. 24-26).

Figure 24-23 Ontogeny of the luteinizing hormone-releasing hormone (LHRH) neurons in the rhesus monkey. In the 36-day embryo the LHRH cells (black dots) are located deep in the nasal septum along the path of the nerves to the brain. By day 38 LHRH cells are clustered along the dorsal region of the olfactory bulbs and nerves terminating with a few cells arching back along the ventral surface of the forebrain. By 55 days the LHRH neurons are in the process of migration, but clusters of LHRH cells have entered the central nervous system and reached the basal hypothalamus. BH, basal hypothalamus; LT, lamina terminalis; NV, nasal ventricle; NA, nasal area; NE, nasal epithelium; NT, nervus terminalis; OB, olfactory bulb; OC, optic chiasm; Tu, olfactory tubercle. (Adapted from Ronnekleiv OK, Resko JA. Ontogeny of gonadotropin-releasing hormonecontaining neurons in early fetal development of rhesus macaques. Endocrinology 1990; 126:498511. © by The Endocrine Society.)

Figure 24-24 Ontogeny of the hypothalamic-pituitary portal circulation. Fig. 24-25 Functional hypothalamic-pituitary portal system.
Fetal FSH and LH secretion in the ovine fetus is not autonomous. By midgestation, the secretion of fetal LH and FSH is pulsatile and mediated by the hypothalamic LH-RH pulse generator. The ovine fetal hypothalamic-pituitary gonadotropin unit has the capacity to respond to gonadal steroid negative feedback by 0.6 gestation. A sex difference in gonadotropin secretion occurs in both the ovine and the human fetus as demonstrated by the fact that orchiectomy (but not oophorectomy) in the ovine fetus leads to an increase in pulsatile secretion of LH (and to a lesser degree FSH). Opioidergic neurons have a tonic suppressive effect on the pulsatile release of LH-RH in the fetus. The excitatory amino acid analogue N-methyl-D-aspartate (NMDA) evokes an LH pulse mediated by LH-RH, which provides additional evidence for the functional integrity of the fetal LH-RH neurosecretory neurons and the capacity of the excitatory amino acids, glutamate and aspartate, to stimulate, directly as well as indirectly, the fetal LH-RH pulse generator (Fig. 24-27). Glutamate is present in abundance in the hypothalamus and is released from glutaminergic neurons by excocytosis in an adenosine triphosphate and calcium-dependent process. Furthermore, FSH stimulates inhibit synthesis by the ovine testis and ovary, and administration of an inhibit-rich extract inhibits fetal FSH but not LH secretion, evidence of the functional capacity of the FSH-fetal gonadal inhibin feedback system. These observations in the human and ovine fetus, including the pattern of change of FSH and LH, provide support for an operational hypothalamic LH-RH-pituitary gonadotropin unit by at least 0.3 gestation in the human fetus and 0.4 gestation in the ovine fetus and for the central role of the CNS in this process.

The Human Neonate and Infant

In both sexes, the concentration of plasma FSH and LH is low in cord blood as a consequence of the inhibitory effect of the high levels of placenta-derived estrogens. The hypothalamic regulatory mechanisms for pituitary gonadotropins, as for other pituitary hormones, are not fully developed at birth. Within a few minutes after birth in the male neonate, the concentration of LH increases abruptly in peripheral blood (about 10-fold) compared with that in cord blood. This short-lived surge in LH release is followed by an increase in serum testosterone concentration during the first 3 hours that persists for 12 hours or more. In the female neonate this increase in LH release does not occur; FSH levels during the first neonatal days are low in both sexes. After the fall in circulating levels of steroids of placental origin (especially estrogens) during the first few days after birth, the concentration of serum FSH and LH increases and exhibits a pulsatile pattern with wide perturbations during the first few months.

![Image 1](https://example.com/image1.png)

**Figure 24-24** Comparison of the pattern of change of serum testosterone, human chorionic gonadotropin (hCG), and serum and pituitary luteinizing hormone (LH) (LER-869) and follicle-stimulating hormone (FSH) levels in the human male fetus during gestation in relation to the morphologic changes in fetal testis. The top graph illustrates the regression curve for the increment (between baseline plasma LH and FSH level and the 15-minute response to administration of LH-RH) to the male fetus plotted as a function of gestational age. The scale marks the slight increase in plasma FSH. Data were recalculated from Takagai and colleagues. The evidence supports the hypothesis that the hypothalamic LH-RH pulse generator is functional early in gestation and mediates the rise in serum concentration of fetal pituitary gonadotrophins. To convert plasma hCG to international units per liter, multiply by 1.7. Other conversions are in the legends of Figure 24-17 and Figure 24-18. Modified from Kaplan SL, Grumbach MM. Pituitary and placental gonadotropins and sex steroids in the human and subhuman primate fetus. Clin Endocrinol Metab 1978; 7:487511; and Gluckman PD, Grumbach MM, Kaplan SL. The human fetal hypothalamus and pituitary gland. In Tulchinsky D, Ryan KJ [eds]. Maternal-Fetal Endocrinology. Philadelphia, WB Saunders, 1980, pp 196232.

FSH pulse amplitude is much greater in the female infant and is associated with a larger FSH response to LH-RH throughout childhood. LH pulses are of greater magnitude in the male (Fig. 24-28). This striking sex difference is also present in agonal infant and male infants and in the infant rhesus monkey. It is postulated that this difference in the pattern of pulsatile gonadotropin secretion in infancy is related, among other possibilities, to the effect of testosterone in the male fetus on the development and function of the hypothalamic-pituitary apparatus. The high gonadotropin concentrations (postnatal gonadotropin surge) are associated with (1) a proliferation of Sertoli cells and gonocytes (and their transformation into spermatogonia), (2) a transient second wave of differentiation of fetal-type Leydig cells, and (3) increased serum testosterone levels in male infants during the first few postnatal months and increased estradiol levels intermittently elevated during the first year of life and part of the second year in females. The mean FSH concentration is higher in females than in males during the first few years of life. By approximately 6 months of age in the male and 2 to 3 years of age in the female, the concentration of plasma gonadotropins decreases to the low levels that are present until the onset of puberty. Thus, the restraint of the hypothalamic LH-RH pulse generator is supported by at least 0.3 gestation in the human fetus and 0.4 gestation in the ovine fetus and for the central role of the CNS in this process.

![Image 2](https://example.com/image2.png)

**Figure 24-25** Pattern of change of serum follicle-stimulating hormone (FSH), luteinizing hormone (LH), and human chorionic gonadotropin (hCG) levels; concentration of pituitary FSH and LH; and increment (between baseline plasma LH and FSH level and the 15-minute response to administration of LH-RH) to the human male fetus during gestation with the development of the fetal ovary. See legends of Figure 24-17, Figure 24-18, and Figure 24-24 for conversions to SI units. Modified from Kaplan SL, Grumbach MM. Pituitary and placental gonadotropins and sex steroids in the human and subhuman primate fetus. Clin Endocrinol Metab 1978; 7:487511.)

and the suppression of pulsatile LH-RH secretion (and thus LH release) do not attain the prepubertal level of quiescence until late infancy or early childhood and earlier in boys than girls.

Although the rise in circulating gonadotropins, sex hormones, and inhibin in both sexes during infancy, the so-called postnatal surge, is quite well documented, the significance of this transitional state before the onset of the juvenile pause remains speculative. The increase in circulating testosterone derived in the normal male infant can be accompanied by facial comedones and even acneiform lesions and the increase in gonadotropins by a transient increase in testicular size, but there may be more subtle changes. For example, Sertoli cell

![Image 3](https://example.com/image3.png)

**Figure 24-26** Pulsatile luteinizing hormone (LH) secretion in the ovine fetus. GA, gestational age. The length of gestation is 145 days in the sheep. (From Clark SL, Ellis N, Styne DM, et al. Hormone ontogeny in the ovine fetus. XVII. Demonstration of pulsatile luteinizing hormone secretion by the fetal pituitary gland. Endocrinology 1984; 115:17411779. © by The Endocrine Society.)

![Image 4](https://example.com/image4.png)

**Figure 24-27** Left, The effect of the ovine fetus of administration for 7 days of luteinizing hormone-releasing hormone (LH-RH) agonist (10 µg intravenously daily) on the acute LH response to LH-RH analog; right, Recovery of the LH response was impaired 8 days after discontinuing LH-RH agonist administration to the ovine fetus. (From Grumbach MM, Kaplan SL. The neuroendocrinology of human puberty: an ontogenetic perspective. In Grumbach MM, Sizonenko PC, Aubert ML, [eds]. Control of the Onset of Puberty. Baltimore, Williams & Wilkins, 1990, pp 168. © 1990, the Williams & Wilkins Co., Baltimore.)
and germ cells proliferate for about 100 days after birth and then decline after about 6 months of age, coincident with the waning of gonadotropins and testosterone. The significance of this increase in Sertoli cells and spermatogonia in relation to adult fertility is not known. It has been speculated that the postnatal testosterone may have an imprinting effect on the CNS, but we are not aware of convincing evidence to support this conjecture.
Neural Control

The neural control of puberty involves two major factors: the timing of puberty and the mechanisms involved in control of the transition from the prepubertal or sexually infantile state through complete sexual maturation.

Timing and Onset of Puberty

The time of onset of puberty and its course are influenced by genetic factors and are modified by environmental factors operating through the CNS. The latter include socioeconomic factors, nutrition, general health, geography, and altitude. Although the action of multiple genes (quantitative or polygenic inheritance) on the time of onset of puberty (or, for example, on stature) has long been recognized, little is known about the gene loci involved in this complex quantitative trait or the effect of gene interactions (epistasis) on this paradigm of complex traits.

The specific mechanisms involved in the timing of puberty are complex. Frisch and Revelle have suggested that in healthy girls, despite different ages, there is an "invariant mean weight" (48 kg) for the initiation of the pubertal spurt in weight, the maximal rate of weight gain, and menarche and that this association is primarily related to fatness. The Frisch hypothesis has generated controversy and has been criticized by some authorities. In part because the empirical estimations and the equations used to determine fat mass have been challenged. The role of nutritional factors and body composition in the onset of menarche is supported by the earlier age of menarche in moderately obese girls; delayed menarche in states of malnutrition and chronic disease, in twins, and after early rigorous athletic or ballet training; and the relationship of weight and diminished body fat to changes in gonadotropin secretion and amenorrhea in girls with anorexia nervosa, voluntary weight loss, and strenuous physical conditioning. Long-term studies of girls who had had malnutrition in infancy suggest that no permanent delay in puberty persists after early treatment.

The proposed causal relationship of "critical body weight," "critical metabolic rate," and body fat to the time of onset of puberty in girls has not been substantiated by direct measurements. When increments in the excretion of urinary gonadotropins were correlated with changes in body composition at puberty, both developmental events occurred simultaneously rather than sequentially. Moreover, menarche is a late event in the pubertal process and is remote from the factors that initiate the hormonal events and the first signs of sexual maturation. Thus, the results of a longitudinal study relating the beginning of testicular and breast development to percentage of body fat are of particular interest; although puberty began at different chronologic ages in the 469 girls studied for 5 years, there was a similar percentage of body fat associated with the onset of puberty. Other studies reported no change in body fat or body composition at the time of menarche that accompany the change in hormone production; there are, however, substantial changes in early puberty and premenarche in girls.

It has long been postulated that some alteration of body metabolism linked to energy metabolism may affect the CNS restraints on pubertal onset and progression. The relationship of adipose tissue mass, fat metabolism, and energy balance to reproduction was illuminated by the discovery of the genes encoding leptin, an 167-amino-acid cytokine-like protein produced mainly, but not exclusively, by adipose tissue. Its receptor, a member of the gp family of cytokine receptors, occurs in several isoforms, of which in the hypothalamus and other areas of the brain only the leptin receptor splice variant with the long intracellular domain (Ob-Rb) contains the protein motifs that possess signaling properties. In the human, leptin circulates in both a free and a high-molecular-weight bound form. Leptin is secreted in a pulsatile fashion and exhibits a diurnal rhythm with a peak at night and a nadir in the morning. Leptin binding activity in serum is highest in childhood and decreases to relatively low levels during puberty. Considerable interest has focused on the potential role of leptin in the control of the onset of puberty from a proposal that it was an essential, if not a key, factor in triggering the onset of puberty to one in which it had a more subsidiary role.

Leptin is a highly conserved 167-amino-acid cytokine-like protein produced mainly, but not exclusively, by adipose tissue. Its receptor, a member of the gp family of cytokine receptors, occurs in several isoforms, of which in the hypothalamus and other areas of the brain only the leptin receptor splice variant with the long intracellular domain (Ob-Rb) contains the protein motifs that possess signaling properties. In the human, leptin circulates in both a free and a high-molecular-weight bound form. Leptin is secreted in a pulsatile fashion and exhibits a diurnal rhythm with a peak at night and a nadir in the morning. Leptin binding activity in serum is highest in childhood and decreases to relatively low levels during puberty. Considerable interest has focused on the potential role of leptin in the control of the onset of puberty from a proposal that it was an essential, if not a key, factor in triggering the onset of puberty to one in which it had a more subsidiary role.

Leptin is a well-established afferent saliety factor in the human that acts on the hypothalamus, including nuclear controlling appetite, to suppress appetite. In rodents it stimulates energy expenditure among a variety of other actions. A mutant gene was first isolated from the extremely obese ob/ob mouse by positional cloning; soon thereafter, the leptin receptor was cloned and the mutation identified in the obese db/db mouse.

Ob/ob and db/db mice have the same phenotype. They are not only obese but also fail to achieve puberty and fertility because of hypogonadotropic hypogonadism, providing evidence for an important role of leptin in

[Figure 24-29 Change in the pattern of pulsatile follicle-stimulating hormone (FSH) and luteinizing hormone (LH) secretion in early infancy, childhood, and puberty. The data for early infancy are derived from Walther and colleagues. Note the pulsatile secretion in the infant and the striking difference in the amplitude of FSH and LH pulses between male and female infants. After infancy, the amplitude and frequency of gonadotropin pulses decrease greatly for almost a decade (juvenile pause) until the onset of puberty. (From Grumbach MM, Kaplan SL. The neuroendocrinology of human puberty: an ontogenetic perspective. In Grumbach MM, Sosenko PC, Aubert ML, eds. Control of the Onset of Puberty. Baltimore, Williams & Wilkins, 1990, pp 168. © 1990, the Williams & Wilkins Co., Baltimore.)]

[Figure 24-30 The pulsatile action of leptin secreted by adipocytes on the hypothalamic luteinizing hormone-releasing hormone (LHRH) pulse generator. Its indirect action through hypothalamic neural networks is illustrated as well as direct action. Leptin appears to function as a permissive factor, not a trigger, in the onset of human puberty. Although leptin is reported to advance puberty in rodents, its role in "triggering" puberty in humans has not been established and is speculative. FSH, follicle-stimulating hormone; mRNA, messenger ribonucleic acid. (See text.)]
reproduction. This association was strengthened by the collaboration of the infertility in ob/ob mice by the administration of recombinant leptin. The first reports on the effect of leptin administration to normal prepubertal mice described decreased appetite and body weight and the advancement of vaginal opening, but the concentration of serum leptin in normal mice did not increase with the onset of and during sexual maturation. Studies in the rat showed that pubertal delay associated with food restriction was partially reversed by the administration of leptin and suggested that a fall in circulating leptin below a critical level would occur with 30% to 75% food restriction. However, leptin administration to normal prepubertal rats did not advance the time of onset of puberty. Cheung and colleagues summarizing the studies in rodents, concluded that although a critical threshold level was necessary for puberty to begin and advance, leptin alone (as in leptin administration to normal rodents) was insufficient to promote puberty.

Among the various sites of action of leptin in the hypothalamus, one site seems to involve a direct action on hypothalamic LHRH neurons: in the rodent these neurosecretory neurons contain leptin receptors (including the Ob-Rb isoform) and release LHRH in response to leptin. Leptin may also act indirectly on the LHRH pulse generator (Fig. 24-30).

There are significant species differences in the role of leptin in reproductive function not only among rodents but also between human and nonhuman primates. In the male rhesus monkey, leptin levels were similar through the advancement of prepuberty to puberty. In the peripheral rhesus monkey 3 to 5 years old, fasted for 2 days, the administration of leptin prevented the decrease in plasma gonadotropins detected in the untreated animals. Continuous infusion of leptin into the lateral ventricle of gonadial male monkeys failed to evoke an increase in LHRH on gonadotropin secretion.

Cheung and colleagues have reassessed the role of leptin in the onset of puberty in the mouse and rat. In well-designed studies, they reported that leptin is not a metabolic trigger for onset of puberty in the rodent but is one among several permissive factors. This assessment is supported by the studies of Bronson and Risman in female mice. Does leptin in the human provide a peripheral, somatic trigger for the onset of puberty to the CNS, or does it have a permissive role, signaling the hypothalamus and the LHRH pulse generator that a critical energy store (obesity) has been attained? Leptin reflects body fat and hence energy stores and has an important role in the control of body weight and the regulation of metabolism. In 1997 the first longitudinal study of plasma leptin in normal boys before and during puberty suggested that there was a brief rise in circulating leptin that provided hormonal evidence of the onset of puberty. Later, two large cross-sectional studies and a longitudinal study of serum leptin levels in prepubertal and pubertal boys and girls showed that leptin increased gradually during the prepubertal years; the levels were similar in both sexes. However, during puberty, leptin continued to rise in girls, whereas in boys the leptin mean levels peaked at Tanner stage 2 and decreased to prepubertal concentrations by genital stage 5. The decrease was attributed to the effect of testosterone on leptin secretion.

Among other variables, adipose tissue mass, percentage body fat, and age were correlated with leptin levels. A correlation was not found between 24-hour serum estradiol and leptin concentrations in nonobese and obese prepubertal and early pubertal girls. People with a homozygous mutation in the leptin gene or the leptin receptor have not only morbid obesity but also a striking delay in puberty owing to hypogonadotropin hypogonadism. Two 19-year-old sisters with a mutation in the leptin receptor that resulted in a truncated leptin receptor were sexually infantile as a consequence of hypogonadotropic hypogonadism. Similarly, in a large Turkish pedigree affected by a stop codon mutation in the gene encoding leptin, a 23-year-old man had failed to attain puberty because of hypogonadotropic hypogonadism. The two affected women were prepubertal and amenorrheic until ages 29 years and 36 years, respectively, when the 23-year-old began to have irregular scant periods and the 36-year-old began to menstruate monthly. A 3-year-old girl with congenital leptin deficiency following treatment with recombinant leptin lost weight and had an early pubertal pattern of LH release to the administration of LHRH. The effects of the mutations indicate that the virtual absence of leptin or of a functional leptin receptor leads to sexual infantilism related to hypogonadotropic hypogonadism, similar to that in ob/ob and db/db mice. The extremely late signs of pubertal development in the two leptin-deficient women are unexplained.

These observations suggest that severe leptin deficiency causes hypogonadotropic hypogonadism and that a critical level of leptin and a leptin signal are required to achieve puberty. However, a rise in leptin is not required to trigger puberty. Boys with a constitutional delay in growth and puberty can enter puberty without an increase in circulating leptin; these boys have lower than expected mean levels of leptin and weight. Further, in two women with congenital lipodystrophic diabetes (Berardinelli-Seip syndrome), which is associated with absence of both subcutaneous and visceral adipose tissue, the severe hypoleptinemia did not lead to a delay in menarche. One of the women had three unaffected children.

In sum, accumulating evidence supports the function of leptin as a permissive factor (tonic mediator) and not a trigger (phasic mediator) in the onset of human puberty (Table 24-18). A major qualitative locus on human chromosome 2 is linked to serum leptin levels and fat mass (the structural gene that encodes leptin is located on chromosome 7), an additional indication that a variety of factors affect leptin. Further, in relation to puberty as an energy-dependent process, a group of 9- to 10-year-old boys observed for 18 months, during the months leading up to an increase in morning salivary testosterone concentrations, had a relatively constant basal metabolic rate (BMR)/lean body mass ratio and an increase in the ratio of BMR/total daily energy expenditure. It is suggested that a subtle energy-dependent process is in play, possibly related to an increase in brain BMR as a secondary phenomenon at the initiation of puberty, or that a central rise in BMR is a signal for the onset of puberty. The timing of puberty has been linked to the vagus, but generally accepted, concept of maturation of the CNS; the maturation is the outcome or consequence of the totality of environmental and genetic factors that retard or accelerate the onset of puberty. It is a provocative but unproven hypothesis that a metabolic signal related to body composition is an important factor in the maturation or activation of the hypothalamic LHRH pulse generator and not a result of the early hormonal and body composition changes in human puberty. In either event, clinical and experimental data support the contention that the factors influencing the timing of puberty are expressed finally through CNS regulation of the onset of puberty.

Mechanisms of Control

In some species, the neural control mechanism is exquisitely sensitive to the environment; for example, in the rat and the mouse, eurexcorporate factors and cues, including light, olfaction, and pheromones, have an important influence, by way of the CNS, on gonadotropin secretion. In seasonal breeding species,
such as sheep, the length of the light-dark cycle is critical and the pattern of gonadotropin secretion is different. In contrast, male and female primates exhibit an estrogen-provoked LH surge. In brief, diverse strategies and adaptive mechanisms have evolved to control puberty in different species.

In both the human and the subhuman primate, the increase in LH and FSH secretion in early infancy is followed by a...
presence of a CNS inhibitory mechanism that, independent of gonadal steroid secretion, restrains the hypothalamic LHRH pulse generator during this pause.

This mechanism suppresses LHRH and as a consequence gonadotropin synthesis and pulsatile secretion and restrains the onset of puberty. The fall in gonadotropin secretion in agonadal children cannot be explained by gonadal steroid feedback (because functional gonads are lacking) or by increased secretion of adrenal steroids (because concentrations are low and glucocorticoid suppression of the adrenal does not augment the concentration of circulating gonadotropins). Therefore, a steroid-independent inhibitory mechanism for suppression of the hypothalamic LHRH pulse generator, located within the CNS, seems to be the dominant factor in restraint of puberty between ages 4 and 11. Gradual loss of this intrinsic CNS inhibitory mechanism would lead to disinhibition or reactivation of the LHRH pulse generator at puberty.

Interaction of Negative Feedback Mechanism and Intrinsic Central Nervous System Inhibitory Mechanism

We believe that both of these mechanisms interact to restrain puberty (see Fig. 24-34). During the first 2 to 3 years of life, the gonadal steroid negative feedback mechanism seems dominant, as evidenced by the striking difference in gonadotropin secretion between the agonal and the intact infant and young child. Extrapolating from the changing pattern of plasma FSH and LH levels in agonal infants and children, beginning at about 3 years of age the intrinsic CNS inhibitory mechanism becomes dominant and remains so during the rest of the juvenile pause, as evidenced by the fall in FSH and LH levels between ages 3 and 10 despite the lack of functional gonads. During this segment of the juvenile pause, the negative feedback mechanism is operative; agonal patients in this age group have higher mean plasma FSH levels than normal prepubertal children and a greater FSH and LH response to the acute administration of LRH. However, the negative feedback mechanism probably plays a secondary role.

As puberty approaches, the CNS inhibitory mechanism gradually wanes, initially during nighttime sleep, and the hypothalamic LHRH pulse generator becomes less sensitive to gonadal steroid negative feedback (Fig. 24-35). After the onset of puberty, gonadal steroid negative feedback attains the setpoint characteristic of the adult and is again the dominant mechanism in restraining gonadotropin secretion (along with inhibin), as reflected in the increased gonadotropin concentrations characteristic of the adolescent with severe primary hypogonadism (see Fig. 24-34). A similar pattern has been described in the infant monkey. The postulated ontogeny of this dual mechanism of restraint of puberty is illustrated in Figure 24-35. Many neural, neurotransmitter-neuromodulator, hormonal, growth, and metabolic factors as well as exteroceptive influences and cues can influence the activity of the LHRH pulse generator, but the nature of the intrinsic CNS inhibitory mechanism remains uncertain.

Indirect evidence for an inhibitory neural network that arises or projects through the posterior hypothalamus and suppresses the LHRH pulse generator has been derived from studies of children with organic forms of true (or central) precocious puberty. For example, a suprasellar arachnoid cyst can cause true precocious puberty by compressing and distorting the hypothalamus. In some children with such cysts, the puberty is reversed with regression of the hormonal and physical features of puberty after decompression of the cyst (Fig. 24-36). We suggest that the disinhibition of the CNS inhibitory mechanism was reversed by treatment of the cyst. The LHRH-secreting hypothalamic hamartoma, a heterotopic mass of nervous tissue that contains LHRH neurosecretory neurons, attached to the tuber cinereum or the floor of the third ventricle, can cause true precocious puberty. In children with such cysts, the puberty is reversed with regression of the hormonal and physical features of puberty after decompression of the cyst. In the rhesus monkey, despite the damping of the LHRH pulse generator during the juvenile pause, the content of hypothalamic LHRH and the LHRH mRNA during this phase is similar to that in the infant and adult monkey. It should be emphasized that quiescence of the LHRH pulse generator during the juvenile pause is not absolute. Low-amplitude LH and FSH pulses are detectable by sensitive and specific immunoradiometric assays. The end of the juvenile pause is marked by an increase in LH pulse amplitude most evident during the early hours of sleep.

Potential Components of the Intrinsic Central Nervous System Inhibitory Mechanism

Indirect evidence for an inhibitory neural network that arises or projects through the posterior hypothalamus and suppresses the LHRH pulse generator has been derived from studies of children with organic forms of true (or central) precocious puberty and studies in the female and male monkey. Children with true precocious puberty associated with posterior hypothalamic neoplasms (usually a pilocytic astrocytoma), radiation of the CNS, midline CNS developmental abnormalities such as septo-optic dysplasia with deficiency of one or more pituitary hormones, or other CNS lesions provide indirect evidence for an inhibitory neural component located in or projecting through the posterior hypothalamus. As a consequence of these lesions, the neural pathway inhibiting the hypothalamic LHRH pulse generator is compromised, resulting in its disinhibition and activation. For example, a suprasellar arachnoid cyst can cause true precocious puberty by compressing and distorting the hypothalamus. In some children with such cysts, the puberty is reversed with regression of the hormonal and physical features of puberty after decompression of the cyst. We suggest that the disinhibition of the CNS inhibitory mechanism was reversed by treatment of the cyst. The LHRH-secreting hypothalamic hamartoma, a heterotopic mass of nervous tissue that contains LHRH neurosecretory neurons, attached to the tuber cinereum or the floor of the third ventricle, can cause true precocious puberty. The LHRH neurons within the hamartoma with their axon fibers projecting to the median eminence secrete LHRH in pulsatile fashion. We consider the hypothalamic hamartoma an "ectopic LHRH pulse generator" that functions independently of the CNS inhibitory mechanism that normally restrain the hypothalamic LHRH pulse generator. An analogy can be drawn between the LHRH-secreting hypothalamic hamartoma and the rescue of fertility in the LHRH-deficient hypogonadal mouse (hyp/hyp) by transplantation of fetal or neonatal hypothalamic tissue into the third ventricle. Some rare, large hypothalamic hamartomas that cause true precocious puberty contain few or no LHRH neurosecretory neuron but contain TGF-. It is postulated that the secretion of TGF-, an astroglia-derived growth factor, may interact directly or indirectly to stimulate LHRH release.

Moreover, the ontogeny of the fetal LHRH pulse generator suggests that its initial unrestrained function is followed by differentiation of inhibitory mechanisms in late gestation. Similarly, the immortalized LHRH neurosecretory neuronal cell line exhibits spontaneous, synchronized autorhythmicity in the release of LHRH. Taken together, these observations suggest that a stimulatory input is not required for pulsatile LHRH secretion. In addition, precocious sexual maturation can be induced in the juvenile female rhesus monkey by posterior hypothalamic lesions; such lesions advance the age at onset of a pubertal increase in LH secretion and the time of the first positive feedback effects of estrogen.

Table 24-19 lists some of the neural and neurotransmitter-neuromodulator factors that may play a role in the restraint of the LHRH pulse generator during the juvenile pause. Noradrenergic, dopaminergic, serotoninergic, and opioidergic pathways; inhibitory neurotransmitters (e.g., GABA); excitatory amino acids (e.g., glutamic and aspartic acids); nitrergic transmitters; other brain peptide, including neurotrophic and growth factors; and corticotropin-releasing hormone affect the hypothalamic LHRH pulse generator.

The studies of Plant and of others in the human exclude melatonin as a critical restraining factor in primates. Many studies have assessed the role of endogenous opioid peptides as possible mediators of the juvenile pause. None provide support for an important role of this family of neuro peptides in the juvenile pause. A critical and landmark advance in our understanding of the nature of the juvenile phase and central inhibition of the
The onset of puberty in rhesus monkeys is characterized by a decrease in GABAergic (and possibly NPY) inhibition of excitatory NMDA amino acid neurotransmitter receptors. These receptors are widely distributed throughout the CNS, including the hypothalamus. Among the facilitory neurotransmitters that affect the release of LHRH, norepinephrine, NPY, and galanin do not appear to have a critical role in the control of the developmental switch of GABAergic synaptic transmission from excitatory to inhibitory. The results suggested that there is a potent inhibition by GABA and GABAergic neurons of the LHRH pulse generator in prepuberty and that exogenous administration of GABA in prepuberty is ineffective because of the high local endogenous GABA levels. Glutamic acid decarboxylase (GAD) is the enzyme that catalyzes the conversion of glutamate to GABA. Two classes of GAD, GAD 65 and GAD 67, are present in mammalian brain. Both GAD mRNAs are detectable in the mediobasal hypothalamus, the site of the LHRH pulse generator. Moreover, antisense oligodeoxynucleotides for GAD 67 and GAD 65 mRNA s inhibited the stimulation of LHRH release whereas nonsense o-oligodeoxynucleotides did not.

The slowest maturation of the hypothalamic LHRH pulse generator is associated with the dominance of the intrinsic CNS inhibitory mechanism and results in true precocious puberty. Two possible mechanisms are proposed. The effect of GABA quite likely has a direct action on the LHRH pulse generator neuron as GABA acting through both GABA_A and GABA_B receptors affects LHRH secretion in the perfused mouse GT1 LHRH-releasing neuronal cell line, which contains GABA_A receptors. Conversely, the chronic repetitive administration of bicuculline into the base of the third ventricle of a prepubertal monkey at 15 months of age caused premature menarche and the onset of the first ovulation.

GABA is the principal inhibitory neurotransmitter in the juvenile and adult brain, but early in brain development extending (at least in the experimental animal) through the postnatal period, GABAergic synaptic transmission is excitatory and increases the intracellular 

### Table 24-19 -- Potential Components of the Intrinsic Central Nervous System Inhibitory Mechanism ("Juvenile Pause")

<table>
<thead>
<tr>
<th>1. Inhibitory</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Inhibitory central neurotransmitter-neuromodulatory pathways</td>
</tr>
<tr>
<td>1. -Aminobutyric acid (the main inhibitory factor)</td>
</tr>
<tr>
<td>2. Endogenous opioid peptides</td>
</tr>
</tbody>
</table>
the hypothalamic LHRH pulse generator and increased release of glutamate, the major excitatory amino acid neurotransmitter in the hypothalamus. Plant and associates induced pubertal or adult-like patterns of episodic LHRH release in both the prepubertal and pubertal female rhesus, NMDA administered centrally and peripherally induced the release of LHRH. The sensitivity to the stimulatory glutaminergic input into the LHRH pulse generator increases strikingly after the onset of puberty, but it is the reduction in GABAergic inhibition that is the critical factor in disinhibition of the LHRH pulse generator. Parenthetically, NMDA administration acutely stimulated the LHRH pulse generator in the ovine fetus, providing additional evidence for the role of NMDA receptors in fetal LHRH neurons and the functional capacity of the fetal pulse generator. Immortalized LHRH neurons contain inotropic NMDA receptors that mediate the release of LHRH by NMDA.

In sum, with the onset of puberty, the disinhibition and reactivation of the LHRH pulse generator are associated with a fall in GABAergic inhibitory neurotransmission and a concomitant increase in the input of excitatory amino acid neurotransmitters (e.g., glutamate) as well as other neurotransmitters (e.g., norepinephrine and NPY). Ojeda and Ma have proposed that in addition to the transsynaptic mechanisms, astrocytes regulate LHRH neurosecretion through astroglia-derived growth factors (including transforming growth factors, IGF-I, basic fibroblast growth factor, neuregulins, and epidermal growth factorked peptides that act directly or indirectly to stimulate LHRH release and, accordingly, may contribute to the reactivation of the LHRH pulse generator at puberty. These observations provide additional evidence that the hypothalamic LHRH neurosecretory neuron is not a limiting factor in puberty. The LHRH pulse generator joins the anterior pituitary gland, gonads, and gonadal steroid end organs as elements that are functionally intact prepuber tally and in the fetus and can be fully activated before puberty as well as in the fetus by the appropriate stimulus. Hence, the CNS restraint of puberty lies above the level of the autorhythmic LHRH neurons or hypothalamus and extends to the area postrema, which contains LHRH neurons and regulates sleep. Sleep-Related Luteinizing Hormone Release and Onset of Puberty

Episodic, or pulsatile, secretion is the fundamental mode of release of pituitary LH and FSH and is evoked by the pulsatile LHRH signal originating from the hypothalamic LHRH oscillator. Discrete episodic bursts of LH release occur approximately once every 120 minutes (about 12 episodes over a 24-hour period) in adult men and about once every hour during the midfollicular phase in women. In sensitive radio-immunoassays, secretory pulses of LH are detectable in prepubertal children and are of lower amplitude than those in pubertal children or adults. The low concentrations of plasma LH and FSH make it difficult to demonstrate pulsatile secretion for methodologic and statistical reasons, but this problem has been overcome by the use of highly sensitive immunoassays.

In adult men and women during most phases of the menstrual cycle, little difference in the amplitude or frequency of these episodic pulses is apparent during a 24-hour period. In pubertal children, however, Boyar and colleagues described the mainly sleep-associated pulsatile release of LH in early puberty and midpuberty; only in late puberty were prominent LH secretory episodes detected during the day, although the absolute differences were small. In prepubertal, early, and midpubertal, and even prepubertal children, an increase in the amplitude of LH pulses occurs during sleep (Fig. 24-40); in late puberty, the daytime LH pulses increase in amplitude but are still less than during sleep until the adult pattern is finally achieved.

The infant exhibits episodic gonadotropin secretion with a striking sex difference: the amplitude of the pulses is large and correlates with the increased plasma gonadotropin levels during the first 6 months in boys and the first 1 to 2 years in girls. After this age, pulsatile secretion is more difficult to detect before the pubertal period but is demonstrable at low amplitude mainly at night; a diurnal rhythm of serum LH, FSH, and testosterone is already demonstrable in 5- to 6-year-old children and becomes more prominent in boys with advancing bone age. In boys, augmented LH release during sleep leads to increased excretion of urinary LH in pubertal children at night than during the day, although the absolute differences were small. In prepubertal, early and midpubertal, and even prepubertal children, an increase in the amplitude of LH pulses occurs during sleep (Fig. 24-40). This pattern of sleep-associated LH secretion occurs in agonalad patients during the pubertal age period, suggesting that it is not dependent on gonadal function. Furthermore, augmented sleep-related gonadotropin release is demonstrable in children with idiopathic true precocious puberty, and in glucocorticoid-treated children with congenital adrenal hyperplasia who have an advanced bone age and early onset of true puberty.

Sleep-enhanced LH secretion can be viewed as a maturational phenomenon related to changes in the CNS and in the hypothalamic restraint of LH release. However, the neural factors involved in the initiation of this circadian rhythm are unclear. Episodic release of gonadotropins is suppressed by anti-LHRH antibodies and by the administration of gonadal steroids or of certain catecholaminergic agonists and antagonists and is augmented by the opioid antagonist naloxone. Naloxone does not alter the testosterone-mediated suppression of LH, nor does it alter the testosterone effects on LH pulsatility in early to midpubertal boys. We have suggested that an increase in endogenous LHRH secretion at puberty has a priming effect on the gonadotroph and leads to increased sensitivity of the pituitary to LHRH (either endogenous or exogenous). In the monkey, a striking increase in the pulse amplitude and a lesser increase in pulse frequency occur between prepuberty and puberty. Sleep-associated LH release in the peripubertal period correlates with the increased sensitivity of the pituitary gonadotrophs to administration of LHRH in the peripubertal period and puberty.
The increased LH release at night in both sexes is evidence that the hypothalamic LHRH pulse generator is initially less inhibited during sleep even in prepubertal children. The much augmented LH pulse amplitude entrained during sleep is the neuroendocrine hallmark of the onset of puberty.

Puberty encompasses orderly maturational changes that involve, sequentially, the extramedial basal hypothalamus, the hypothalamic LHRH pulse generator, the pituitary, the gonads, and the gonadal steroid target organs. At each level these structures may exhibit differences in responsiveness to neural or trophic stimuli, depending on their sensitivity and on the particular hormonal milieu. If the increased secretion of gonadotropins at the beginning of puberty is a consequence of changes in both neural and hormonal restraints on the synthesis and pulsatile secretion of LHRH, disinhibition and reaugmentation of the LHRH pulse generator should lead to increased amplitude and frequency of pulses initially, followed by priming of the gonadotrophs, increased pulsatile gonadotropin secretion from the pituitary, and finally augmented output of steroids by the gonad.

LHRH release is not directly measurable in the human, but endogenous LHRH secretion can be estimated indirectly and qualitatively by determining the pulsatile pattern of LH and by the gonadotropin response to exogenous LHRH. The pituitary sensitivity to synthetic LHRH and the dynamic reserve or readily releasable pool of pituitary gonadotropins have been studied at different stages of the menstrual cycle, in various disorders of the hypothalamic-pituitary-gonadal system. The results support the concept that the prepubertal state is characterized by functional LHRH deficiency. The release of LH after administration of LHRH is minimal in prepubertal children before infancy, increases during the peripubertal period and puberty, and is still greater in adults (depending on the phase of the menstrual cycle in women). The change with maturation in the pattern of FSH release is different from that of LH and results in a striking reversal of the FSH/LH ratio after administration of LHRH to both males and females beyond infancy, increases during the peripubertal period and puberty, and is still greater in adults (Fig. 24-41), and is still greater in adults (see Fig. 24-41). Moreover, there is a sex difference in the FSH response, with pubertal and prepubertal females releasing more FSH than males at all stages of sexual maturation.

These observations suggest a striking change in pulsatile sensitivity in LHRH to prepubertal and pubertal individuals as well as a sex difference in the "dynamic reserve" of pituitary FSH. The sex difference in LH and FSH response to LHRH suggests that the pituitary gonadotrophs of prepubertal females are more sensitive to LHRH than those of prepubertal males, even though the concentration of circulating gonadal steroids is low in both sexes at this stage of maturation. Prepubertal girls have a larger readily releasable pool of pituitary FSH than prepubertal or pubertal males, possibly related in part to the higher concentration of inhibin B in pubertal boys (see Fig. 24-41). The sex difference in sensitivity to LHRH and releasable FSH and the low inhibin levels in the prepubertal female may be factors in the higher frequency of idiopathic true precocious puberty in girls and in the occurrence of premature thelarche. The available data are consistent with the hypothesis that less LHRH is required for FSH than for LH release. These findings also point out the difference between pulsatile sensitivity and the actual secretory rate of FSH and LH.

The responses to LHRH in peripubertal children who do not yet exhibit physical signs of sexual maturation provide evidence that the self-priming effect of endogenous LHRH augments pulsatile responsiveness to exogenous LHRH and is an important factor in the increased gonadotropin secretion at puberty. This change in responsiveness of the gonadotrophs is apparently mediated by increased pulsatile secretion of LHRH and the increased LH response to synthetic LHRH is one of the earliest hormonal markers of puberty onset.

The degree of previous exposure to gonadotropes to endogenous LHRH appears to affect both the magnitude and the quality of LH responses to a single intravenous dose of LHRH. Studies of the effects of acute and chronic administration of synthetic LHRH in hypergonadotropic hypogonadism, hypogonadotropic hypogonadism, constitutional delayed growth and adolescence, and idiopathic precocious puberty support this concept of self-priming. The prepubertal pituitary gland has a smaller pool of releasable LH and decreased responsiveness to the acute administration of synthetic LHRH. With the approach of puberty, the derepression of the hypothalamic LHRH pulse generator and the increased pulsatile secretion of LHRH augment pulsatile sensitivity to LHRH and enlarge the reserve of LH. The reason for the discordance in FSH and LH release prepubertally is not clear, but the frequency of LH pulses may be a

![Figure 24-40](image-url) Plasma luteinizing hormone (LH) and testosterone sampled every 20 minutes in a 14-year-old boy in pubertal stage 2. The histogram displaying sleep stage scores is depicted above the period of nocturnal sleep. Sleep stages are rapid eye movement (REM) stages I to IV shown by depth of line graph. Plasma LH is expressed as mIU/mL. Plasma testosterone is expressed as nanograms per 100 mL. To convert LH values to international units per liter, multiply by 1.0. To convert testosterone values to nanomoles per liter, multiply by 0.3467. (From Buxar RM, Rosendahl RD, Kaplan S, et al: Human puberty: simultaneous augmented secretion of luteinizing hormone and testosterone during sleep. Reproduced from the Journal of Clinical Investigation, 1974, vol. 54, pp. 609576 by copyright permission of the American Society for Clinical Investigation.)

![Figure 24-41](image-url) Changes in plasma luteinizing hormone (LH) (top) and follicle-stimulating hormone (FSH) (bottom) levels in prepubertal, pubertal, and adult individuals. Note the limited LH response in prepubertal children compared with that of pubertal and adult subjects. The FSH response to LH-releasing hormone (LHRH) is similar in prepubertal, pubertal, and adult males. In females, the FSH response is significantly greater than that of prepubertal, pubertal, or adult males. For conversion to SI units, see the legend of Figure 24-17. (Modified from Grumbach MM, Roth JC, Kaplan SL, et al: Hypothalamic-pituitary regulation of puberty in man: evidence and concepts derived from clinical research. In Grumbach MM, Gravh DD, Mayer FE [eds]: Control of the Oesest of Puberty. New York, John Wiley & Sons, 1974, pp. 115166.)
factor in the adult rhesus monkey with ablative hypothalamic lesions that eliminate endogenous LHRH secretion, reduction of the frequency of exogenous LHRH pulses from one per hour to one every 3 hours increased the FSH/LH ratio. Furthermore, inhibin and endogenous gonadal steroids may also affect this ratio through action on the hypothalamus, the pituitary gland, or both.

These observations and the previously discussed role of the intermittence of the LHRH signal to the gonadotrophs as an essential factor in the neural control of gonadotropin secretion have important implications for the induction of puberty. Pulsatile administration of LHRH to prepubertal monkeys promptly initiated puberty (and, in females, ovulatory menstrual cycles) and restored complete gonadal function in adult monkeys with hypothalamic lesions. Similar studies in the human yielded comparable results in prepubertal children and in adults with hypothalamic hypogonadotropic hypogonadism. These results provide further support for reactivation of the hypothalamic LHRH pulse generator as the first hormonal change in the onset of puberty.

Responsiveness of the gonads to gonadotropins also increases during puberty. For example, the augmented testosterone secretion in response to administration of hCG at puberty in boys is probably a consequence of the priming effect of the increase in endogenous secretion of LH (in the presence of FSH) on the Leydig cell.

Maturation of Positive Feedback Mechanism

In normal women, the midcycle surge in LH and FSH secretion is attributed to the positive feedback effect of an increased concentration of plasma estradiol for a sufficient length of time during the latter part of the follicular phase. Estradiol has both negative and positive feedback effects on the hypothalamic-pituitary system. Although the suppressive effect is probably operative from late fetal life on, the positive action of endogenous (or exogenous) estradiol on gonadotropin release has not been demonstrated in normal prepubertal and early pubertal children. Hence, acquisition of positive feedback, a requisite for ovulation, is a late maturational event in puberty and, from the present evidence, probably does not occur before midpuberty in normal girls.

Among the requirements for a positive feedback action of estradiol on gonadotropin release at puberty are (1) ovarian follicles primed by FSH to secrete sufficient estradiol to reach and maintain a critical level in the circulation, (2) a pituitary gland that is sensitized to LHRH and contains a large enough pool of releasable LH to support an LH surge, and (3) controversial in the human but not in lower animals, sufficient LHRH stores for the LHRH neurosecretory neurons to respond with an acute increase in LH secretion in addition to the usual adult pattern of pulsatile LH secretion.

The main site of action of estradiol is at the level of the anterior pituitary, but estrogen has dual sites of action including a negative as well as positive feedback action on the hypothalamus. Knobil and colleagues have shown in the rhesus monkey that positive as well as negative feedback can occur in adult ovariectomized females in whom the medial basal hypothalamus is surgically disconnected from the remainder of the CNS. In monkeys with hypothalamic lesions, unvarying, intermittent LHRH administration leads to sufficient estradiol release from the ovary to induce an ovulatory LH surge in the absence of an increase in the dose of the LHRH pulses. Estradiol has a positive feedback effect directly on the pituitary gland in normal women, and prolonged administration of estradiol is accompanied by an increased LH response to LHRH administration in women. The fact that the major positive feedback action on the pituitary gland is demonstrable in the absence of an increase in pulsatile LHRH secretion suggests that the failure to elicit a positive feedback action with administration of estradiol to prepubertal girls could be related to the inadequate LHRH pulses or insufficient LH reserve, respectively, or both components.

That gonadotropin cyclic and estradiol-induced positive feedback can be demonstrated by midpuberty and before menarche does not imply that the positive feedback loop is complete. Indeed, the modulating effect of the pubertal ovary and its output of estradiol on the hypothalamic-pituitary gonadotropin unit may be insufficient to induce an ovulatory LH surge even when there is an adequate pituitary store of readily releasable LH and FSH. The ovary, because of lack of sufficient gonadotropin stimulation, decreased responsivity, or other local factors, does not secrete estradiol at a high level or long enough to induce an ovulatory LH surge. We visualize the process leading to ovulation as a gradual one in which the ovary (the Zeitgeber for ovulation) and the hypothalamic-pituitary gonadotropin complex become progressively more integrated and synchronous until, finally, an ovary primed for ovulation secretes sufficient estradiol to induce an ovulatory LH surge.

Studies of basal body temperature and of plasma progesterone concentrations suggest that as many as 55% to 90% of cycles are anovulatory during the first 2 years after menarche and that the proportion decreases to less than 20% of cycles by 5 years after menarche. A cyclic surge of LH occurs during some anovulatory cycles in adolescence, but the mechanism of ovulation seems unstable and immature and does not appear to have attained the fine-tuning and synchronization requisite for maintenance of regular ovulatory cycles. There is a rise in BMI, waist circumference, hip circumference, serum LH, androstenedione, testosterone, and DHEAS in the few years following menarche.
Summary of Present Concept

Our present concept of the role of the hypothalamic-pituitary-gonadal system in the control of the onset of puberty is illustrated in Table 24-20. Clearly, the understanding of these complex maturational processes is incomplete. Puberty is not an immutable process; it can be arrested or even reversed. Environmental factors and certain disorders that affect the onset or progression of puberty mediate their effects by direct or indirect suppression of the hypothalamic LHRH pulse generator and its periodic oscillatory signal, LHRH. For example, strenuous physical conditioning in girls (but not boys) and anorexia nervosa can delay or arrest puberty or lead to the reversion of the hypothalamic-pituitary unit to a prepubertal state, depending on the magnitude of the functional LHRH insufficiency. With a decrease in physical activity in the former and with resumption of weight gain and attainment of sufficient body mass in the latter, the pubertal process is reactivated. In rare instances, true precocious puberty caused by an extrinsic mass lesion that impinges on the hypothalamus can be reversed by decompression or removal of the mass (a subarachnoid cyst, for example).
ADRENAL ANDROGENS AND ADRENARCHE

The adrenal component of pubertal maturation (the adrenarche) and the interactions between adrenal and gonadal hormones are poorly understood. Speculation has focused on the mechanism of adrenarche, the fact that adrenarche occurs earlier than gonadarche (the maturation of the hypothalamic-pituitary-gonadal system), and the interaction between adrenal and gonadal hormones at puberty.

Nature and Regulation of Adrenal Androgens

The major adrenal androgen precursors secreted by the adrenal cortex are DHEA, DHEAS, and androstenedione; apparently, none of these C_{19}-steroids directly activate the androgen receptor. By extraglandular metabolism, the so-called adrenal androgens contribute to physiologically active testosterone and estradiol. In normal adult women, only androstenedione is an important precursor; DHEA and DHEAS contribute little to circulating testosterone and estradiol but can be converted locally to these steroids in some peripheral tissues. However, scant information is available on the metabolism and kinetics of DHEA and DHEAS in prepubertal children. Androstenedione is the major androgen secreted by the ovary during and after puberty. It is more readily converted to potent androgens than DHEA or DHEAS. However, DHEA and especially DHEAS (which binds avidly to serum proteins, particularly albumin) are useful biochemical markers of adrenal androgen secretion and the onset of adrenarche.

Cross-sectional and longitudinal studies have demonstrated a progressive increase in the plasma concentration of DHEA and DHEAS in boys and girls by about the age of 6 (to 8 years skeletal age) that continues through puberty (age 13 to 15). Reaches a peak at ages 20 to 30, and then gradually decreases. These age-related and developmental (e.g., fetal) changes in DHEA are unique and not matched by other steroid hormones (Fig. 24-42). The 20-fold increase in the concentration of DHEAS between the onset of adrenarche and midpuberty, without a concomitant increase in plasma cortisol, is accompanied by increased excretion of urinary 17-ketosteroids, especially 11-deoxy C_{19}-steroids. This increase serves as a mark of the onset of adrenarche and begins approximately 2 years before the increase in gonadotropin and gonadal steroid secretion. The increase is not associated with increased sensitivity of the pituitary gonadotrophs to LHRH or with sleep-associated LH secretion and occurs at an age when the hypothalamic-pituitary-gonadal complex is functioning at a low level. Associated with the increase in the adrenal secretion of DHEA and DHEAS (and independent of a change in the secretion of cortisol or aldosterone) are the appearance and growth of the zona reticularis coincident with adrenarche (see Fig. 24-42). The zona reticularis, the innermost zone of the adrenal cortex, is the last step in zonation of the adrenal cortex and the sole source of DHEA and DHEAS. In contrast to the zona glomerulosa and zona fasciculata, four main features distinguish the zona reticularis:

1. An exceedingly low level of expression of 3-hydroxysteroid/4,5-isomerase type 2 and cytochrome P450 mRNAs and minimal enzyme activities.
2. Abundant dehydroepiandrosterone (hydroxysteroid) sulfotransferase activity. Unlike cortisol secretion, the secretion of DHEA and DHEAS in response to ACTH administration varies with age. Dissociation of adrenarche and gonadarche occurs in a variety of disorders of sexual maturation.

3. Relative increase in 17,20-lyase to 17-hydroxylase activity of P450c17, the enzyme that catalyzes both activities and cytochrome b5 synthesis. These characteristics are shared by the fetal zona of the fetal adrenal cortex.

4. Expression of major histocompatibility complex class II (human leucocyte antigen DR) antigens (these antigens are not expressed in the fetal zona of the fetal adrenal cortex).

The CYP17 gene encodes the microsomal enzyme P450c17, the single enzyme that catalyzes both adrenal 17-hydroxylase and 17,20-lyase activities. In contrast to the zona fasciculata, the zona reticularis has an increased 17,20-lyase/17-hydroxylase ratio. Tissue-specific regulation of the 17,20-lyase activity between the human adrenal and testis is well described. In the human Leydig cell (and quite likely the zona reticularis) the 5 pregnenolone substrate is 17-hydroxylated and the 17-hydroxyprogrenenolone is oxidized to DHEA, whereas in the zona fasciculata the 17-hydroxyprogrenolone is converted to 17-hydroxyprogesterone and 21-hydroxylated and the oxidation to DHEA is inhibited. In the rat, site-directed mutagenesis of arginine 346 to alanine in P450c17 led to retention of 80% of 17-hydroxylation activity but reduced to 7% the 17,20-lyase activity. Similarly, mutation of the corresponding residue in human P450c17 (arginine 347 to alanine) resulted in strikingly decreased 17,20-lyase activity but retention of 17-hydroxylase activity. Two X/Y phenotypic females with hypergonadotrophic hypergonadism and normal mineralocorticoid and glucocorticoid function had isolated 17,20-lyase deficiency because of homoygous mutations at either the arginine 347 residue or arginine 358 in P450c17.

In contrast to these observations of loss of 17,20-lyase activity with retention of 17-hydroxylase activity, Miller and colleagues showed that the ratio of human 17,20-lyase to 17-hydroxylase activities were increased by increased phosphorylation of serine and threonine residues on the P450c17 enzyme and b5 and by the increased abundance of the redox cytochrome P450 oxidoreductase and by cytochrome b5, which preferentially promotes 17,20-lyase activity by allosterically affecting the interaction between P450c17 and P450 oxidoreductase. These studies provide a provisional hypothesis for the mechanisms that appear to be involved in the relatively increased 17,20-lyase activity of the zona reticularis but not its regulation.

There are several hypotheses about the control of adrenal androgen secretion. Evidence, although incomplete and indirect, suggests that the regulation of androgen secretion in the zona reticularis is based on a dual control mechanism: (1) corticotropin (adrenocorticotropic hormone [ACTH]) is obligatory, as evidenced, for example, in the findings in ACTH deficiency or resistance, for (2) the action of an unidentified adrenal andrognstimulating factor, possibly pituitary in origin or from a nonadrenal source, or an intra-adrenal event. This concept is illustrated in Figure 24-44. Rejected alternatives to a unique adrenal androgen stimulating factor are the known pituitary hormones, such as an isolated increase in ACTH or endorphin, pro-opiomelanocortin (residues 79 to 96), and normal mineralocorticoid and glucocorticoid function had isolated 17,20-lyase deficiency because of homoygous mutations at either the arginine 347 residue or arginine 358 in P450c17.

A distinct adrenal androgen stimulating factor, whether of pituitary, intra-adrenal, or other origin, could explain the following observations.

1. The spurt in adrenal growth and the differentiation and growth of the zona reticularis at adrenarche occur independently of an increase in ACTH or cortisol secretion but correlate with the increase in plasma DHEAS (see Fig. 24-42).

2. Cortisol and adrenal androgen secretions vary independently with age, during normal as well as premature adrenarche, and in Cushing's disease, starvation, malnutrition, anorexia nervosa, and chronic disease.

3. Unlike cortisol secretion, the secretion of DHEA and DHEAS in response to ACTH administration varies with age.

4. Dissociation of adrenarche and gonadarche occurs in a variety of disorders of sexual maturation (see Fig. 24-44), including premature adrenarche (onset of pubic or axillary hair before age 6), chronic adrenal insufficiency, true precocious puberty (when the onset is before age 6), primary hypergonadism, isolated gonadotropin deficiency, and anorexia nervosa.

A longitudinal study of 42 children demonstrated that an increase in BMI (not the value itself at any age) is related to the rise in the urinary excretion of DHEAS, suggesting that a change in nutritional status is one physiologic regulator of adrenarche.
Adrenal Androgens and Puberty

The earlier onset of adrenarche than gonadarche and the contribution of adrenal androgens to the growth of public and axillary hair have led some to suggest that in normal children adrenal androgens are an important factor in the onset of puberty and the maturation of the hypothalamic-pituitary-gonadal complex. Although true precocious puberty may occur in circumstances in which the prepubertal child has previously been exposed to excessive levels of androgens from an endogenous or an exogenous source (e.g., after the initiation of glucocorticoid therapy in congenital virilizing adrenal hyperplasia or after removal of a sex steroid-secreting adrenal or gonadal neoplasm), there is little evidence that adrenal androgens play an important qualitative or rate-limiting role in the onset of puberty in normal children. Most patients with premature adrenarche, who secrete excessive amounts of adrenal androgens for their age, enter puberty and experience menarche within the normal age range. Moreover, prepubertal children who have congenital or acquired chronic adrenal insufficiency (Addison's disease) and, consequently, have deficient or absent adrenal androgen secretion usually have a normal onset of and progression through puberty when given appropriate glucocorticoid and mineralocorticoid replacement therapy.

Thus, early activation of adrenal androgen secretion does not commonly lead to sexual precocity, nor is deficient or absent adrenal androgen output usually associated with delayed puberty. Furthermore, growth studies in children with chronic adrenal insufficiency, isolated gonadotropin deficiency, hypergonadotropic hypogonadism, and androgen resistance suggest that in girls and boys adrenal androgens are not essential for the adolescent growth spurt, whereas gonadal steroids secreted by the testis and ovary are and act in concert with GH. A transient increase in height velocity (about 1.5 cm/year in both sexes) that occurs in middle childhood (6 to 7 years) and lasts about 2 years has been attributed by some to adrenarche. However, the middle childhood spurt, which terminates while serum DHEAS continues to increase, is related to the cyclic pattern of prepubertal growth and to genetic regulation of growth rather than an increase in either adrenal androgen or GH secretion. This hypothesis, which has been controversial, is supported by a longitudinal analysis of urinary excretion of DHEAS and total 17-ketosteroid sulfates with prepubertal growth.
In patients with constitutional delay in growth and puberty, bone age correlates better with the time of onset of and stage of puberty than does chronologic age. These patients have a retarded bone age at presentation but, on achieving a bone age of approximately 12 to 14 years for boys and 11 to 13 years for girls, they can be expected to show the earliest stages of sexual maturation. The U.S. Health Examination Survey showed that 5.7% of boys with a bone age of 14 years lacked pubic hair (stage 1) and 4% were in genital stage 1, whereas at age 15 years only 0.2% were still in pubic hair stage 1 and 0.8% were still in genital stage 1. Unfortunately, the study started at an age at which the same descriptive information could not be determined for girls. 

In summary, healthy individuals who spontaneously enter puberty after the age of 13 for girls and 14 for boys have constitutional delay in growth and adolescence. Affected individuals are usually short (2 SD below the mean value for height for age) at presentation but, on achieving a bone age of approximately 12 to 14 years for boys and 11 to 13 years for girls, they can be expected to show the earliest stages of sexual maturation. The U.S. Health Examination Survey showed that 5.7% of boys with a bone age of 14 years lacked pubic hair (stage 1) and 4% were in genital stage 1, whereas at age 15 years only 0.2% were still in pubic hair stage 1 and 0.8% were still in genital stage 1. Unfortunately, the study started at an age at which the same descriptive information could not be determined for girls.

Healthy individuals who spontaneously enter puberty after the age of 13 for girls and 14 for boys have constitutional delay in growth and adolescence. Affected individuals are usually short (2 SD below the mean value for height for age) at presentation but, on achieving a bone age of approximately 12 to 14 years for boys and 11 to 13 years for girls, they can be expected to show the earliest stages of sexual maturation. The U.S. Health Examination Survey showed that 5.7% of boys with a bone age of 14 years lacked pubic hair (stage 1) and 4% were in genital stage 1, whereas at age 15 years only 0.2% were still in pubic hair stage 1 and 0.8% were still in genital stage 1. Unfortunately, the study started at an age at which the same descriptive information could not be determined for girls.
Other causes
- Langerhans' histiocytosis
- Postinfectious lesions of the CNS
- Vascular abnormalities of the CNS
- Radiation therapy
- Congenital malformations especially associated with craniofacial anomalies
- Head trauma
- Lymphocytic hypophysitis

**Isolated Gonadotropin Deficiency**
- Kallmann's syndrome
  - With hyposmia or anosmia
  - Without anosmia
- LH/FSH receptor mutation
- Congenital adrenal hypoplasia (DAX1 mutation)
- Isolated LH deficiency
- Isolated FSH deficiency
- Prohormone convertase 1 deficiency (PCI)

**Idiopathic and Genetic Forms of Multiple Pituitary Hormone Deficiencies Including PROP-1 Mutation**

**Miscellaneous Disorders**
- Prader-Willi syndrome
- Laurence-Moon and Bardet-Biedl syndromes
- Functional gonadotropin deficiency
  - Chronic systemic disease and malnutrition
    - Sickle cell disease
    - Cystic fibrosis
    - Acquired immunodeficiency syndrome (AIDS)
    - Chronic gastroenteric disease
    - Chronic renal disease
    - Malnutrition
    - Anorexia nervosa
    - Bulimia
    - Psychogenic amenorrhea
  - Impaired puberty and delayed menarche in female athletes and ballet dancers (exercise amenorrhea)
- Hypothyroidism
- Diabetes mellitus
- Cushing's disease
- Hyperprolactinemia
- Marijuana use
- Gaucher's disease

**Hypergonadotropic Hypogonadism**

**Males**
- The syndrome of seminiferous tubular dysgenesis and its variants
  - (Klinefelter's syndrome)
- Other forms of primary testicular failure
  - Chemotherapy
  - Radiation therapy
  - Testicular steroid biosynthetic defects
  - Sertoli-only syndrome
  - LH receptor mutation
  - Anorchia and cryptorchidism
- Trauma/surgery

**Females**
- The syndrome of gonadal dysgenesis (Turner's syndrome) and its variants
  - XX and XY gonadal dysgenesis
  - Familial and sporadic XX gonadal dysgenesis and its variants
  - Familial and sporadic XY gonadal dysgenesis and its variants
  - Aromatase deficiency
- Other forms of primary ovarian failure
  - Premature menopause
  - Radiation therapy
  - Chemotherapy
  - Autoimmune oophoritis
Galactosemia
Glycoprotein syndrome type 1
Resistant ovary
FSH receptor mutation
LH/CG resistance
Polycystic ovarian disease
Trauma/surgery
Noonan’s or pseudo-Turner’s syndrome
Ovarian steroid biosynthetic defects

and these patients are the ones who most often seek medical advice. The combination of delayed pubertal maturation and decreased stature during adolescence, superimposed on a strong familial tendency toward short stature, leads to conspicuous shortness, especially in the peripubertal period, more often than with either condition alone. We emphasize that no one test distinguishes between constitutional delay in growth and puberty and hypogonadotropic hypogonadism.

Growth rate before the actual onset of puberty in these patients is often suboptimal for chronologic age, but growth velocity usually increases to normal levels after puberty begins. A theoretical model of growth in delayed puberty is proposed on the basis of the infancy-childhood-puberty growth chart. A transient decrease in GH secretion is present in some individuals with constitutional delay but is usually normal. GH release as well as GH secretagogues including the administration of GH-releasing hormone may be decreased in children with constitutional delay in puberty. The amplitude of GH secretion and the GH response to GH-releasing hormone increase after the administration of exogenous androgens or estrogens. Thus, constitutional delay in puberty may constitute a state of functional temporary GH insufficiency for chronologic age but not for bone age; this does not constitute a rationale for treatment of constitutional delay with GH. There is an interaction of IGF-I and gonadotropins in the ovary and testis, and the relatively low secretion of GH (and presumably intraglandular IGF-I) in constitutional delayed puberty may impair the gonadal secretion of gonadotropins. Affected boys seem to be more distrest by short stature than by the delay in sexual development. Occasionally, individuals with constitutional delay in puberty are of normal stature. In such instances, the genetic tendency for growth is greater than in cases characterized by short stature. In these patients, diagnostic and therapeutic decisions focus on the pubertal status alone. Patients with constitutional delay in adolescence and growth often do not reach their predicted height. This has been attributed in a retropective study to a lack of growth of the spine in relation to leg length and is said to result in eunuchoid proportions when the subject reaches adulthood (associated with a decreased upper to lower segment ratio); the greater this segmental disproportion is, the closer the patient is to reaching target height. An alternative explanation for reduced adult stature is that the patients most likely to be referred and reported are those who combine genetic short stature with constitutional delay in growth and puberty. The magnitude of the catch-up linear growth during puberty in boys is a major determinant of adult height.

Although the use of exogenous androgens may improve self-image and start the secondary sexual changes of puberty, low-dose androgens neither improve final height nor change the

<table>
<thead>
<tr>
<th>TABLE 24-22</th>
<th>Constitutional Delay in Growth and Adolescence</th>
</tr>
</thead>
<tbody>
<tr>
<td>A variation of normal</td>
<td></td>
</tr>
<tr>
<td>Males more often seek assistance</td>
<td></td>
</tr>
<tr>
<td>Family history of delayed menarche or delayed secondary sexual characteristics</td>
<td></td>
</tr>
<tr>
<td>Height is often below the fifth percentile, but growth rate is normal for skeletal age</td>
<td></td>
</tr>
<tr>
<td>Onset of adrenarche is delayed</td>
<td></td>
</tr>
<tr>
<td>The combination of genetic short stature and constitutional delay leads to more profound short stature</td>
<td></td>
</tr>
<tr>
<td>Final height is less than predicted</td>
<td></td>
</tr>
</tbody>
</table>

eunuchoid proportions. Further, the use of GH therapy has not been shown to increase the final height in constitutional delay even if growth velocity does transiently increase with GH therapy. Girls with constitutional delay in growth and puberty present more rarely than boys, but as a group they also reach suboptimal height. They usually reach their predicted adult height but not their target height, whereas some boys may reach neither, and as a group patients with constitutional delay have a mean deficit of 2.4 cm below the mean predicted height. Individual differences in final height are striking, however, varying about 10 cm from predicted adult height. In counseling, the physician must be aware of the limitations in the accuracy of a prediction of final height in a patient with current methods. Some studies have combined LHRH agonist therapy with GH treatment to increase final height in children who are normal except for genetic short stature, but the preliminary results are either inconclusive or show increased predicted or near final height (which does not necessarily translate into increased adult height). This approach to treatment remains experimental and exceptionally expensive ($14,000 per centimeter gained by one estimate).

The discovery of the critical role of estradiol in skeletal maturation in boys as well as girls led to the suggestion that in boys with constitutional delayed growth and puberty, treatment with a potent aromatase inhibitor would improve adult height by inhibiting skeletal maturation. Preliminary evidence was obtained in a double-blind, randomized, placebo-controlled study involving boys with constitutional delay in puberty in which treatment with a course of monthly intramuscular repository testosterone and placebo was compared with testosterone and once-daily oral letrozole (a potent fourth-generation aromatase inhibitor). In the letrozole plus testosterone group, in which serum estradiol was strikingly suppressed during the 12-month period, the mean increase in predicted adult height was 5.1 cm while the predicted adult height was unchanged in the testosterone placebo group. These early observations suggest that adult height can be increased by inhibiting estradiol synthesis and dampening the rate of skeletal maturation without affecting the development of male secondary sex characteristics. Normal volumetric bone density has been described in young men previously affected by constitutional delay, in contrast to the decreased peak bone mass related to
Insufficient pulsatile secretion of LH RH and the resulting FSH and LH deficiency lead to delayed sexual maturation. The magnitude of the LH RH deficiency and hence the phenotype can vary from severe sexual infantilism to instances in which the separation from constitutional delay of puberty is difficult. LH RH deficiency may be secondary to a genetic or developmental defect present at birth but undetected until the age of expected puberty, or it may be due to a tumor, inflammatory process, vascular lesion, radiation, or trauma.

The deficiency of pulsatile LH RH may be quantitatively absolute or relative or qualitative, especially in females; it may involve abnormalities in the amplitude or frequency of LH RH or L H pulses or both. 

...comparatively, gonadotropin deficiency may arise from lesions or defects that involve the pituitary gland directly. When GH is affected as well as gonadotropins, impaired growth is...
Sexual infantilism may be caused by other extrasellar tumors that arise in or encroach on the hypothalamus. Germinomas (previously termed pinealomas, ectopic pinealomas, atypical teratomas, or dysgerminomas) or other germ cell tumors of the CNS are the extrasellar tumors that most commonly cause sexual infantilism, although, when all primary CNS tumors are considered, germinomas are rare. The diagnosis is usually made during the second decade of life. Polydipsia and polyuria are among the most common symptoms followed by visual difficulties and abnormalities of growth and puberty. The most common endocrine abnormalities are deficiencies of vasopressin and GH, but other anterior pituitary hormone deficiencies (including gonadotropin deficiency) and elevated serum prolactin levels are frequent. The concentrations of hCG, in spinal fluid especially and in serum, and of -fetoprotein are useful tumor markers in children and adolescents with a CNS germ cell tumor. Rather than delaying puberty, germ cell tumors in boys may cause gynecomastia precociously by secretion of hCG (see section on sexual precocity). A similar problem is seen in suprasellar teratomas that produced mild sexual precocity in a 6-year-old girl who had nondetectable serum concentrations of LH and FSH was reported. This pubertal development was thought to be possibly related to aromatase activity of the teratoma; with therapy and regression of the tumor, the breast budding disappeared.

A germ cell tumor may arise in the suprasellar hypothalamic region, in the pineal region, or in another area of the CNS. Subependymal spread along the lining of the third ventricle is common, and seeding may lead to involvement of the lower spinal cord and corda equina. MRI scans with contrast enhancement are useful in the diagnosis of tumors more than 0.5 cm in diameter and detection of isolated enlargement of the pituitary stalk, an early finding on MRI scans. Periodic MRI monitoring for further development of a tumor is indicated whenever thickening of the pituitary stalk is encountered. Hypothalamic-pituitary abnormalities on MRI are related to functional defects such as diabetes insipidus. Unlike the size of the pituitary gland, which increases 100% between years 1 and 15, the size of the pineal gland does not change after age 1 in normal individuals, and thus any enlargement after that time is suspicious of a mass lesion. Pure germ cell tumors are low in radioactivity, and therefore the preferred treatment; the clinical features and the response to radiation therapy are so characteristic that surgery is rarely indicated except for biopsy to establish a tissue diagnosis. When a mixed germ cell tumor is found, both radiation therapy and chemotherapy are recommended.

Hypothalamic and optic gliomas or astrocytomas occurring either as part of neurofibromatosis (von Recklinghausen's disease) or incidentally, can also cause sexual infantilism.

Pituitary Tumors

Pituitary adenomas are rare in childhood and adolescence as only 2% to 6% of all pituitary adenomas occur in this age group. In one study, 50% of pituitary adenomas occurring before adulthood were prolactinomas, 20% were GH-secreting adenomas, and 30% were chromophobe adenomas. Hyperprolactinemia related to microprolactinomas or macroprolactinomas of the pituitary is uncommon in childhood and adolescence and is a rare cause of delayed puberty in both boys and girls. Among our patients, only 2 of 29 had delayed onset of puberty, although primary amenorrhea was the presenting symptom in 13 of 20 prolactinomas. Galactorrhea may be absent by history but is often demonstrable by manual manipulation of the nipples (because serum prolactin may rise after manipulation of the nipples, samples should be obtained before examination or many hours later). Transphenoidal resection of microprolactinomas in children and adolescents was an effective treatment with an 89% cure rate. The dopamine agonist bromocriptine is used by some as a method of decreasing serum prolactin concentrations and decreasing the size of the tumors; we use this approach in children and adolescents in whom resection of the adenoma is incomplete and to reduce the size of large macroprolactinomas before attempted surgical removal.

Pupertal progression in affected boys and girls as well as normal menstrual function in girls usually follows the reduction in serum prolactin levels. In certain prolactinomas in children and adolescents, there was a preponderance of macroadenomas in girls and of macroadenomas in boys, and these larger tumors led to local symptoms related to their size.

Other Central Nervous System Disorders Leading to Delayed Puberty

Langerhan's Cell Histiocytosis (Hand-Schüller-Christian Disease, or Histiocytosis X)

This disorder, now thought to be a clonal proliferative disorder of Langerhan's histiocytes or their precursors, is characterized by the infiltration of lipid-laden histiocytic cells or foam cells in the skin, viscera, and bone. Diabetes insipidus, usually resulting from infiltration of the hypothalamus or the pituitary stalk or both, is the most common endocrine manifestation. However, GH deficiency and delayed puberty may occur. There may be visceral involvement including the lung, liver, and spleen. Other findings include cyst-like areas in flat bones of the skull, the ribs, the pelvis, and the scapula; in the long bones of the arms and legs, and in the dorsolumbar spine. Lesions of the mandible may lead to the radiographic impression of "floating teeth" within rarefied bone and the clinical finding of delayed closure of the fontanelles. Lesions of the mandible may lead to the radiographic impression of "floating teeth" within rarefied bone and the clinical finding of delayed closure of the fontanelles.

Postinfectious Inflammatory Lesions of the Central Nervous System, Vascular Abnormalities, and Head Trauma

These are unusual causes of hypogonadotropic hypogonadism. Rarely, tuberculous or sarcoid granulomas of the CNS are associated with delayed puberty. Radiotherapy to the head may be hypoplastic, presumably because of the lack of hypothalamic-pituitary factors, and in some patients the neurohypophysis may have an ectopic location.

In our series, the syndrome is associated with decreased maternal age. A mutation in the HESX1 gene is a rare cause of septo-optic dysplasia. Other developmental defects of the anterior pituitary gland associated with hypothalamic-hypogonadism and other pituitary hormone deficiencies are caused by autosomal recessive mutations in homeobox genes encoding transcription factors involved in the early aspects of pituitary development. These include, in addition to HESX1, mutations in LHX3 and PROP1.
Isolated Gonadotropin Deficiency

Isolated hypogonadal hypogonadism is characterized by selective deficiency of gonadotropins owing to a defect at the level of the hypothalamus involving the LHRH pulse generator or the gonadotrophs, or both, without an anatomic lesion (Table 24-23 and Table 24-24). As a consequence, signs of puberty fail to occur by age 14 years in boys and age 13 years in girls or the pubertal maturation is incomplete or transient. In boys, micropenis or undescended testes or both signs are evidence of a fetal testosterone deficiency. The heterogeneous disorders that lead to isolated hypogonadotropic hypogonadism are typically associated with a prepubertal concentration of gonadal sex steroid values (testosterone in boys; estradiol in girls) and low or normal gonadotropin levels. In the severe form, the concentration of gonadal sex steroids and gonadotropins is low, pulsatile secretion of LH is absent or virtually so, and the LH response to the administration of LHRH is deficient. The tests, if palpable, are small and the concentration of serum inhibin B and estimate of seminiferous tubule function are low.

Isolated gonadotropin deficiency may occur in families (about 20% to 30% of patients) or sporadically. The pattern of inheritance in affected families is that of an autosomal dominant, autosomal recessive, or X-linked recessive trait. In contrast to patients with CNS tumors, who usually have associated GH deficiency and growth failure, and to patients with constitutional delay in growth and adolescence, who are short for chronologic age, patients with isolated gonadotropin deficiency are usually of appropriate height for their age. Because their concentrations of gonadal steroids are too low for the epiphyses to fuse at the normal age, these patients develop increased arm span for height and decreased upper/lower ratios (eunuchoid body proportions) and, if untreated, usually become tall adults. An autosomal recessive form has been described in the mouse (hypf) in which there is a deletion of a part of the LHRH gene. The mutant RNA is incapable of generating functional LHRH.

Kallmann's Syndrome

This genetically heterogeneous syndrome (Table 24-25) is the most common form of isolated hypogonadotropic hypogonadism with delayed puberty in which anosmia or hyposmia resulting from agenesis or hypoplasia of the olfactory lobes or sulci, or both, is associated with LHRH deficiency. The prevalence in boys is about four times that in girls. Although the extent of the defect in olfaction usually seems to correlate with the degree of LHRH deficiency, even in patients with complete anosmia the LHRH deficiency may be partial (fetal eunuch syndrome). Rarely, affected men who had a severe delay in puberty may recover spontaneously, experience an increase in testicular size, and enter full puberty.

The magnitude of the LHRH deficiency correlates with the size of the testes. A variety of deletions and mutations of the KAL1 gene have been described, including large and small (exon) deletions, point mutations, and mirror movements of the upper extremities (synkinesia) (see Table 24-24). All of these structures and organs are sites of expression of the KAL1 gene in the human fetus. Associated defects inconsistently present are cleft lip, cleft palate, imperfect facial fusion, seizure disorders, short metacarpals, pes cavus, neurosensory hearing loss, mirror movements of the upper extremities (synkinesia), and growth failure, and to patients with constitutional delay in growth and adolescence, who are short for chronologic age, patients with isolated gonadotropin deficiency are typically associated with a prepubertal concentration of gonadal sex steroid values (testosterone in boys; estradiol in girls) and low or normal gonadotropin levels. In the severe form, the concentration of gonadal sex steroids and gonadotropins is low, pulsatile secretion of LH is absent or virtually so, and the LH response to the administration of LHRH is deficient. The tests, if palpable, are small and the concentration of serum inhibin B and estimate of seminiferous tubule function are low.

TABLE 24-23 — Isolated Gonadotropin Deficiency

<table>
<thead>
<tr>
<th>Classification</th>
<th>Age and Range (yr)</th>
<th>Testicular Enlargement</th>
<th>Undescended Testes</th>
<th>Gynecomastia</th>
<th>Ocular Anomalies</th>
<th>Other Anomalies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Euosmic</td>
<td>3 5/12; 20 6/12;</td>
<td>3/10</td>
<td>3/10</td>
<td>2/10</td>
<td>3/10</td>
<td>6/10</td>
</tr>
</tbody>
</table>

TABLE 24-24 — Isolated Gonadotropin Deficiency: Clinical Features in 20 Adolescent Boys

Males more commonly affected

Familial or sporadic

Height normal for age; tall adult height if untreated

Eunuchoid skeletal proportions

Delayed bone age

Small, often cryptorchid testes; Diameter <2.5 cm prepubertal size; phallus may be small

Normal adrenarche

Examine for anosmia or hyposmia (Kallmann's syndrome)

Look for associated malformations (facial, central nervous system, skeletal, renal)
deletions in this region of the X chromosome can lead to an association of Kallmann's syndrome with X-linked ichthyosis caused by steroid sulfatase deficiency, mental retardation, and chondroplasia punctata.

Autosomal dominant inheritance of the phenotype is supported by some early studies, and this pattern of inheritance is further suggested by a report of an affected male who fathered an affected son after treatment with hCG. Apparent autosomal recessive inheritance characterizes other kindreds. Thus, the various forms of Kallmann's syndrome are due to heterogeneous mutations in which the phenotype can vary. For example, a 20-year-old man with the complete picture of Kallmann's syndrome had an identical twin brother (proved by genetic fingerprinting) with anosmia but a normal adult phenotype and normal plasma testosterone and gonadotropin concentrations.

In classical, X-linked Kallmann's syndrome, fetal LHRH neurosecretory neurons fail to migrate from the olfactory placode, where they arise, to the medial basal hypothalamus, where they constitute the LHRH pulse generator (see earlier discussion). The defect may be absolute or relative. The fetal LHRH-containing cells and neurites are arrested in their migration to the brain and end in a tangle around the cribriform plate and in the dural layers adjacent to the meninges beneath the hypothalamus, where they constitute the LHRH pulse generator (see earlier discussion). The defect may be absolute or relative. The fetal LHRH-containing cells and neurites are arrested in their migration to the brain and end in a tangle around the cribriform plate and in the dural layers adjacent to the meninges beneath the hypothalamus, where they constitute the LHRH pulse generator (see earlier discussion).

As noted previously, in some patients this abnormality can be seen in cranial MRI scans. Hence, the MRI: aplasia or hypoplasia of olfactory bulbs and/or sulci

**TABLE 24-25 -- Features of Kallmann Syndrome**

**Clinical**

- LHRH deficiency: absent or arrested puberty
- Anosmia or hypop sia
- In infancy: microphallus; cryptorchidism
- Normal stature and growth in childhood
- Normal adrenarche
- Eunuchoid proportions
- Associated midline defects (e.g., cleft lip, cleft palate, midline cranial anomalies)
- MRI: aplasia or hypoplasia of olfactory bulbs and/or sulci

**Prevalence:** approximately 1 in 7500 males, 1 in 50,000 females; one-tenth prevalence of Klinefelter's syndrome

**Inheritance:** sporadic and familial cases; genetic heterogeneity

- **X linked**
- X-linked recessive (Kallmann et al.
- X chromosome deletion: Xp22.3 (Ballabio et al.

- **Autosomal**
- Dominant (sex limitation) (Santen and Paulsen
- Recessive (White et al.

- **Anatomy:** developmental field defect
- Aplasia or hypoplasia of olfactory bulb and sulcus
- Arrested migration of LHRH neurosecretory neurons from olfactory placode to medial basal hypothalamus
- LHRH, luteinizing hormonereleasing hormone.

Kallmann's syndrome can be considered a developmental malformation caused by a field defect.
Other Forms of Isolated Hypogonadotropic Hypogonadism

Hypogonadotropic hypogonadism may be transmitted by autosomal recessive inheritance with none of the other features of Kallmann’s syndrome. Males with cerebellar ataxia and deficient gonadotropin production are reported in kindreds with X-linked inheritance (possibly a variant form of Kallmann’s syndrome), and hypogonadotropic hypogonadism may be associated with the multiple lentigines and basal cell nevus syndromes.

Luteinizing Hormone/Releasing Hormone Receptor Mutations

A mutation in the human LHRH gene itself has not been reported, in contrast to the mutation described in the hyp/hyp

mouse. However, familial and sporadic patients have been reported with mutations in the gene encoding the GnRH receptor, the G protein-coupled seven transmembrane segments, which lead to various degrees of hypogonadotropic hypogonadism with normosmia.

More than six families as well as sporadic patients have been described in this autosomal recessive disorder, with heterozygous or homozygous mutations in the LHRH receptor. Mutations are dispersed throughout the coding region of the gene. The clinical presentation is heterogeneous even within the same pedigree and especially in patients with compound heterozygous mutations. The patients may present with severe features of isolated hypogonadotropic hypogonadism, sexual infantilism, long-delayed puberty, relatively mild hypogonadism, and infertility. The clinical variants include the fertile euarch variant and reversal in adulthood.

The impairment of signal transmission by the mutant LHRH receptor is highly variable; in one affected woman, pulsatile LHRH treatment induced ovulation and made a successful pregnancy possible.

In all types of congenital gonadotropin deficiency, male patients are likely to manifest micropenis (penile length less than 2.5 cm at birth and in infancy) because of lack of fetal gonadotropin stimulation of fetal testes during the last half of gestation. Rarely, boys with congenital GH deficiency have micropenis even if gonadotropin levels are normal. Infants and children with microgenitalia related to hypothalamic deficiencies may be treated with one or two 3-month courses of testosterone enanthate, 25 mg per month given intramuscularly, to enlarge the size of the penis.

Although concern was raised that early testosterone therapy might not allow the attainment of a normal adult penile size, experience has proved otherwise. Further, the concern that the penis might not respond to androgens later in life if exposed to testosterone in childhood, a pattern noted in the rat, proved incorrect. Thus, it is appropriate to treat male infants and children with micropenis related to gonadotropin or GH deficiency with short courses of androgens to enlarge the penis into the normal childhood range. It is not appropriate to reverse the sex of a male infant because of microphallus owing to such causes as fetal testosterone deficiency.

X-Linked Congenital Adrenal Hypoplasia and Hypogonadotropic Hypogonadism

This uncommon X-linked recessive disorder of adrenocortical organogenesis is due to a deletion or mutation in the DAX1 gene (dosage-sensitive, sex reversal, adrenal hypoplasia congenita, X chromosome gene 1). The DAX1 gene, a member of the nuclear receptor superfamily, encodes an orphan receptor that is a putative transcriptional repressor. The DAX1 locus undergoes X inactivation. It maps to the DSS (dosage-sensitive sex reversal) locus (Xp21); a double dose of DAX1 is associated with a female phenotype or ambiguous genitalia in 46,XY males.

DAX1 protein has a novel domain in the amino terminus that contains two putative unique zinc finger motifs, and the carboxy terminus contains a conserved ligand-binding domain that binds deoxyribonucleic acid (DNA), localizes in the nucleus, and contains a transcriptional silencing domain that antagonizes the steriodogenic factor-1 (SF-1) trans-activation function. DAX1 has an SF-1 response element in the Sprotomer region.

Abnormalities of the DAX1 gene are characterized by severe glucocorticoid, mineralocorticoid, and, at puberty, androgen deficiency. A mature adrenal cortex is lacking and the abnormal structure of the adrenal cortex resembles that of the fetal zona made up of disorganized vacuolated cytoplasmic cells. The major effect of DAX1 deficiency is a severe primary adrenal insufficiency that with hyponatremia, hyperkalemia, acidosis, and hypocholesterolemia (failure to thrive, vomiting, poor feeding, dehydration, circulatory collapse, increased pigmentation) is lethal if untreated early in life. The concentration of plasma ACTH and plasma renin activity are high; plasma cortisol and aldosterone levels are low. Less commonly, the onset of symptomatic adrenal insufficiency is delayed into later childhood, an early sign of which is increased skin pigmentation.

In the infant male, salt wasting are usually the most prominent feature but cortisol deficiency is present and detectable, and the adrenal insufficiency includes deficient secretion of the zona reticularis steroids, DHEA and its sulfocoujugate. The testses are undescended in less than half of the patients; micropenis is rare, but urogenital abnormalities and hearing loss are occasionally present. Most commonly, because of hypogonadotropic hypogonadism, signs of sexual maturation at the age of puberty are lacking, including absence of pubic and axillary hair and testicular enlargement; the concentrations of serum FSH, LH, and testosterone are
There are exceptions to this typical presentation, related in part to the nature of the DAX1 mutation; for example, the adrenal dysfunction may be subtle and delayed puberty, owing to hypogonadotropic hypogonadism, the prominent clinical feature.

Boys who do not present with clinical evidence of adrenal insufficiency in infancy often have a more insidious onset during childhood. In rare instances, the adrenal insufficiency is not detected until adulthood and the hypogonadotropic hypogonadism is partial. In a pedigree in which two affected boys had a hemizygous DAX1 nonsense mutation and neonatal onset of adrenal insufficiency, a maternal aunt who was homozgyous for the mutation had sexual infantilism and primary amenorrhea but even after decades of follow-up maintained normal adrenal function. A maternal grandfather who carried the same mutation was asymptomatic. This pedigree again highlights the limitations and complexities of genotype and phenotype correlations. A survey of 106 individuals with variants of puberty delay or hypogonadotropic hypogonadism indicated that a mutation of DAX1 is rare and most unlikely to be the cause in such cases unless there is a history of adrenal insufficiency.

For many years, the nature of the hypogonadotropic hypogonadism in this condition has been uncertain. Two advances clarified this issue. First, as discussed subsequently, intragenic mutations in the DAX1 gene indicate that the hypogonadotropic hypogonadism is an intrinsic characteristic of the disorder, a manifestation of the single gene mutation and not related to the involvement of a contiguous gene. Second, the DAX1 gene is expressed not only in the adrenal cortex and testes (and weakly in the ovary) but also in the hypothalamus and pituitary. More commonly, there is evidence of both LHRH deficiency and an abnormality in the gonadotrophs, giving a mixed picture of both hypothalamic and intrinsic gonadotroph defects, but in some instances one or the other defect is primary, usually at the pituitary level. The pulsatile secretion of LH is absent or erratic; basal immunoreactive LH and FSH levels may be normal but the gonadotropins seem to lack bioactivity. Of interest, in affected male infants in whom hypotalamic/pituitary gonadotropin-pituitary function was assessed at birth and in infancy (as late as 140 days), the serum concentration of testosterone was normal (i.e., elevated to puberty levels), as was that of FSH and LH and there was the expected pubertal response to LHRH administration. In addition to the glucocorticoid and mineralocorticoid deficiencies, the serum DHEAS levels were low. These observations suggest that, at least in some affected boys, the LHRH pulse generator/pituitary gonadotropin apparatus is intact and functional in infancy and early childhood and that the LHRH-gonadotroph defects are not manifest until later in childhood or the peripubertal period.

A tall 2-year-old boy with a DAX1 mutation resulting in a frameshift and premature stop codon had as his first clinical manifestation penile enlargement and public hair that proved to be due to LHRH-independent sexual precocity. Serum testosterone values were in the early pubertal range; basal and LHRH-stimulated LH and FSH were prepubertal. The sexual precocity was not modified by LHRH agonist treatment; the LH receptor gene had a normal sequence. At age 3 years, he became progressively pigmented and cortisol and aldosterone deficiencies were documented as well as extremely high levels of plasma ACTH. On replacement therapy with cortisol acetate and 9-fluorocortisone, the size of his testes decreased and the concentration of serum testosterone fell into the prepubertal range. The observations suggested that the exceedingly high circulating ACTH levels, possibly acting through the human melanocortin 1 receptor present in human Leydig cells, were the underlying cause of the increased steroidogenesis and testosterone secretion, which was reversed by glucocorticoid treatment. In addition, as DAX1 inhibits the SF-1 trans-activation, a regulator of steroidogenic genes, the loss of DAX1 inhibition of SF-1 transcriptional activity may also have had a role.

Data on seminiferous tubule function are limited. Azoospermia unresponsive to gonadotropin treatment was detected in a few affected men. Delayed puberty is a manifestation in some female carriers of a DAX1 mutation. In addition to a deletion of DAX1, most mutations in the gene are nonsense mutations and frameshifts; missense mutations with a change in a single amino acid are relatively uncommon. In a review of 86 patients reported to have intragenic mutations in DAX1, the most common were frameshift mutations (49%); nonsense mutations were described in 28% and missense mutations in 20%, almost all of which were located in the ligand-binding domain.

Contiguous gene syndromes are not uncommon in association with X-linked congenital adrenal hypoplasia, the gene of which maps to Xp21, distal to the glycerol kinase (DMD) gene. Most mutations in the DMD gene are nonsense mutations and frameshifts; missense mutations with a change in a single amino acid are relatively uncommon. In a review of 86 patients reported to have intragenic mutations in DAX1, the most common were frameshift mutations (49%); nonsense mutations were described in 28% and missense mutations in 20%, almost all of which were located in the ligand-binding domain.

Isolated Luteinizing Hormone Deficiency

Isolated LH deficiency (the fertile eunuch syndrome) is associated with deficient testosterone production (which responds to administration of hCG) in the presence of variable spermatogenesis; in most instances, it is an incomplete form of isolated gonadotropin deficiency. The disorder may be idiopathic or secondary to a hypogonadotrophic pituitary neoplasm.

Isolated Bioactive Luteinizing Hormone

Rarely, isolated bioactive LH is due to a mutation in the gene encoding the LH subunit. The only known patient had a striking discrepancy between immunoreactive and bioactive LH. A 17-year-old male with a history of delayed puberty with increased immunoreactive serum LH levels but no LH bioactivity and normal serum FSH concentrations had a homozygous mutation in exon 3 of the LH subunit gene (glutamine 54 arginine). Biopsy of the testes showed absent Leydig cells and arrested spermatogenesis. Treatment with hCG increased testosterone secretion and spermatogenesis. The serum LH of the hypergonadotrophic mother exhibited only 55% of normal binding to the LH receptor. He had normal male sex differentiation, most likely because of the action of hCG during the second trimester and extending into the third trimester.

The patient had low testosterone concentrations; it is uncertain that he completely lacked LH bioactivity in late fetal life as he did not have micro-/485, a common finding in male infants with congenital LH deficiency.

Isolated Follicle-Stimulating Hormone Deficiency

Mutations in the FSH subunit, either homozgyous or compound heterozygous, have been reported in three females and two males. The three affected women presented with delayed puberty, lack of breast development or poorly developed secondary sex characteristics, and primary amenorrhea but normal adrenarche. Immunoreactive FSH was not detected in the serum; the LH concentration was elevated and serum estradiol was low. There are three female cases of disordered puberty with FSH subunit mutations. Two of three women had a homozygous nonsense mutation in the FSH subunit gene (Val 61 X) and the other was a compound heterozygote (Cys 51 Gly/Val 61 X).

Two affected men have been reported. Both had azoospermia; small, soft testes; and absence of serum FSH. In one man with normal puberty and LH and testosterone values, the missense mutation was a Cys 82 Arg substitution. The other man, with a nonsense mutation (Val 61 X), had delayed puberty, low testosterone and inhibited, and high LH.

Two males with this mutation had a normal (or only slightly delayed) onset and progression of puberty, but because of abnormal Serum development, the testes were small and soft and there was azoospermia. A woman with a mutation in the prohormone convertase 1 (PC1) gene had extreme childhood obesity, hypocortisolism, defects in conversion of proinsulin to insulin, and isolated hypogonadotropic hypogonadism.

Idiopathic Hypogonadotropic Dwarfism

Idiopathic hypopituitarism is usually caused by a deficiency of hypothalamic releasing factors. In the untreated state, patients with deficient LHRH with rare exceptions have delayed puberty. In contrast, patients with isolated GH deficiency ultimately undergo spontaneous pubertal development, without exogenous gonadal steroids, when the bone age reaches the pubertal stage of 11 to 13 years. Those who have associated gonadotropin deficiency do not
undergo spontaneous puberty, even when the bone age advances to the pubertal stage during GH therapy.

Common to many patients with idiopathic hypopituitarism is early onset of growth failure; late onset of diminished growth suggests the presence of a CNS tumor. There is an association between breech delivery, especially in males, perinatal distress, and idiopathic hypopituitarism, and malformations of the pituitary gland are demonstrable by MRI in common with such patients. The familial forms of multiple pituitary hormone deficiencies with either autosomal recessive or X-linked inheritance are less common.

The degree of hormone deficit and the age of onset of pituitary hormone deficiencies may vary within a single kindred with the same genetic defect.

The absence of GH and gonadotropins may allow long-term but slow growth to increase final height; the height at the onset of puberty and the height in relation to bone age determine the final height that is reached. One patient was reported to be taller than expected for the family after the diagnosis of panhypopituitarism was made at 25 years of age and treatment was given. Treatment with GH in prepubertal children with isolated GH deficiency can increase the rate of pubertal development. Alternatively, if GH treatment is instituted in children already in puberty who have a limited height potential, limitation of the amount of growth attained with GH treatment often results. In these instances, the use of LHRH agonists to suppress pubertal development in addition to the use of GH can increase final height.

The judicious use of low-dose testosterone in affected boys of pubertal age with associated gonadotropin deficiency does not seem to impair growth achieved by GH replacement.

Miscellaneous Conditions

Prader-Willi Syndrome

This autosomal dominant syndrome early-onset childhood hyperphagia; pathologic obesity and carbohydrate intolerance; infantile central hypotonia and lethargy; delayed onset and poor fetal activity, a tendency for intrauterine growth retardation; short stature by 15 years of age; small hands and feet; mild to moderate mental retardation; emotional instability including perseveration, obsessions, and compulsions; and characteristic facies with almond-shaped eyes, triangular mouth, and narrow bitemporal diameters associated with delayed puberty and hypogonadotropic hypogonadism caused by hypothalamic dysfunction. In spite of this, there is a tendency to early adrenarche.

Affected boys often have microgenia and cryptorchidism. Weight reduction may lead to menarche in some females. Hence, severe obesity may play a role in the impaired puberty in some patients. Plasma GH responses to provocative stimuli and to sleep are usually low but may be normal. The assessment of GH secretion has been confounded by the obesity; it is often decreased in nonGH-deficient obese individuals.

The role of relative GH deficiency in this disorder is uncertain and controversial (see Eiholzer and colleagues for a contrary view). In the past, GH deficiency has been advocated to support the treatment of Prader-Willi syndrome with recombinant human GH (rGH). This justification is no longer needed. In June 2000 the U.S. Food and Drug Administration (FDA) approved Prader-Willi syndrome as an indication for rGH treatment in affected children without a requirement for assessing GH secretion; genetic testing is required to confirm the clinical diagnosis.

The decision to approve rGH treatment was strongly influenced by long-term randomized control trials in Prader-Willi syndrome. GH treatment decreases body fat and increases fat utilization, lean body mass, linear growth, and energy expenditure and there are possible improvements in physical strength and motor development. The recommended dose is 1.0 to 1.5 mg/m²/day (0.03 to 0.05 mg/kg/day). The optimal dose, maintenance of positive clinical effects, optimal age for initiation of treatment, and frequency of adverse side effects including type 2 diabetes mellitus remain to be established.

A striking discovery is the discovery that patients with Prader-Willi syndrome have 4- to 5-fold higher fasting concentrations of plasma ghrelin (the "hunger hormone") than in equivalently obese controls. Ghrelin, a novel enteric hormone secreted among other tissues by the stomach, increases food intake (a powerful orexigen) and its body weight, and growth hormone secretion. This finding raises the possibility of new pharmacologic and surgical approaches to treatment.

The ghrelin gene is widely expressed in human tissues, the highest level is in the fundus of the stomach, where its novel acylated peptide ghrelin is localized to X/A cells, a distinctive endocrine cell population in the oxyntic mucosa.

This distinct genetic disorder with a frequency of about 1 in 20,000, rarely familial (the recurrence risk depends on the type of genetic defect), is caused by abnormalities involving the long arm of chromosome 15 in the region q11-q13. Approximately 70% of Prader-Willi cases are caused by a paternal deletion of 15q11-q13 (commonly about 3 to 5 megabase pairs in size); 20% to 25% of cases have maternal uniparental disomy (either isodisomy or heterodisomy) in which both chromosomes 15 are derived from the mother, possibly by nondisjunction during maternal meioisis, and represent a striking example of genomic imprinting. In 2% to 5%, an imprinting center defect has been detected. The lack of a functional paternal 15q11-q13 region, caused by any of a variety of genetic mechanisms, can result in the syndrome. One candidate imprinted gene, among several, that maps to this region, SNRPN (small nuclear ribonucleoprotein-associated polypeptide Smn), implicated in splicing pre-rRNA, is expressed in the brain including the hypothalamus and has been advanced as one explanation of the syndrome. Little is known about the fine structure of the brain in this syndrome. A study of the hypothalamic paraventricular nucleus described a decrease in the number of immunoreactive oxytocin-containing cells in both male and female. Its physiology and disorders. In Yen SSC, Jaffe RB (eds). Reproductive Endocrinology, 2nd ed. Philadelphia, WB Saunders, 1996, pp. 313384.)

Laurence-Moon and Bardet-Biedl Syndromes

The Laurence-Moon syndrome has frequently been incorrectly combined with the Bardet-Biedl syndrome although they are now regarded as distinct entities. Both are rare autosomal recessive traits and both combine retinitis pigmentosa and hypogonadism of various etiologies. Many of the Bardet-Biedl patients have developmental delay, as do all of the Laurence-Moon patients. The Laurence-Moon syndrome, however, is associated with spastic paraplegia, whereas the Bardet-Biedl syndrome involves postaxial polydactyly, onset of obesity usually in early infancy, renal dysplasia, and a relatively high prevalence of the Bedouin of the Middle East. Similar findings are present in the Biedou syndrome II with iris coloboma, hypogonadism, obesity, polydactyly, and developmental delay, but it too is a distinct entity. The genetically and phenotypically heterogeneous Bardet-Biedl syndrome is linked to six loci that map to chromosomes 2, 3, 11, 15, 16, and 20; in most cases three mutant genes are required for a phenotype.

Functional Gonadotropin Deficiencies

Severe systemic and chronic disorders and malnutrition are associated with delayed puberty or failure to progress through the stages of puberty. It is necessary to distinguish the effects of malnutrition, which can lead to functional hypogonadotropic hypogonadism, from the primary effects of the disease. For example, a group of malnourished rural children from Kenya had chronicologic delay in pubertal development and excreted less urinary FSH and LH than well-nourished urban children of the same age. When the two groups were matched by pubertal stage rather than chronologic age, there was no different a difference in gonadotropin excretion.

A study of girls who had previously suffered from kwashiorkor demonstrated no delay in breast development or peak height velocity but a delay in pubic hair growth.
Anorexia nervosa, a common cause of gonadotropin deficiency in adolescence, is a functional disorder, apparently increasing in prevalence in girls but rare in boys. Cushing's disease can be associated with delayed onset or arrest of gonadarche, which is usually corrected by trans-sphenoidal removal of an ACTH-secreting tumor. Hypothyroidism may delay the onset of puberty or menarche; treatment with levothyroxine reverses this pattern. There is likely to be a permanent loss of height if pubertal growth spurt that can result in a normal final height. Celiac disease decreases the growth rate in childhood and adolescence, but with appropriate dietary restrictions final adult height appears normal.

Further, boys with cystic fibrosis have an autoimmune to their sperm that appears at the time of puberty and the appearance of spermatogenesis. Boys with cystic fibrosis have antispermat immunoglobulin M (IgM) antibodies when prepubertal and IgA, IgM, and IgG antibodies during puberty; men with congenital absence of the vas deferens, for comparison, had IgM and IgG antisperm antibodies predominantly.

Boys with sickle cell anemia often exhibit impaired Leydig cell function caused by ischemia of the testes, gonadotropin deficiency, or both factors. Chronic gastrointestinal disease such as Crohn's disease is often accompanied by delayed puberty; therapy to restore nutrition, if successful, enables puberty to progress through the stages of puberty. As GH secretion is usually not affected in AIDS, the poor growth appeared more related to the delay in pubertal development.

Chronic renal disease has been associated with delayed pubertal development and decreased pulsatile gonadotropin secretion related to a decrease in the mass of bioactive and immunoactive LH secreted rather than an alteration of the frequency; successful renal transplantation usually restores gonadotropin secretion.

Patients with nephrotic syndrome have poor pubertal growth, poor secondary sexual development, and deficient gonadotropin secretion in a pattern resembling constitutional delay in puberty. Treatment of glucocorticosteroids with alternate-day glucoocorticoid therapy leads to a late, diminished but prolonged pubertal growth spurt that can result in a normal final height. Children with end-stage renal disease receiving renal or peritoneal dialysis are often delayed in reaching sequential pubertal stages and deficient in linear growth; this loss of growth leads to decreased final height despite the improved growth experienced after renal transplantation. Immuneactive gonadotropin concentrations may be elevated, presumably because of impaired renal clearance, but the response to LHRH is blunted in severe renal impairment. TeBG is elevated in chronic renal failure and free testosterone is low. Although improved growth and pubertal development usually ensue after renal transplantation, the glucocorticoid treatment that follows transplantation presents its own problems. Survivors of renal transplantation who have immune suppression and alternate-day steroid treatment often have delayed onset of puberty and decreased pulsatility of GH and gonadotropins at night.

Advances in the treatment of leukemia have improved the prognosis. Children with early onset and long-term remission experience puberty at an appropriate age or with only slight delay, whereas patients with initial symptoms of leukemia in late childhood may have considerable delay of pubertal development.

The type of therapy for malignancy also influences the age of puberty; radiation to the head may cause hypogonadotropic hypogonadism or GH deficiency, or both, and radiation to the abdomen or pelvis and certain types of chemotherapy, especially if administered during puberty, may impair gonadal function and cause primary hypogonadism. Total-body irradiation for bone marrow transplantation may lead to a decrease in growth in the presence of normal GH secretion. Girls with leukemia treated with CNS radiation demonstrated a diminished pubertal growth spurt and diminished final height; this effect appears to be true whether 24 or 18 Gy is used at a later age, but the final height in boys was reduced only with the higher dose. Further, treatment with 24 Gy before the age of 6 years carries a high risk of short stature. Remarkably, chemotherapy combined with radiation therapy limited to the CNS carries a risk of short stature mainly related to decreased growth of the spine. However, patients treated with chemotherapy without radiation had a normal final height following a period of catch-up growth after the chemotherapy. Secondary malignancies, such as papillary thyroid carcinoma and pulmonary fibrosis, are additional risks with total-body radiation. A follow-up study of over 100 patients treated for acute lymphocytic leukemia before puberty indicated an incidence of obesity of over 45%, indicating the need for dietary counseling in such cases. Hypothyroidism may delay the onset of puberty or menarche; treatment with levotiroxine reverses this pattern. There is likely to be a permanent loss of height if diagnosis is delayed; with thryoxine (T_4) replacement, growth may continue for a longer period after menarche than is the norm but the deficit in height is not regained. Poorly controlled diabetes mellitus can lead to poor growth, fatty infiltration of the liver, and sexual infantilism (Mauriac's syndrome). probably related to poor nutritional status; prepubertal children are most vulnerable to poor glycemic control, and pubertal subjects exhibit normal growth unless severe hyperglycemia occurs. The degree of control necessary to avoid these complications cannot be exactly quantified, but adolescents with even moderately poor control frequently manifest some degree of growth impairment and delayed puberty or irregular menses. Serum IGF-I is decreased in children and adolescents with poorly controlled diabetes mellitus regardless of pubertal stage.

Cushing's disease can be associated with delayed onset or arrest of gonadarche, which is usually corrected by trans-sphenoidal removal of an ACTH-secreting pituitary adenoma. The corticotroph adenoma is the most common prepubertal adenoma. Anorexia nervosa, a common cause of gonadotropin deficiency in adolescence, is a functional disorder, apparently increasing in prevalence in girls but rare in boys, characterized by a distorted body image, obsessive fear of obesity, and food avoidance that can cause severe self-induced weight loss (to less than 85% of normal
weight for age and height or BMI < 17.5 kg/m² after cessation of growth), primary or secondary amenorrhea, widespread endocrine disorders, and even death (specific diagnostic details are in the DSM-IV criteria of the American Psychiatric Association). Other common features include onset in middle adolescence, hyperactivity, defective thermoregulation with hypothermia and sensitivity to cold, constipation, bradycardia and hypotension, decreased basal metabolic rate, dry skin, fine downy hypertrichosis, peripheral edema, and parotid enlargement. The clinician should be aware of the subclinical form. The pathogenesis is multifactorial and includes a genetic factor and a well-characterized psychological component. Anorexia nervosa may rarely occur in association with a primary psychiatric disorder.

It is important to rule out organic disease before the diagnosis of anorexia nervosa is assigned; one girl with macroprolactinoma presented with signs consistent with anorexia nervosa. The prevalence of anorexia nervosa is increased in Turner's syndrome. Hypogonadotrophic hypogonadism is documented in many patients with anorexia nervosa and, in at least part, is related to weight loss. However, unidentified factors may contribute to the amenorrhea of anorexia nervosa, especially when the onset of amenorrhea precedes the onset of severe weight loss. It is not uncommon for a patient with anorexia nervosa to be referred months after the onset of the condition; growth failure may be the first sign of anorexia nervosa, and this condition must be considered in the differential diagnosis of growth failure.

In anorexia nervosa, the concentrations of plasma FSH, LH, leptin, and estradiol and the excretion of urinary gonadotropins are characteristically low. In adult women, there may be a reversion to a circadian rhythm of LH secretion and to the sleep-associated increase in episodic LH secretion characteristic of puberty; in severe cases, the amplitude of the pulsatile episodes is diminished and resembles the pattern in prepubertal children. Similarly, the LH response to LRHR correlates with the severity of the weight loss. In patients who weigh less than 75% of the appropriate weight and have strikingly reduced BMI and percent body fat, there is either a blunted or an absent LH response to the administration of synthetic LRHR and undetectable or small LH pulses. Administration of intravenous LRHR at 90- to 120-minute intervals can stimulate the pituitary to produce LH pulses that are indistinguishable from the normal pubertal pattern. This response further supports the important role of functional LRHR deficiency in the amenorrhea of anorexia nervosa. Serum leptin levels are low, remarkably so with severe malnutrition consistent with the strikingly decreased mass of adipose tissue, and increase when weight is regained. Other hormonal changes include an increased mean concentration of plasma GH and plasma cortisol; low levels of plasma IGF-I, DHEAS, and plasma thyroid stimulating hormone (TSH); a decreased rise in serum prolactin after the administration of thyrotropin-releasing hormone (TRH) or insulin-induced hypoglycemia; and a diminished capacity to concentrate urine.

The restoration of normal endocrine and metabolic function after weight gain suggests that many of these changes are secondary to starvation and severe weight loss; nevertheless, the amenorrhea may persist for months after weight gain, suggesting persistent hypothalamic dysfunction. Treatment of this disorder requires skillful management, understanding, patience, and psychiatric consultation. Various approaches have been used to increase the food intake. In view of the associated mortality, parental alimentation may be indicated in resistant patients with severe weight loss, especially in the presence of infection or an electrolyte imbalance.

In functional amenorrhea, LH and estradiol can be normal, with normal plasma estrogen levels. In swimmers, menstrual cycles were frequently irregular and anovulatory rather than absent, and the plasma concentrations of DHEAS and LH were higher than in normal, with normal plasma estrogen levels.

Exercise, Hypo-Ovarianism, and Amenorrhea (the Female Athlete)

In the late 1970s, reports of amenorrhea in female long-distance runners and delayed menarche in other female athletes, including ballet dancers, figure skaters, and gymnasts, appeared. In 1992 the American College of Sports Medicine defined the female athletic triad of primary or secondary amenorrhea, disordered eating, and osteoporosis or osteopenia derived from the chronic lack of estrogen. Although there are substantial endocrine effects of excessive athletic training in girls, elite prepubertal and pubertal athletes suffer relatively few physical injuries.

Increasing information strengthened the link between increased physical activity and abnormalities of puberty. An incidence of 15% of bulimia, anorexia nervosa, or anorexia athletica was found in 603 Norwegian girls; the athletes most affected were those engaged in sports emphasizing weight. In healthy ballet dancers and female athletes, factors other than

Figure 24-53 47.XXY Klinefelter's syndrome in 17-year-old identical twins. At age 15 gynecomastia was noted. The twins had a eunuchoid habitus and poorly developed male secondary sexual characteristics. Both were 187 cm in height; arm spans were 187 cm and 189.5 cm; the voices were high-pitched; the testes measured 1.8 × 1.5 cm; penis length was 7.5 cm.
intensively trained adolescent girls and their mothers showed a positive correlation between the delayed menarche found in the girls and the age of menarche of their mothers. Likewise, another study of 96 girls demonstrated a relationship between the age of onset of puberty and menarche in athlete's mothers; there is a relationship between the choice of sport and constitutional factors, according to these authors, with no indication that the sport causes changes in growth rate or height. These studies emphasize the role of genetic factors in the age of menarche, pubertal development, and growth rather than the role of energy expenditure. Thus, the biology of pubertal delay in female athletes remains controversial and uncertain.

Although men are less affected than women, males may also be affected by rigorous physical training. Males may have decreased LH response to LHRH and decreased spontaneous LH pulse frequency and amplitude; the serum testosterone is normal or low.

Ballet dancers have a higher incidence of scoliosis than the general population and, as already noted, often have delayed puberty; idiopathic scoliosis in the general population has an association with a statistically earlier age of menarche (0.4 years earlier) and an early adolescent growth spurt. The strongest association with scoliosis is taller stature at the time of the pubertal growth spurt, which, in this study, occurred at an earlier than average age; this combination leads to only a slight increase in final height. Scoliosis usually develops during the pubertal growth spurt and occurs more often in girls with a more rapid pubertal growth spurt. Final height in familial constellations of scoliosis does not vary from the family norm.

Prolactin levels may be elevated in women athletes and could contribute to the delayed menarche found in this group.

Other Causes of Delayed Puberty

Marijuana use has been associated with gynecomastia and is a putative cause of pubertal delay.

Gaucher's disease caused delay in pubertal development in two of three patients in one study.

Girls with familial dysautonomia have delayed menarche and often a severe premenstrual syndrome. The condition is ultimately compatible with pregnancy.

Delayed Puberty and Mood

A prospective study of the intramuscular administration of long-acting testosterone preparations to boys with delayed puberty, aged 14 to 17 years, revealed an effect on mood. The boys were randomly assigned to a course of 200 mg of testosterone enanthate, administered intramuscularly four times at 3-week intervals, or to no treatment. At 1-year follow-up, all of the boys in the testosterone group exhibited excellent growth in stature; growth in the control subjects was significantly lower than that of the testosterone group. Both groups showed improved self-image, and the treated subjects also exhibited notable increases in both school-related and extracurricular social activity. Thus, a relatively brief course of testosterone enanthate had beneficial effects on growth and on inducing or advancing pubertal maturation that was related to improved social function over that achieved simply by the passing of 1 year.

In contrast, however, a study of psychological tests in boys and girls with delayed puberty of mixed causes, including constitutional delay, treated with periods of three doses of steroid alternating with no treatment, demonstrated only one significant treatment effect, namely an increase in withdrawn behavior problems during administration of low-dose estrogen in girls. There were no consistent sex differences. The authors concluded that the administered testosterone or estrogen had minimal effects on behavior problems or mood in adolescents (see earlier for the effects of this treatment schema on sexuality).

Hyponadotropic Hypogonadism: Sexual Infertility Caused by Primary Gonadal Disorders

Primary gonadal failure and the impaired secretion of gonadal steroids lead to decreased negative feedback and elevated LH and FSH levels. The most common forms of primary gonadal failure are associated with sex chromosome abnormalities and characteristic physical findings. Testicular or ovarian dysfunction as an isolated finding is less commonly a cause of pubertal delay.

Klinefelter's Syndrome (Syndrome of Seminiferous Tubular Dysgenesis) and Its Variants

Klinefelter's syndrome, or seminiferous tubular dysgenesis, and its variants occur in approximately 1 in 1000 males and are the most common forms of male hypogonadism. The invariable clinical features include small, firm testes (less than 3.5 cm in length), impaired spermatogenesis, and a male phenotype, usually with gynecomastia and eunuchoid proportions. Elevated gonadotropin levels are found postpubertally; before the age of 12, gonadotropin concentrations are in the prepubertal range. Rarely, low gonadotropin concentrations occur when hyponadotropin hypogonadism is associated with 47,XXY Klinefelter's syndrome.

Hyalinization and fibrosis of the seminiferous tubules and pseudocarcinomatous changes of the Leydig cells develop after puberty; prepubertal testes show only subtle histologic changes, although the testes are small and the germ cell content is reduced. Prepubertally, the disproportionate length of the lower extremities and decreased upper/lower body ratio can identify patients without an increase in arm span. There is variation in Leydig cell function, but the plasma concentration of testosterone tends to be in the normal range until about age 14, after which age it may fail to rise to normal adult levels. The onset of puberty is usually not delayed, but impaired Leydig cell reserve and postpubertal testes may lead to slow progression or arrest of pubertal changes.

Testosterone replacement should be considered when the LH level rises above the normal range of values. Serum estradiol/testosterone ratios and TeBG levels are decreased, but impaired Leydig cell reserve and postpubertal testes may lead to slow progression or arrest of pubertal changes.

Testosterone administration does not appear to reduce the gynecomastia, but dihydrotestosterone and aromatase inhibitors or estrogen receptor antagonists may be effective. If the gynecomastia does not regress, a reduction mammaplasty is required. Tall stature for family size is common in this disorder because of the disproportionate growth of the legs.

Neurobehavioral abnormalities, primarily in language and frontal executive functions, are frequent, and some say universal, in Klinefelter's syndrome. These problems may be severe enough to lead to evaluation in childhood and the postpubertal recognition of the syndrome.

The global I.Q. in unselected populations of subjects with Klinefelter's syndrome is normal or near normal, but verbal I.Q. (VIQ), in contrast to that in Turner's syndrome, is usually lower than performance I.Q. (PIQ). As patients with clinical or psychological problems are referred more often for evaluation, some studies are skewed to suggest more significant deficits than are prevalent in an unselected population of XXY individuals.

Whereas younger patients with Klinefelter's syndrome have a VIQ that is less than their PIQ, there are older adults who have PIQ less than VIQ. It was suggested that PIQ drops in late puberty while VIQ remains stable. Further, prepubertal patients with Klinefelter's syndrome have a VIQ that is less than their PIQ.
relatively high estrogen concentrations characteristic of some patients with Klinefelter's syndrome during puberty, which is reflected as well in the development of gynecomastia. Estrogen is known to enhance verbal skills, and reduced androgen levels decrease visual-spatial function. However, the local aromatization of testosterone to estrogen that is not fully reflected in serum values of these hormones and the varying effect of testosterone and estrogen in Klinefelter's syndrome on physical features such as gynecomastia and habitus make detailed interpretation of these cognitive changes difficult.

Hypotheses have been advanced supporting the effect of prenatal testosterone on cerebral dominance and language and reading pathology, but such explanations are unlikely to explain the difficulties faced by patients with Klinefelter's syndrome as androgen deficiency is not apparent until puberty begins. Anecdotal clinical observation suggests that improvement in psychosocial and self-image problems occurs with testosterone administration, but convincing studies documenting these observations are not yet at hand. Although there is a growing feeling among parents that testosterone treatment in the early pubertal period improves language, reading, and behavior in boys with Klinefelter's syndrome, well-controlled studies supporting this contention are not available. One study did report better mood, less irritability, more energy and drive, less tiredness, more endurance and strength, less need for sleep, better concentration ability, and better relations with others during testosterone treatment of mid-20-year-old adults with Klinefelter's syndrome.

Affected individuals detected by karyotype analysis at birth and in screening studies had minimal impairment (10 to 20 points) in VIQ compared with control subjects and normal full-scale I.Q. Severe retardation is uncommon, although there is an increased prevalence of speech and learning disorders and adjustment problems in adolescence. Psychopathology is rare in most studies, and a 20-year follow-up of 47,XXY individuals showed little or no variation from unaffected controls in employment, social status, mental or physical health, or criminality.

Conditions associated with Klinefelter's syndrome include acrocyanal valvular disease and ruptured berry aneurysms (six times the normal rate), breast carcinoma (20 times the rate in normal men and one fifth that of women); other malignancies such as acute leukemia, lymphoma, and germ cell tumors at any midline site systemic lupus erythematous; and osteoporosis. There is an increased risk of diabetes mellitus and thyroid disease. About 25% of men with Klinefelter's syndrome have osteoporosis.

About 20% of mediastinal germ cells are associated with Klinefelter's syndrome, and they occur at a younger age than the mediastinal germ cell tumors that are not associated with Klinefelter's syndrome (average age 16 versus 27 years in one study). With rare exceptions, these germ cell tumors, which may be located in the midline anywhere from the CNS to the pelvis, secrete hCG and induce sexual precocity. Klinefelter's syndrome needs to be considered in boys with hCG-secreting germ cell tumors, especially if the tumor is located in the mediastinum or CNS.

Survival of childhood cancer is increasing. Most cancer therapy affects testicular function and can lead to adult infertility. Chemotherapeutic agents used in the treatment of nephrotic syndrome or leukemia, such as cyclophosphamide or chlorambucil, have led to Sertoli cell, Leydig cell, and germ cell damage in prepubertal patients; these effects are sometimes reversible. Chemotherapy for childhood Hodgkin's disease, including chlorambucil, vinblastine, mechloethamine (Mustargen), vincristine (Oncovin), procarbazine, and prednisone, may allow spontaneous progression through puberty, but both FSH and LH concentrations may be elevated and the inhibin B concentrations decreased during puberty, indicative of gonadal damage.

The basal serum FSH and the rise in LH and FSH after LHRH are correlated with the dose of cyclophosphamide. Therapy for Hodgkin's disease with COPP and Mustargen, Oncovin, procarbazine, and prednisone (MOPP) can cause severe damage to germinal cells apparently without much effect on Leydig cells even if therapy occurred in the prepubertal period.

Serum FSH is often elevated and inhibit B levels are low in such patients treated with chemotherapy and basal LH is normal although LHRH-stimulated serum LH is elevated; thus, germinal cell damage is evident but Leydig cell function appears normal. In addition, doxorubicin (Adriamycin), bleomycin, vinblastine, and dacarbazine can cause germ cell death. Although some degree of gonadal maturation such as that noted during puberty was considered to be necessary before these drugs could cause gonadal damage, gonadal damage can occur earlier as a result of therapy in the prepubertal period but may not be demonstrable until the age of puberty.

Radiation of the gonads can cause primary testicular failure, usually resulting in azoospermia, although normal testosterone secretion may be associated with elevated LH and FSH values (compensated Leydig cell failure). Because they might be included in a radiation therapy field, the gonads must be shielded from the treatment, if possible. Doses of 0.35 Gy to the testes may lead to temporary aspermia, doses over 2 Gy lead to permanent aspermia, and doses over 15 Gy may cause Leydig cell dysfunction.

Sperm preservation is possible in a boy who will undergo chemotherapy or radiotherapy, although the patient's parents may express concern about the collection of sperm in teenagers by masturbation or electroejaculation. Cryopreservation of sperm in a sperm bank, as carried out for adult patients undergoing orchietomy, is an option. For a prepubertal or early pubertal boy, standard sperm banking techniques may not be appropriate. However, testicular tissue can be preserved freezable and reimplanted later in the subject's own testis or be stimulated to mature for use in intracytoplasmic sperm injection. Other methods of preserving fertility years after treatment for childhood cancers are under consideration. Advances in germ cell preservation are promising, and unexpected developments will undoubtedly emerge over the next two decades.

Male pseudohypoparathyroidism caused by 17-hydroxylase/17,20-lyase (P450c17) deficiency related to mutations in CYP17 is associated with sexual infantilism and a female phenotype; the testosterone biosynthetic defect blocks the synthesis of testosterone and adrenal androgens, impairing masculinization at all stages of development. Associated cortical deficiency and increased mineralocorticoid secretion in this condition lead to hypertension, decreased serum potassium levels, and metabolic alkalosis. Glucocorticoid replacement suppresses ACTH and mineralocorticoid excess and corrects the electrolyte abnormalities, but no sexual development occurs unless exogenous gonadal steroids are administered. Less severe deficiencies are associated with ambiguous genitalia; one case of delayed puberty in a phenotypic male was attributed to partial deficiency of 17,20-desmolase activity. CYP17 mutations leading to isolated 17,20-lyase deficiency are quite rare.

A rare autosomal recessive condition is steroidogenic acute regulatory protein (StAR) deficiency in which the ability to produce C17,20-lyase deficiency are quite rare. Delayed puberty in a phenotypic male was attributed to partial deficiency of 17,20-desmolase activity.

The large adrenal glands may be visualized on sonographic, CT, or MRI scans. Death often occurs in infancy because of unrecognized glucocorticoid and mineralocorticoid deficiency. Affected individuals appear physically to be sexually infantile females, whether their karyotype is 46,XY or 46,XX; because of the absence of gonadal or adrenal androgen production, the affected XY patients develop secondary sexual characteristics including pubic hair. However, surprisingly, XX females even with a null mutation develop female sex characteristics at puberty, including pubic hair and multicystic ovaries, but have either primary or secondary amenorrhea. Apparently, in contrast to the fetal testis, the fetal ovary, which is insensitive to FSH and steroidogenically inactive, is undamaged in fetal life; it remains so until the onset of puberty, when, under FSH stimulation and the recruitment of ovarian follicles, the ovaries undergo progressive

Other Forms of Primary Testicular Failure

Psuedohermaphroditism is a condition in which an individual, who is chromosomically male (46,XY), appears to be female at birth. This condition can occur due to a variety of factors, including genetic, hormonal, and environmental factors. The condition is often associated with internal male reproductive organs and external female genitalia. The cause of pseudohermaphroditism is not always clear, and the condition can be difficult to diagnose.

In most cases, the condition is not progressive and the individual may lead a normal life. However, in some cases, the condition can be associated with fertility issues, and in rare cases, it can be associated with an increased risk of certain types of cancer.

Psychological and Social Aspects

The psychological and social aspects of pseudohermaphroditism can be complex and can vary from individual to individual. The condition can be associated with a range of emotional and psychological issues, including low self-esteem, loneliness, and feelings of isolation.

In some cases, the condition can lead to social and emotional problems, such as social isolation, bullying, and discrimination. In some cases, the condition can also lead to a lack of understanding and acceptance from others, which can further exacerbate the individual's psychological and social issues.

However, with appropriate support and guidance, individuals with pseudohermaphroditism can lead fulfilling and productive lives. It is important to provide them with the necessary support and resources to help them navigate the challenges associated with the condition.

Conclusion

Pseudohermaphroditism is a rare condition that affects individuals with a 46,XY karyotype. Although the cause of the condition is not always clear, it can be associated with a range of physical, psychological, and social issues. With appropriate support and guidance, individuals with pseudohermaphroditism can lead fulfilling and productive lives.
damage and cyst formation.

Luteinizing Hormone Resistance

Presumptive evidence of LH resistance caused by an LH receptor
abnormality on the Leydig cell was reported in an 18-year-old boy with a male phenotype, no male secondary sexual development, gynecomastia, elevated plasma LH levels, and early pubertal plasma testosterone concentrations that did not increase after hCG administration; there was no elevation of testosterone precursor levels.\textsuperscript{1} The testosterone were prepubertal in size and had the microscopic appearance of normal prepubertal testes. Plasma membrane receptor preparations from the testes bound only half as much radiolabeled hCG as control testes.

This autosomal recessive disorder is due to a mutation in the gene encoding the G protein-coupled, seven-transmembrane LH/hCG receptor cell receptor.\textsuperscript{2} in affected males (see Chapter 22): mutations causing a severe compromise in LH/hCG receptor function are associated with XY male pseudohermaphroditism. Another affected phenotype male had a homozygous deletion of exon 10 of the LH receptor.\textsuperscript{3} This deletion causes incomplete loss of function of the LH receptor. Homozygous missense mutations, Ser 616 Tyr, and Ile 625 Lys, are associated with micropenis (but not hypospadias) related to partial impairment of LH receptor function. The Ser 616 Tyr mutant receptor shows an interesting discordance: a poor response to LH but not to hCG.

Nephropathic cystinosis in boys leads to hypergonadotropic hypogonadism.\textsuperscript{4}

Anorchia and Cryptorchidism\textsuperscript{5,6}

In the 46,XY male without palpable testes, it is important to determine whether any testicular tissue is present. The patient may have intra-abdominal testes, which carry an increased risk of malignant degeneration, anorchia (the "vanishing testes" syndrome), in which no testes are found at laparotomy; or retractile testes, a variation of normal.\textsuperscript{7} The presence of a male phenotype and male internal ducts indicates that functioning fetal testes capable of secreting testosterone and AMH were present early during fetal life but degenerated thereafter.

Administration of 3000 U per m\textsuperscript{2} hCG intramuscularly usually evokes an increased concentration of plasma testosterone after 72 hours when functional Leydig cells are present;\textsuperscript{8} lack of a rise in testosterone concentration, in conjunction with an increased plasma concentration of FSH and LH or an augmented gonadotropin response to LRHR,\textsuperscript{9} is evidence for the diagnosis of bilateral anorchia. Alternatively, measurement of AMH indicates the presence of testicular tissue in a range of suspected conditions in prepubertal boys, from anorchia to male pseudohermaphroditism and true hermaphroditism.\textsuperscript{10} Serum inhibin B is a useful indicator of the presence of functional testicular tissue; values were correlated with the testosterone response to hCG administration, values less than 15 pg/mL indicating anorchia.\textsuperscript{11}

Unilateral cryptorchidism versus the presence of a descended testes on one side and none on the other side presents a diagnostic dilemma. In most, but not all (90%)\textsuperscript{12} cases there is testicular compensatory hypertrophy of the descended testes if there is absence of the contralateral testes,\textsuperscript{13} probably because of elevation of FSH secretion. As the finding does not universally predict monorchia, laparoscopy is recommended in this condition.

Cryptorchid testes may descend into the scrotum during more prolonged treatment with hCG (3000 U/m\textsuperscript{2} surface area intramuscularly every other day for six doses),\textsuperscript{14} incremental LRHR,\textsuperscript{15} or a combination of hCG and LRHR treatment.\textsuperscript{16} Although such descent occurs in retractile testes, it can occur in true cryptorchid testes in which descent is not prevented by local anatomical factors. Normally, testicular descent has occurred by 1 year of age; although later descent is described, the incidence is low.\textsuperscript{17,18} Orchidopexy is recommended between 12 and 18 months of age in testes not expected to descend spontaneously.

Two critical steps in the maturation of germ cells are described in the prepubertal testis: (1) at 2 to 3 months of age, the gonocytes (primitive spermatogonia), the fetal stem cell pool, transform into the adult dark spermatogonia, the adult stem cell pool (possibly related to the surge in LH, FSH, and testosterone in early infancy); (2) at 4 to 5 years of age, the onset of meiosis and the appearance of primary spermatocytes occur.\textsuperscript{19} The identification of the gonocyte transformation has influenced recommendations concerning the timing of orchidopexy. Postpubertal orchidopexy is associated with a high (>85%) prevalence of azoospermia or oligospermia.\textsuperscript{20,21} In a study in Copenhagen, the risk of neoplasia was 5% in patients with an intra-abdominal testis, abnormalities of the external genitalia, or an abnormal sex chromosome karyotype compared with 10% (1185) in patients with cryptorchidism who lacked these characteristics.\textsuperscript{22} It has been surmised that cryptorchid testes, even if replaced in the scrotum, may never have normal spermatogenic function as a consequence of an early abnormality in germ cell maturation. vascular damage to the testicular circulation during orchidopexy, or an intrinsic testicular defect.\textsuperscript{23} However, the phenomenon may have been based on a sample of boys who underwent orchidopexy later than is currently believed to be optimal. In a follow-up study of men with cryptorchidism who had an orchidopexy between 1955 and 1975, the paternity rate was 65% for men who had had bilateral cryptorchidism compared with 90% for the formerly unilateral cryptorchid and 93% for control men. The reduction in fertility was supported by semen and hormone analyses.\textsuperscript{24}

Successful fertilization by the use of intracytoplasmic injection of sperm extracted from the testes of cryptorchid men who had orchidopexy after puberty was reported.\textsuperscript{25,26}

Early orchidopexy seems to reduce the risk of carcinoma of the testes,\textsuperscript{27} although dysgenetic testes, even if located in the scrotum, carry an increased risk of malignant transformation.\textsuperscript{28} Undescended testes remain at a higher temperature than descended testes and have a maturation arrest at the conversion of the gonocyte to the spermatogonia, which appears to direct the testes toward malignant degeneration.\textsuperscript{29} Estimates place the incidence of testicular carcinoma at 0.5 per 100,000 boys with an increase of 6% per year in adolescents.\textsuperscript{30} Skakkebaek and co-workers\textsuperscript{31} expressed concern that adverse environmental factors may be important in the apparent increase in testis cancer, cryptorchidism, hypospadias, and low semen quality. A study of 794 men with testicular cancer in England reported that the increasing performance of orchidopexy before 10 years of age appears to have reduced the increased risk of testicular carcinoma associated with undescended testes.\textsuperscript{32} There is a small risk of carcinoma of the testes in puberty, but the absence of carcinoma in situ in puberty is not an assurance that carcinoma will not develop in adult life. Periodic sonography of the testis is recommended after the onset of puberty.\textsuperscript{33} At present, the earlier the orchidopexy is carried out, the better for ultimate function and reduced risk of malignant degeneration.\textsuperscript{34} One year is a useful age at which to consider orchidopexy for undescended testes because it is an age at which the likelihood of spontaneous descent lessens but the benefits to the testes of orchidopexy remain.

The risk of breast cancer is increased in men with a history of undescended testes, orchidopexy, orchitis, testicular injury, infertility, or any cause of delayed puberty. This risk is associated with the gynecomastia that occurs in these conditions.\textsuperscript{35}

Syndrome of Gonadal Dysgenesis and Its Variants (Turner's Syndrome)

See also Chapter 22.\textsuperscript{36} The most common form of hypergonadotropic hypogonadism in the female is the syndrome of gonadal dysgenesis or Turner's syndrome and its variants, a sporadic disorder with an incidence of 1 per 2500 liveborn girls.\textsuperscript{37} In which all (X chromosome monosomy with haploinsufficiency) or part of the second sex chromosome (partial sex chromosome monosomy) is absent. About 99% of 45,X concepts abort spontaneously and 1 in 15 spontaneous abortions has a 45,X karyotype.\textsuperscript{38} The 45,X karyotype is associated with female phenotype, short stature, sexual infantilism, various somatic abnormalities, and frequent fetal demise.

Sex chromosome mosaicism or structural abnormalities of an X or Y chromosome may modify the features of this syndrome, although about 40% of individuals with the five features noted previously have mosaicism or structural abnormalities of the X chromosome. Thus, we may view the syndrome of gonadal dysgenesis and its variants as a continuum ranging from the typical 45,X phenotype to a normal male or female phenotype.\textsuperscript{39} Comprehensive recommendations for the diagnosis and management of Turner's syndrome have been presented by an international committee.\textsuperscript{40}
This karyotype is found in approximately 60% of cases of Turner's syndrome. \[22\] The short stature is due to loss of a homeobox-containing gene located on the pseudoautosomal region (PAR 1) of the short arms of the X (Xp22) and Yp11.3 chromosomes. \[23\] The gene is called \textit{SHOX} (short stature homeobox-containing gene) \[24\] or \textit{PHOG} (pseudoautosomal homeobox osteogenic gene). \[25\] Because it is located on the pseudoautosomal region of the short arm of the X and Y chromosomes, it escapes X inactivation.

\textit{SHOX} haploinsufficiency is responsible for, in addition to abnormal growth, mesomelic growth retardation and Madelung's deformity of the wrist (bilateral bowing of the radius with a dorsal subluxation of the distal ulna) \[26\] in Leri-Weill dyschondrosteosis (\textit{SHOX} haploinsufficiency). Langer mesomelic dysplasia, which includes severe dwarfism with striking hypoplasia or aplasia of the ulnar and fibula, is due to \textit{SHOX} nullizygosity. \textit{SHOX} haploinsufficiency appears to be responsible for -2.0 SD of the approximately -3.0 SD deficit in stature and the skeletal abnormalities in Turner's syndrome. \[27\] A patient with complete gonadal dysgenesis and tall stature had a 45,X/46,der(X) and three doses of the \textit{SHOX} gene because of the \textit{SHOX} duplication on the der(X) chromosome. \[28\]

Turner's syndrome may be seen in the newborn period. In 45,X abortuses there have edema and large hygromas of the neck that may be seen with prenatatal ultrasound studies; this lymphectic defect is the basis for the loose skinfolds that ultimately form the webbed neck (pterygium colli). Affected newborn infants may also have lymphedema of the extremities; the term Bonnevie-Ullrich syndrome has been applied to newborn infants with these features of Turner's syndrome. It is important to determine whether coarctation of the aorta or a bicuspid aortic valve or both are present because of the risk of hypertension and aortic rupture (see Chapter 22).

Features noted during childhood are spread to various locations in the body. Frequent features are distinct faces with micrognathia, "fishmouth" appearance, high-arched palate with dental abnormalities, epicanthal folds, ptosis, low-set or deformed ears, short neck with low hairline and webbing (pterygium colli), and recurrent otitis media often leading to hearing impairment; about 25% of affected adults require hearing aids. \[29\] A broad, shield-like chest leads to the appearance of widespread nipples; the areolae are often hypoplastic. Skeletal defects include short fourth metacarpals and cubitus valgus (which may develop after birth), Madelung's deformity of the wrist (in about 7%), genu valgum, and scoliosis. The skin demonstrates extensive pigmented nevi, a tendency to keloid formation, and hypoplastic nails. \[30\] Lymphatic obstruction leads not only to the infantile puffiness of extremities and pterygium colli but also to a distinctive shape of the ears. Cardiovascular anomalies include coarctation of the aorta in about 10% (40% have associated webbing of the neck), aortic stenosis, and bicuspid aortic valves; the latter individuals are at risk for a dissecting aortic aneurysm. \[31\]

An echocardiogram of the cardiovascular system must be obtained. Prophylactic antibiotics are indicated if an anatomic abnormality is demonstrated. Abnormal phevocalceal collecting systems, abnormal position or alignment of the kidneys, and abnormalities of the gonadals are found in Turner's syndrome. The X chromatin pattern was negative, and the karyotype was 45,X. She was short (height 134.5 cm; height age 9.5 years) and sexually infantile except for the appearance of sparse pubic hair, and exhibited characteristic stigmata of the syndrome: a short webbed neck, shield-like chest with widely separated nipples, bilateral metacarpal signs, ptosis over the dorsum of the fingers, cubitus valgus, increased number of pigmented nevi, characteristic facies, and low-set ears. The bone age was 13 /2 years; urinary 17-ketosteroids 5.1 mg/day; urinary gonadotropin greater than 100 mIU/day. Vaginal smears and the urocytogram showed an immature pattern in which only squamous cells were absent. With estrogen therapy, female secondary sexual characteristics were induced; the cyclic administration resulted in periodic estrogen withdrawal bleeding. Right, A 45,X, 9/11/2-year-old patient with Turner's syndrome. Apart from short stature (height 118 cm; age 8 10/2 years), increased pigmented nevi, and subtle changes in the fingers and toes, she had few somatic anomalies. In contrast to the patient at the left, the main clinical feature was short stature.

Abnormal vascular supply to the kidney are encountered in 30% to 60% of patients, and recurrent urirary tract infections are uncommon. \[32\] Defects of the gastrointestinal system include intestinal telangiectasias and hemangiomatose that rarely lead to massive gastrointestinal bleeding. Furthermore, the prevalence of inflammatory bowel disease, chronic liver disease, and colon cancer is increased. \[33\] The uterus and fallopian tubes are infantile. Pelvic ultrasonography or MRI usually permits the detection of even a small uterus in these patients and commonly streak gonads. Autoimmune diseases, such as Hashimoto's thyroiditis (a 16-fold relative risk) and Graves' disease, are common, \[34\] and an association with juvenile rheumatoid arthritis and psoriatic arthritis is described. Glucose intolerance resulting from increased insulin resistance is also common after the age of puberty; in some, this may be due to associated obesity. The risk of type 2 diabetes mellitus is increased. \[35\]

Affected patients are usually small at birth because of intrauterine growth retardation with a mean deficit in length of 2.6 cm (-1.24 SD) and exhibit a slow childhood growth rate that results in a loss of about 8 to 9 cm (-3.0 SD) by age 3 years. \[36\] This study, derived from longitudinal measurements on 47 full-term patients with Turner's syndrome, indicates that the first 3 years of life contribute a major part of the height deficit. Further, there is a decrease in growth rate at the time of expected puberty and failure to undergo a pubertal growth spurt. \[37\] Individuals with Turner's syndrome in the United Kingdom and United States have a mean final height of approximately 142 to 143 cm, about 20 cm less than the average height of normal women; the adult stature of these patients correlates with midparental height and with the height of unaffected women of the same ethnic group. \[38\] Their pattern of growth does not suggest that these individuals are GH-deficient. \[39\] Rather, haploinsufficiency of the \textit{SHOX} gene located in the pseudoautosomal region on the short arm of X and Y (see earlier) is estimated to contribute two thirds of the height deficit. It is postulated that a second gene on the short arm of X that does not undergo X inactivation contributes the other one third of the deficit.

Specific growth curves are available for plotting the growth of affected children. \[40\] In a group of girls with Turner's syndrome and spontaneous puberty, height velocity was transiently higher during puberty than in girls with ameneorrhea, but final adult height was not different (see Chap. 22). GH treatment is now approved by the FDA for Turner's syndrome to increase height. Trials of the effect of treatment with rhGH in children with Turner's syndrome began in 1983. It soon became apparent that hGH administration increased the rate of growth, and data are now available on final height including the results of some randomized dose-response trials from groups around the world.

The average height gain has varied from 4 to 16 cm. \[41\] This variability in gain in height is incompletely understood, but many factors have been implicated including the age of initiation of therapy, dose duration, age of beginning estrogen replacement (especially number of years from beginning hGH treatment), number of injections per week, compliance, and whether the last measured height represented final height. \[42\] The weekly dose of hGH is 0.375 mg/kg divided into seven daily doses. It is important to individualize the dose. The Dutch Advisory Group on Growth Hormone reported that a treatment regimen that gradually increased the dose 0.63 mg/kg per week elicited a 16.0 to 4.1 cm increase in height. It is now apparent that the early initiation of hGH therapy (e.g., 2 to 8 years of age) and a mean duration of treatment of about 7 years can lead to the majority of treated patients achieving a final height greater than 150 cm; in the Dutch study, the mean final height was 162.3 ± 6.1 cm with the high-dose schedule. \[43\] The timing of the introduction of estrogen can have an important effect on final height, \[44\] but with an early age of initiation of GH therapy, low-dose estrogen can be introduced at an appropriate age (about age 13) without compromising adult height. At present, hGH treatment of patients with Turner's syndrome has been safe; untoward events are infrequent.

It is essential that the parents are fully informed about the pros and cons of hGH treatment and that the child is informed in an age-appropriate manner about the use of hGH. The long-term protocol is laborious and expensive and the treatment is invasive but, with the use of new devices for administration, minimally painful.
The appearance of pubic hair is often delayed in the syndrome of gonadal dysgenesis, even though adrenarche, as assessed by the increase in concentration of plasma DHEAS, occurs at the normal age. The pubic hair of affected individuals is sparse, but estrogen therapy increases the growth of pubic hair despite a lack of increase in adrenal androgen secretion. The streak gonads result in sexual infantilism; rarely, probably in about 10% of cases, puberty, menarche, and, even more rarely, pregnancy may occur.

As described earlier, skeletal abnormalities are a common feature and may affect areal BMD determinations. Most untreated children have a normal volumetric bone density (by phalangeal radiographic absorptiometry), in contrast to the less accurate areal BMD in patients with Turner's syndrome, compared with normal girls. Recent studies have not confirmed an increased prevalence of wrist fractures. In an HGH-treated group of girls in whom estrogen therapy was begun at 12 years of age, BMD SD scores were above the mean value. When peripheral quantitative CT, a state-of-the-art technique for assessing volumetric BMD, was used to determine volumetric BMD at two radial sites, a decrease in radial bone mass was found because of a reduction in cortical bone thickness (the endocortical bone surface) with a corresponding increase in bone marrow cross-sectional area, but trabecular volumetric BMD was normal. These reports reiterated the limitations of estimates of areal BMD by DEXA scans in contrast to volumetric BMD in Turner's syndrome. Human GH treatment in childhood appears to have a positive effect on BMD. Estrogen therapy is critical for the prevention and repair of osteoporosis. Although estrogen therapy is important in adolescents and adults, the optimal dose preparation and site of delivery for the prevention of osteoporosis are not known. Nor are data available on the usefulness or adverse effects of continuing of rhGH treatment after final height is achieved in adolescence.

The IQ is normal when verbal ability including comprehension and vocabulary is considered, but spatiotemporal processing, visuomotor coordination, mathematical ability (particularly in geometry) may be impaired, leading to a decrease in the performance IQ. The neurocognitive phenotype associated with Turner's syndrome maps to distal Xp.

Girls with Turner's syndrome have a normal VIG but a discrepancy between VIQ and PIQ, with the latter about 1 SD below the mean value. This is opposite to the pattern found in Klinefelter's syndrome (see earlier). Turner's syndrome is associated with impaired visuconstructional or visual-perceptual abilities in association with executive dysfunction and decreased attention span, which can lead to learning difficulties. In most studies visual-spatial abilities are impaired in girls with Turner's syndrome, but one study found no specific deficits in visual-spatial or tactile-spatial tasks.

Normal girls perform better on motor tasks as they get older, a trend that is not found in Turner's syndrome. Although one study found diminished lateralization of motor tasks, a later study found no difference in lateralization and performance of motor tasks in individuals with Turner's syndrome compared with control subjects; there was superior performance of the dominant or right hand in contrast to the nondominant or left hand.

Anatomic changes of the brain and neurocognitive changes occur in Turner's syndrome. There are consistent MRI abnormalities in the right parietal lobe and the occipital lobes that show decreased volumes in these areas implicated in visuospatial processing. Using positron emission tomography, Murphy and colleagues showed decreased glucose metabolism in the right parietal and occipital lobes. These anatomic data relate to the difficulties in visual-spatial skills found in most studies of girls with Turner's syndrome because these problems are most closely linked to the right parietal region.

Girls with Turner's syndrome resemble normal girls in verbal and language skills, but there are frequently difficulties with memory and attention and decreased arithmetic skills, related to mistakes on operation and alignment processes; and girls with 45,X mosacom with a 46,XX cell line, 45,X/46,XX, scoring closer to normal than those with other types of mosaicsisms. However, girls with Turner's syndrome can score higher on reading achievement tests than predicted by I.Q. or age; this hyperliteracy is a strength in many girls with this disorder. Only 3.3% of girls with Turner's syndrome have mental retardation in the absence of a variant of the syndrome caused by a ring X chromosome.

Because most girls lack ovarian estrogen production and are treated during the teenage years with exogenous estrogen (hormone replacement therapy), Turner's syndrome provides an opportunity to determine the effects of estrogen on neurocognitive function. There appear to be estrogen-dependent and estrogen-independent tasks that are affected in Turner's syndrome. However, there is controversy about which are estrogen-dependent and which are not. For example, visual-spatial perceptual deficits appear to begin in childhood and persist into adult life unaffected by estrogen treatment according to some studies. However, another group of studies show that the effects of estrogen are dependent on dose and time, with untreated patients and those treated with high-dose or long-term therapy performing equally poorly.

Thus, girls with Turner's syndrome improve spontaneously in testing on visuospatial abilities from younger than 12 years to older than 15 years of age, and the older girls treated with estrogen for 3 to 24 months improved further. However, the ability of those treated with estrogen for more than 2 years decreased to that of untreated age-matched Turner's patients. A decrease in spatial abilities has been reported in normal girls going through puberty; it is suggested that this is another example of a biphasic effect of estrogen, with low levels, as in puberty, fostering spatial ability but higher levels, at the end of puberty and thereafter, suppressing such ability. Some motor-related skills and nonverbal processing were performed more rapidly in estrogen-treated girls with Turner's syndrome. Study of event-related potentials in preteen and post-teen girls with Turner's syndrome compared with age-matched controls demonstrated congenital and age-related abnormalities; age-related abnormalities are later ameliorated by estrogen treatment if started early enough. The results of a double-blind study of estrogen versus placebo treatment for 1 to 3 years in 7- to 9-year-old girls with Turner's syndrome show direct effects of estrogen. These girls were part of a larger double-blind study to determine the effect of GH and estrogen on final height. GH therapy in Turner's syndrome led to an increased growth rate. In addition, the girls felt better about their attractiveness, intelligence, and popularity; they perceived that they experienced less teasing, and there was no effect on school performance with GH therapy. Girls with Turner's syndrome younger than 6 years did not perceive that they had a problem with height, but by 7 to 12 and especially 13 to 15 years, affected girls have a strong desire for GH therapy and even unrealistic expectations of what GH therapy might accomplish in terms of adult height. Estrogen therapy improved self-esteem even if there remained a significant difference in height between Turner girls and RGW normal range. GH did not affect the nonverbal neuron-cognitive defects in Turner's syndrome, nor did it affect IQ or achievement scores.

Placebo-treated girls with Turner's syndrome performed less well than either control or estrogen-treated individuals with Turner's syndrome in recall of digit span backward and immediate and delayed recall of the children's word list, suggesting that estrogen replacement therapy improves verbal and nonverbal memory. This observation raises the possibility that a prepubertal deficit in estrogen may affect performance of these tasks as well as suggesting a potential role for low-dose estrogen replacement therapy in late prepuberal girls with Turner's syndrome. These results are similar to the improvement in short-term and long-term verbal memory found in postmenopausal or surgically castrated women treated with estrogen replacement therapy.

Treatment with estrogen for more than 4 years appeared to move the scores for girls with Turner's syndrome on self-esteem and psychological well-being toward normal control values as they reached 16 years of age, compared with a significant difference documented at 12 years of age. No such change occurred in the nonestrogen-treated group, suggesting that the estrogen effects caused the change rather than the passage of years.
Several mechanisms are proposed to explain these estrogen effects: (1) estrogen acts as a neuromodulator in a transient time frame, (2) estrogen alters synapse formation and remodeling in a permanent time frame, or (3) both mechanisms. Thus, estrogen may function as an organizational agent in the brain of the young but as a stabilizing agent in older individuals.

The increase in mental health problems documented in Turner's syndrome may be rooted in the increased peer ridicule experienced by girls with Turner's syndrome as opposed to a biologic abnormality. The teasing can by itself lead to decreased self-image and depression. There is an increased risk of impaired social adjustment in Turner's syndrome. 45,X females with Turner's syndrome have higher ratings of social and attention problems and withdrawn behaviors than their own sisters. The risk is higher if the single X chromosome comes from the mother (X0) rather than the father (X0). Individuals with Turner's syndrome have difficulty in inferring affective intention from facial appearance. As an explanation for these phenomena, there appears to be a locus on the X chromosome on Xq or close to the centromere on Xp that escapes X inactivation and affects social cognition. When the locus was inherited from the father, the 45,X individuals were significantly better adjusted, with superior verbal and higher order executive function skills, which mediate social interactions. If expressed only on the paternally derived X chromosome, the existence of this putative locus may explain, in part, why 46,XY males (whose single X chromosome is maternal) are more vulnerable to developmental disorders of language and social cognition, such as autism, and 46,XX males. Additionally, in addition to haploinsufficiency and its consequences, sex chromosome imprinting appears to be a factor in the Turner syndrome phenotype.

In the first example of an imprinted gene on the X chromosome, the locus resides in the pericentric region of the short arm or on the long arm of the X chromosome. Skuse and coworkers postulated that this imprinted gene may play a role in male-female differences in social behavior and developmental disorders. It is useful to monitor the patient's progress in high school mathematics. Gender identity and sexual orientation are female. Although it has been generally accepted that these patients do well in their psychological development, consistent with the report of Skuse and associates. A study of 103 children with Turner's syndrome demonstrated a significant decrease in social competence score, an increase in total behavior problems, and social and attention problem scales with difficulty in schooling, in peer relationships, and in concentration as well as immaturity, hyperactivity, and nervousness; the origin of the X chromosome was not determined. Structural abnormalities of the X chromosome were associated with more behavior problems than a missing X chromosome or mosaicism of the X chromosomes. Hyperactive behavior usually improves after the age of puberty. Mental retardation and a "severe" phenotype are associated with a small ring X chromosome to undergo X inactivation resulting in X chromosome disomy for genes that undergo X inactivation.

The origin of the single X chromosome affects memory performance as well because those with a maternally derived X have increased verbal forgetting but normal nonverbal forgetting whereas those with a paternally derived X chromosome have the opposite pattern of problems. Sex Chromatin Positive Variants of the Syndrome of Gonadal Dysgenesis

Mosaicism of 45,X/46,XX, 45,X/47,XXX, or 45,X/46,XX/47,XXX chromosomes is associated with a chromatin-positive buccal smear and usually fewer manifestations of the syndrome of gonadal dysgenesis. Likewise, structural abnormalities of the X chromosome can be associated with fewer phenotypic features of the syndrome. Lack of genetic material on the long or the short arm of the second X chromosome can cause decreased gonadal function; loss of all or part of the short arm of the X leads to the physical findings of Turner's syndrome (see Chapter 22). Depending on the location and extent of the deletion on the short arm of the X chromosome, these patients are more likely to have modest pubertal growth and some spontaneous pubertal development.

Sex Chromatin Negative Variants of Gonadal Dysgenesis

These variants include 45,46,XY mosaicism and structural abnormalities of the Y chromosome. Affected individuals vary in phenotype from that of classical gonadal dysgenesis to that of ambiguous genitalia to phenotypic males. Patients present with short stature, delayed puberty, and a history of hypospadias repair. There is variable testicular differentiation, ranging from a streak gonad to functioning testes. Patients with mosaicism involving a Y cell line or abnormalities of the Y chromosome are at risk for neoplastic transformation of the dysgenetic testes. Gonadoblastomas, benign nonmetastasizing tumors, may arise within the gonad and produce either testosterone or estrogen; the neoplasm may be calcified sufficiently to be detected on an abdominal radiograph. Thus, the appearance of feminization or virilization in a patient with dysgenetic gonads and a Y cell line may indicate gonadoblastoma formation. Of greater significance is the increased prevalence of malignant germ cell tumors arising within the dysgenetic gonad or gonadoblastoma. Such tumors occur more often in postpubertal subjects and rarely in children. The management of gonads in patients with a Y cell line is discussed in Chapter 22.

46,XX and 46,XY Gonadal Dysgenesis

Pure gonadal dysgenesis refers to phenotypic females with sexual infantilism and a 46,XX or 46,XY karyotype without chromosomal abnormalities. Familiar and Sporadic 46,XX Gonadal Dysgenesis and Its Variants

The usual phenotype of 46,XX gonadal dysgenesis includes normal stature, sexual infantilism, bilateral streak gonad, normal female internal and external genitalia, and primary amenorrhea. The streak gonad occasionally produces estrogen or androgens, but malignant transformation is rare. Incomplete forms of this condition may result in hypoplastic ovaries that produce enough estrogen to cause some breast development and a few menstrual periods followed by secondary amenorrhea. This heterogeneous syndrome occurs sporadically or with autosomal recessive inheritance and in some instances is associated with other congenital malformations; some familial cases have been associated with sensorineural deafness (Perrault's syndrome) (see later).

Familiar and Sporadic 46,XY Gonadal Dysgenesis and Its Variants

A phenotype that includes female genitalia with or without clitoral enlargement, normal or tall stature, bilateral streak gonads, normal müllerian structures, sexual infantilism, and a eunuchoid habitus is typical of 46,XY gonad dysgenesis. About 15% of the patients have a deletion or mutation in the SRY gene. If the dysgenetic testes produce significant amounts of testosterone, slight clitoral enlargement may occur at birth and virilization may ensue at puberty. The incomplete form of 46,XY gonadal dysgenesis may involve any degree of ambiguity of the external genitalia and internal ducts. The risk of neoplastic transformation of the streak gonads or dysgenetic testes is increased, and gonadectomy is indicated. The disorder is usually transmitted as an X-linked or sex-limited autosomal dominant trait, less commonly as an autosomal recessive trait. A novel homozygous missense mutation in exon 1 of the desert hedgehog gene was reported in a patient with polyneuropathy associated with partial 46,XY gonadal dysgenesis.

Other Causes of Primary Ovarian Failure

Primary ovarian failure is increasing in prevalence as a consequence of the long-term effects of cytotoxic chemotherapy and radiation as these agents prolong life in children and adolescents with cancer. Radiation Therapy

Radiation therapy that includes the ovaries within the field can cause primary ovarian failure; a dose of 4 Gy to the ovaries leads to sterility in 30% of young women and 100% of older women. It is useful surgically to move the ovaries out of the radiation field; ovarian transposition before radiation therapy is compatible with normal menses, pubertal development, and pregnancy in most cases. The uterus may also be affected by radiation and may not expand normally during pregnancy. Careful endocrine follow-up of these children and adolescents is essential.
Successful treatment of childhood acute lymphoblastic leukemia is now commonplace. In a large study by Quigley and colleagues, after cytotoxic chemotherapy boys and girls had enhanced germ cell damage as evidenced by increased FSH secretion and boys had decreased testicular size for the stage of puberty. The concentration of plasma inhibin B is usually decreased, a sensitive indicator of damage to the germinal epithelium, and the girls at puberty had evidence of a compensated decrease in ovarian follicular function. Quite likely as a result of cranial radiation, the mean age of menarche was advanced about 12 months despite the primary ovarian damage; puberty was not advanced in the boys. The type of chemotherapy is related to the effects on the gonads. Nitroso compounds (carmustine and lomustine) or procarbazine for the treatment of brain tumors has been linked to primary gonadal failure manifested by elevated plasma gonadotropin levels in boys and girls; the boys had small testes but were able to secrete adequate testosterone for their pubertal stage. Adjuvant chemotherapy for localized osteosarcoma in the prepubertal period is compatible with ovarian function and fertility.

Although it was previously thought that cancer therapy does not cause gonadal damage in prepubertal individuals, current evidence suggests otherwise. Prepubertal boys and girls treated with abdominal radiation for Wilms' tumor plus chemotherapy (dacarbazine, vincristine with Adriamycin, or cyclophosphamide in most) may experience gonadal damage, whereas those given chemotherapy alone usually do not. The ovary is less vulnerable to the effects of radiation and chemotherapy than the testis. The prevalence of ovarian damage does not appear to be as high as testis. Regular menses are reported in a majority of females treated as children. Nevertheless, age at treatment has a significant role: treatment between 13 and 19 years of age was associated with a more than twofold increase in premature ovarian failure during the third decade. Attempts to protect the gonads by suppressing the pituitary-gonadal axis with gonadal steroids or LHRH agonists are ineffective.

Autoimmune Oophoritis

Premature menopause may occur at any age before the normal climacteric and has been reported in adolescent girls; cessation of ovarian function usually occurs as secondary amenorrhea. Autoimmune oophoritis can cause ovarian failure leading to primary amenorrhea, oligomenorrhea, arrest of puberty, and occasionally cystic enlargement of the ovaries. Most often it is associated with other autoimmune endocrinopathies, especially autoimmune Addison's disease, in which it may precede the onset of adrenal insufficiency, but it rarely, if ever, occurs in isolated premature ovarian failure.

Autoimmune oophoritis is present in more than 20% of patients with autoimmune adrenal insufficiency. Various autoantibodies have been detected in autoimmune oophoritis, including autoantibodies to cytotome P450 steroidogenic enzymes (P450scc, P450c17, P450aromatase), some are organ-specific, whereas others react with antigens in more than one tissue and more than one cell type. Glucocorticoid therapy may improve, at least temporarily, ovarian function.

Miscellaneous Causes of Ovarian Failure

Homogamous galactosemia is commonly associated with primary ovarian failure (from failure to develop puberty to primary or secondary amenorrhea and premature menopause), but puberty is usually normal in males and the risk of testicular dysfunction is low; compound heterozygotes have normal onset of puberty.

Dyspasia and Premature Ovarian Failure

A rare autosomal dominant disorder involving eyelid dysplasia and premature ovarian failure is due to haploinsufficiency of the FOXL2 gene, a member of the winged helix/tailless family of transcription factors. The eyelid abnormalities include small palpebral fissures, ptosis, and a skinfold extending inward and upward from the lower lid (epicanthus inversus). The gene is expressed in the follicular cells and the mutations that lead to haploinsufficiency are associated with an increased rate of follicular atresia; the degree of ovarian failure is variable from primary amenorrhea to irregular menses and premature ovarian failure, ranging from normal-appearing ovaries on ultrasonography to streak gonads and on ovarian biopsy a variable number of primordial follicles. The infertility component of the syndrome is limited to the female.

Congenital disorders of glycosylation-1 (carbohydrate-deficient glycoprotein syndrome type I) constitute an autosomal recessive disorder associated with circulating glycoproteins deficient in their terminal carbohydrate modifications, including a wide range of glycoproteins, enzymes, binding proteins, and coagulation factors. A typical isoform pattern of serum transferrin on isoelectric focusing is used as a diagnostic test. The dominant clinical features are neurologic manifestations of involvement of the central and peripheral nervous system. Among the other organ systems affected is the pituitary-gonadal system.

The hypergonadotropic hypogonadism is more severe in females than males who virilize at puberty. There are two interesting aspects: both the ovary and the pituitary gland are affected. The affected girls have sexual infantilism; the ovaries are hypoplastic or atrophic. High serum FSH and LH levels exhibit normal electrophoretic patterns but appear to have decreased but not absent FSH bioactivity in an FSH bioassay, and in the only three girls tested, a response to the administration of human menopausal gonadotropin was indicated by an increase in serum estradiol and, in one patient, ovarian follicular growth. These observations suggest an abnormality in both the FSH molecule and the ovary, in the latter case a defect in the configuration and activation of the FSH receptor itself and the binding of ligand or a postreceptor defect.

Resistant ovary is a heterogeneous cause of primary hypogonadism, a syndrome associated with elevated concentrations of plasma FSH and LH and ovaries that contain primordial follicles. The syndrome is usually idioptic, but an increasing number of genetic abnormalities have been described in addition to the more common X-chromosomal defects.

Follicle-Stimulating Hormone Receptor Gene Mutations and Hypergonadotropic Hypogonadism

The FSH receptor is a member of the G protein-linked receptor seven-transmembrane superfamily; it has a large, extended extracellular ligand-binding domain. The eye of abnormalities include small palpebral fissures, ptosis, and a skinfold extending inward and upward from the lower lid (epicanthus inversus). The gene is expressed in the follicular cells and the mutations that lead to haploinsufficiency are associated with an increased rate of follicular atresia; the degree of ovarian failure is variable from primary amenorrhea to irregular menses and premature ovarian failure, ranging from normal-appearing ovaries on ultrasonography to streak gonads and on ovarian biopsy a variable number of primordial follicles. The infertility component of the syndrome is limited to the female.

The FSH receptor mutation in the Finnish patients is not a null mutation. It remains to be determined whether the loss or complete inactivation of the FSH receptor leads to failure of puberty and sexual infantilism or to estrogen synthesis by the immature ovarian follicles described in the FSH subunit knockout mouse. Affected males in these families are normally masculinized at puberty but tend to have small testes. They have a variable degree of spermatogenic insufficiency but not azoospermia, increased plasma concentrations of FSH and LH, decreased inhibin levels, and normal plasma testosterone values.
could be associated with delayed puberty. In the affected female, on the other hand, LH/hCG resistance does not affect pubertal maturation but leads to amenorrhea with high serum LH levels but normal FSH and estradiol concentrations.\footnote{1508}

**Polycystic Ovary Disease**

Polycystic ovary disease or functional ovarian hyperandrogenism does not delay the onset of puberty but often delays menarche or causes menstrual abnormalities\footnote{1509} (see Chapter 22).

**Noonan's Syndrome (Pseudo-Turner's Syndrome, Ullrich's Syndrome)**

Individuals with Noonan's syndrome (see Chapter 29) have webbed neck, phtosis, down-slaning palpebral fissures, low-set ears, short stature, cubitus valgus, and lymphedema, and hence this phenotype has been called pseudo-Turner's syndrome.\footnote{1510} Features that differentiate these individuals from those with Turner's syndrome include triangular faces, pectus excavatum, right-sided heart disease (e.g., pulmonic stenosis), often with valve dysplasia, or atrial septal defect, hypertrophic cardiomyopathy, varied blood clotting defects, and an increased incidence of mental retardation. Females with Noonan's syndrome have normal ovarian function. Males have normal differentiation of external genitalia but may have undescended testes; germinal aplasia or hypoplasia and impaired Leydig cell function may be present.\footnote{1511} Puberty may be delayed an average of 2 years.\footnote{1512} Stature is decreased, usually following the -2 SD curve; the pubertal growth spurt is often delayed or attenuated but final height is usually at the low limits of normal.\footnote{1513} Administration of hGH increased the growth rate, but long-term effects are not yet available.\footnote{1514}

Adult males with Noonan's syndrome are reported to have osteopenia, which has been attributed to estrogen deficiency because estrogen administration improves the decreased bone mineral content.\footnote{1515} Patients who have sufficient ovarian function to undergo spontaneous pubertal development and menarche have adequate BMD compared with those in whom puberty was induced by estrogen and who exhibited osteopenia despite hormone replacement.\footnote{1516} Noonan's syndrome is usually inherited as an autosomal dominant trait.\footnote{1517} A gene implicated in Noonan's syndrome has been localized to the long arm of chromosome 12 (12q 24.1).\footnote{1518} Mutations in PTPN11, the gene encoding the non-receptor protein tyrosine phosphatase SHP-2, which contains RE Src homology-2 (SH2) domains, account for about 50% of cases. The incidence is estimated as 1 in 1000 to 1 in 5000. One parent may have features of the syndrome in 40% to 60% of cases. About 50% of cases are thought to be the result of new mutations.

**Frasier Syndrome**

Chronic renal failure is combined with gonadal dysgenesis in the Frasier syndrome.\footnote{1519} This diagnosis should be considered in any phenotypic female with end-stage renal disease (related to focal segmental glomerulosclerosis) and sexual infantilism; the karyotype may be 46,XY or 46,XX.\footnote{1520}

**Diagnosis of Delayed Puberty and Sexual Infantilism**

See Figure 24-55 and Figure 24-56 and Table 24-26. Signs of puberty have not yet appeared in 0.4% of normal boys by age 13 years old or 15% of normal girls by age 13 years old.\footnote{1521}

The diagnosis of hypergonadotropic hypogonadism is readily established by elevation of random plasma LH and FSH concentrations. However, the differential diagnosis of hypergonadotropic hypogonadism versus constitutional delay in growth and adolescence is more difficult because of the overlap in physical and laboratory findings in the two conditions, including inability to differentiate between normal and low concentrations of serum gonadotropins  (Table 24-27). No single test consistently makes this distinction. The majority of boys with pubertal delay have a self-limited variant in the tempo of growth and pubertal onset. The task of the physician is, on the one hand, to avoid a costly investigation of an essentially healthy boy while, on the other, identifying those who have an underlying disorder compromising the hypothalmic-pituitary-gonadal system.

Unfortunately, many patients with Kallmann's syndrome wait years for the correct diagnosis to be made even in the presence of classical findings\footnote{1522}; a high index of suspicion by the physician is important.

A presumptive diagnosis can usually be formed during the initial evaluation on the basis of the history and physical examination. Has puberty failed to occur or did it begin but fail to progress or regress? History taking must elicit all symptoms of chronic or intermittent illnesses and all details pertaining to growth and development as well as questioning about the patient's sense of smell. Disorders of pregnancy, abnormalities of labor and delivery, and birth trauma, if present in the patient's history, suggest that a congenital or neonatal event may be related to the delay in puberty. Poor linear growth and poor nutritional status during the neonatal period and childhood may reflect long-standing abnormalities of development. A growth chart is plotted to represent graphically the increase in stature and to assess growth velocity from birth. Late onset of growth failure usually indicates a serious condition requiring evaluation as soon as it is noted. Family history may reveal disorders of puberty or infertility, anosmia, or hyposmia in relatives as well as delay in the age of onset of puberty in parents or siblings. Recalled age of pubertal onset is relatively reliable in women\footnote{1523} but less often accurate in men. A history of consanguinity is important in the detection of autosomal recessive disorders.

The physical examination starts with determination of height and weight; the upper/lower segment ratio or sitting height is calculated, and the arm span is measured and compared with

---

**Figure 24-55** The evaluation of delayed puberty in boys.

**Figure 24-56** The evaluation of delayed puberty in girls.
**TABLE 24-26 -- Differential Diagnostic Features of Delayed Puberty and Sexual Infantilism**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Stature</th>
<th>Plasma Gonadotropins</th>
<th>LHRH Test LH Response</th>
<th>Plasma Gonadal Steroids</th>
<th>Plasma DHEAS</th>
<th>Karyotype</th>
<th>Olfaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constitutional delay in growth and adolescence</td>
<td>Short for chronologic age, usually appropriate for bone age</td>
<td>Prepubertal, later pubertal</td>
<td>Prepubertal, later pubertal</td>
<td>Low, later normal</td>
<td>Low for chronologic age, appropriate for bone age</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td><strong>Hyponadotrophic hypogonadism</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Isolated gonadotropin deficiency</td>
<td>Normal, absent prepubertal growth spurt</td>
<td>Low</td>
<td>Prepubertal or no response</td>
<td>Low</td>
<td>Appropriate for chronologic age</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td>Kallmann's syndrome</td>
<td>Normal, absent prepubertal growth spurt</td>
<td>Low</td>
<td>Prepubertal or no response</td>
<td>Low</td>
<td>Appropriate for chronologic age</td>
<td>Normal</td>
<td>Anosmia or hyposmia</td>
</tr>
<tr>
<td>Idiopathic multiple pituitary hormone deficiencies</td>
<td>Short stature and poor growth since early childhood</td>
<td>Low</td>
<td>Prepubertal or no response</td>
<td>Low</td>
<td>Usually low</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td>Hypothalamic-pituitary tumors</td>
<td>Late onset decrease in growth velocity</td>
<td>Low</td>
<td>Prepubertal or no response</td>
<td>Low</td>
<td>Normal or low for chronologic age</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td><strong>Primary gonadal failure</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syndrome of gonadal dysgenesis (Turner's syndrome) and variants</td>
<td>Short stature since childhood</td>
<td>High</td>
<td>Hyper-response for age</td>
<td>Low</td>
<td>Normal for chronologic age</td>
<td>45,X or variant</td>
<td>Normal</td>
</tr>
<tr>
<td>Klinefelter's syndrome and variants</td>
<td>Normal to tall</td>
<td>High</td>
<td>Hyper-response at puberty</td>
<td>Low or normal</td>
<td>Normal for chronologic age</td>
<td>47,XXY or variant</td>
<td>Normal</td>
</tr>
<tr>
<td>Familial XX or XY gonadal dysgenesis</td>
<td>Normal</td>
<td>High</td>
<td>Hyper-response for age</td>
<td>Low</td>
<td>Normal for chronologic age</td>
<td>46,XX or 46,XY</td>
<td>Normal</td>
</tr>
</tbody>
</table>

The length and width of the tests are measured or the volume is assessed using an orchidometer. The length and diameter of the stretched penis are determined in boys, and the diameter of glandular breast tissue and areolar size are noted in girls. The presence or absence of galactorrhea is defined. Obese boys often appear to have a small penis because of excessive adipose tissue surrounding the phallus; therefore, the determination of pubic and axillary hair, and the degree of acne, is important in the search for a chronic disorder that may delay puberty.

Laboratory studies (Table 24-28) include determination of plasma LH and FSH concentrations, measurement of the rise in LH level after LHRH administration, determination of testosterone concentrations in boys and estradiol levels in girls, and measurements of T4 and prolactin concentrations in boys and girls if the clinical features warrant. It is important to use one of the few national endocrine laboratories for the determinations of the hormones of puberty because most local laboratories are interested only in differentiating normal, higher, adult values from inappropriately low levels and not the low concentrations characteristic of the early stage of pubertal development. For example, levels of estradiol below 15 pg/mL are not measured routinely or with confidence in many clinical laboratories despite the availability of methods and commercial kits to measure accurately values as low as 1.5 pg/mL.

---

**TABLE 24-28 -- Endocrine and Imaging Studies in Delayed Adolescence**

<table>
<thead>
<tr>
<th>Initial assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plasma testosterone or estradiol</td>
</tr>
<tr>
<td>Plasma FSH and LH</td>
</tr>
<tr>
<td>Plasma thyroxine (or prolactin)</td>
</tr>
<tr>
<td>Bone age and lateral skull roentgenograph</td>
</tr>
<tr>
<td>Test of olfaction</td>
</tr>
</tbody>
</table>

**Follow-up studies**

- Karyotype (short, phenotypic females)
- MRI with contrast enhancement
- Pelvic ultrasonography (females)
- LHRH test
- hCG test (males)
- Pattern of pulsatile LH secretion
- Visual acuity and visual fields
FSH, follicle-stimulating hormone; hCG, human chorionic gonadotropin; LH, luteinizing hormone; MRI, magnetic resonance imaging.

Radiographic examination includes bone age determination and, if the diagnosis is at all consistent with a CNS lesion, an MRI of the brain with specific attention to the pituitary and hypothalamic area using contrast; only advanced pituitary tumors or significantly calcified craniohypophysis appear on lateral skull films, and, although a positive result is useful, a negative radiograph cannot rule out a CNS defect. In contrast to MRI scans, CT scanning can detect calcification. Ultrasonic evaluation of the uterus and ovaries is not usually indicated initially in work-up of delayed puberty but provides useful information about the state of development of these structures. Again, it is important that the ultrasonographer has experience with children and young adolescents. Regrettably, individuals with normal internal genital organs have been told that they lack a uterus or ovaries, or both, by ultrasonographers inexperienced with this age group. One study demonstrated streak gonads in 50% of the patients with Turner’s syndrome who were choosing orif therapy. Assessment of chromosomal abnormalities should be considered in all short girls, even in the absence of somatic signs of Turner's syndrome and especially if puberty is delayed, and in boys with suspected Klinefelter stigma or behavior.

A presumptive diagnosis of constitutional delay in growth and adolescence is made if the history and growth chart reveal a history of short stature but consistent growth rate for bone age (and no signs or symptoms of hypothalamic lesions), if the family history includes parents or siblings with delayed puberty, if the physical examination (including assessment of the olfactory threshold) is normal, if optic discs and visual fields are normal, and if the bone age is significantly delayed. In classical cases, an MRI scan of the hypothalamic-pituitary region may not be necessary. The rate of growth in these patients is usually appropriate for bone age; a decrease in growth velocity occurs in some normal children just before the appearance of secondary sexual characteristics and may awaken concerns if such a pattern occurs in these subjects. Further, in these individuals the onset at puberty correlates better with bone age than with chronologic age. Elevated concentrations of gonadotropins and gonadal steroids to early pubertal levels precede secondary sexual development by several months; thus, measurements of serum LH, FSH, estradiol, or testosterone levels may help in predicting future development. The third-generation LH assays are reported to be sufficiently sensitive to allow the determination of the onset of endocrine puberty with a single blood sample in most boys, but an LH/FSH test is still often performed. An increase in the concentration of LH of more than 7.5 IU/L (2 ng/mL LER-960) determined by conventional polyclonal radioimmunoassay after intravenous administration of 100 pg of LH usually precedes the first physical sign of sexual maturation by less than 1 year.

Clomiphene citrate, an antiestrogen with weak estrogenic effects, decreases secretion of gonadotropins in pubertal patients but increases gonadotropin secretion in pubertal patients and in adults. However, we have not found administration of clomiphene citrate to be useful in the diagnosis of constitutional delay of growth and adolescence.

Various tests have been proposed for differentiating hypogonadotrophic hypogonadism from constitutional delay in puberty. Trials assessing the prolactin response to TRH, chlorpromazine, metoclopramide, or domperidone for differential diagnosis either failed or gave inconsistent results. The combination of the prolactin response to metoclopramide and the gonadotropin response to LHRH has been suggested, as has the use of priming doses of LHRH with evaluation of the gonadotropin response to a subsequent dose of LHRH or to a superactive LHRH agonist. The FSH response is higher in patients with hypothalamic-pituitary deficiencies who undergo development. A sensitive immunofluorometric assay for LH may help to distinguish between constitutional delay of growth and adolescence and hypogonadotropic hypogonadism better than the polyclonal LH radioimmunoassay. Urinary gonadotropin excretion is lower in hypogonadotropic patients than in delayed puberty, but this method of differential diagnosis may require years of observation before the difference is apparent. Although some methods are promising, their efficacy remains to be confirmed. There is a tendency for hypogonadotropic patients to undergo adrenarche at a normal age and to have a higher DHEAS concentration than those with constitutional delay in growth, and this pattern is helpful in the differential diagnosis.

Measurement of L as serum testosterone is proposed to be an accurate indication of impending pubertal development; a value greater than 0.7 nmol/L (20 ng/dL) predicts enlargement of testes to greater than 4 mL by 12 months in 77% of cases and by 15 months in 100% of cases, whereas of those with a value less than 0.7 nmol/L only 12% entered puberty in 12 months and only 25% entered puberty in 15 months. This technique may help predict spontaneous pubertal development but still requires considerable watching and waiting.

At present, there does not appear to be a practical and reliable endocrine test for indisputably differentiating between constitutional delay in growth and adolescence and hypogonadotropic hypogonadism. Watchful waiting remains the procedure of choice.

A typical patient with isolated gonadotropin deficiency is of average height for age and has eunuchoid proportions; low plasma concentrations of gonadal steroids, LH, and FSH, and no increase or a blunted response of LH after LHRH administration. The amplitude and usually the frequency of LH pulses are decreased when serial blood samples are studied over a 24-hour period. In some but not all forms of Kallmann’s syndrome, the sense of smell is absent or impaired. However, differentiation of isolated gonadotropin deficiency in the absence of hyposmia or anosmia from constitutional delay in puberty may be difficult at initial study. Gonadotropin-deficient patients may be as short as those with constitutional delay in growth and adolescence, and concentrations of LH and FSH in hypogonadotropic hypogonadism may be indistinguishable from those of normal prepubertal children or children with constitutional delay. Sometimes years of observation are necessary to detect the appearance of spontaneous and progressive signs of secondary sexual development or to document rising concentrations of gonadotropins or gonadal steroids before the diagnosis is clear. In general, but not in all cases, absence of the first signs of sexual maturation or failure of a rise in gonadotropins or gonadal steroid levels by age 18 in the presence of a normal concentration of serum DHEAS for chronologic age supports the diagnosis of isolated gonadotropin deficiency.

Patients with deficiency of gonadotropins combined with deficiency of other pituitary hormones require careful evaluation for a CNS neoplasm. Visual field or optic disc abnormalities support the diagnosis of CNS tumor; even if these tests are normal, cranial MRI should be performed to evaluate the pituitary gland and stalk and the hypothalamic region. CT scans but especially MRI scans of the head are valuable in detecting mass lesions and developmental abnormalities of the hypothalamic-pituitary region.

Treatment of Delayed Puberty and Sexual Infertility

Treatment of delayed puberty (Table 24-29) depends on the diagnosis and the nature of the disorder. Patients with constitutional delay in growth and adolescence ultimately have spontaneous onset and progression through puberty. Often, reassurance and continued observation to ensure that the expected sexual maturation occurs are sufficient. However, the stigma of appearing less mature than one's peers can cause psychological stress; such individuals may be unable to participate in the dating activities their friends are starting. Smaller size may lead them to avoid participation in sports and physical activities, in some but not all forms of Kallmann’s syndrome, the sense of smell is absent or impaired. However, differentiation of isolated gonadotropin deficiency in the absence of hyposmia or anosmia from constitutional delay in puberty may be difficult at initial study. Gonadotropin-deficient patients may be as short as those with constitutional delay in growth and adolescence, and concentrations of LH and FSH in hypogonadotropic hypogonadism may be indistinguishable from those of normal prepubertal children or children with constitutional delay. Sometimes years of observation are necessary to detect the appearance of spontaneous and progressive signs of secondary sexual development or to document rising concentrations of gonadotropins or gonadal steroids before the diagnosis is clear. In general, but not in all

1200

For psychological reasons, in boys of age 14 or older who show no signs of puberty, a 4- to 6-month course of testosterone enanthate, cypionate, or propionate (50 to 100 mg intramuscularly every 4 weeks) may be helpful. The low dose of testosterone enanthate is generally considered to be safe but can raise LDL and lower HDL cholesterol values as an expected effect.

Oral treatment with 2.5 mg of fluoxymesterone (Halotestin) for 6 to 60 months allows increased pubertal development without adverse effects on final height, although the necessity to take a daily dose may decrease compliance. Low-dose oxandrolone (2.5 mg/day orally) is sometimes used as an oral alternative to intramuscular testosterone enanthate; this agent increases growth through androgenic effects reflected by suppression of LH and FSH but does not stimulate GH secretion as it is not aromatized to estrogen. The temporary increase in growth velocity found with oxandrolone does not affect final height. Short-term treatment with fluoxymesterone (2.5 mg/day orally) was also reported to be a safe treatment that does not compromise adult height. Transdermal testosterone may be applied as a daily patch or a gel, although experience with these forms of androgen is more limited than with the other forms. Preliminary experience suggests that overnight (approximately 8 to 9
hours) or every other night use of 2.5 mg is effective. Testosterone gel is being investigated as a daily topical preparation to advance pubertal development.

For girls of age 13 or older, a 3- to 4-month course of ethinylestradiol (5 µg/day orally) or conjugated estrogens (0.3 mg/day orally) may be used to initiate maturation of the secondary sexual characteristics without unduly advancing bone age or limiting final height. A fourth-generation aromatase inhibitor, letrozole, administered along with testosterone in a randomized controlled trial in boys with constitutional delay in puberty and growth decreased the advancement in bone age, an effect that will presumably lead to a greater adult height by strikingly decreasing the synthesis of estradiol from testosterone. Estradiol is the sex steroid that has the major effect on skeletal maturation. Letrozole does not block the virilizing effects of testosterone. This promising treatment is experimental; it may improve the decreased adult height in some boys with constitutional delay compared with their predicted genetic potential, but the slower rate of growth may be accompanied by a decrease that testosterone cannot overcome.

If, during the 3 to 6 months after discontinuing gonadal steroid therapy, spontaneous puberty does not ensue or the concentrations of plasma gonadotropins and plasma testosterone in boys or plasma estradiol in girls do not increase toward pubertal values, the treatment may be repeated. Usually, only one or two courses of therapy are necessary. When treatment is discontinued after bone age has advanced, for example, to 12 to 13 years in girls or 13 or 14 years in boys, patients with constitutional delay usually continue pubertal development on their own, whereas those with gonadotropin deficiency do not progress and may, in fact, regress.

Alleviating the underlying problem treats functional gonadotropin deficiency associated with chronic disease. Delayed puberty in this situation is usually a result of inadequate nutrition and low weight; when weight returns to normal values, puberty usually occurs spontaneously. Treatment with T allows normal pubertal development in hypothyroid patients with delayed puberty.

Congenital or acquired gonadotropin deficiency as a result of a lesion or surgery requires replacement therapy with gonadal steroids at an age approximating the normal age of onset of puberty. An exception may occur when GH deficiency coexists with gonadotropin deficiency; if bone age advancement and epiphyseal fusion are brought about by testosterone or estradiol replacement before therapy with GH causes adequate linear growth, adult height is compromised. However, if puberty is not initiated early enough, the patient may well suffer psychological damage. It is generally advisable to initiate puberty in such patients with low-dose

---

**TABLE 24-29 — Management and Treatment of Delayed Puberty**

<table>
<thead>
<tr>
<th>Objectives</th>
<th>Therapy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Determine site and etiology of abnormality</td>
<td>Concerned but not anxious or socially handicapped adolescent:</td>
</tr>
<tr>
<td>Induce and maintain secondary sexual characteristics</td>
<td><strong>Reassurance and follow-up (tincture of time)</strong></td>
</tr>
<tr>
<td>Induce pubertal growth spurt</td>
<td><strong>Repeat evaluation (including serum testosterone or estradiol) in 6 mo</strong></td>
</tr>
<tr>
<td>Prevent the potential short-term and long-term psychological, personality, and social handicaps of delayed puberty</td>
<td>Psychosocial handicaps, anxiety, highly concerned:</td>
</tr>
<tr>
<td>Ensure normal libido and potency</td>
<td><strong>Therapy for 4 mo with</strong></td>
</tr>
<tr>
<td>Attain fertility</td>
<td><strong>Boys:</strong> testosterone enanthate 100 mg intramuscularly every 4 wk at 14 yr of age, or overnight transdermal testosterone patch</td>
</tr>
<tr>
<td></td>
<td><strong>Girls:</strong> ethinylestradiol 510 µg daily by mouth or conjugated estrogens 0.3 mg daily by mouth or overnight ethinylestradiol patch at 13 yr of age</td>
</tr>
<tr>
<td></td>
<td><strong>No therapy for 48 mo; reevaluate status including serum testosterone or estradiol; if indicated repeat treatment regimen</strong></td>
</tr>
</tbody>
</table>

---

**TABLE 24-30 — Hormonal Substitution Therapy in Boys with Hypogonadism**

<table>
<thead>
<tr>
<th>Goal: to approximate normal adolescent development when diagnosis is established</th>
<th>Induction therapy: at 13 yr of age, testosterone enanthate (or other long-acting testosterone ester) 50 mg intramuscularly every month for about 9 mo (612 mo)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Over the next 3 to 4 yr: gradually increase dose to adult replacement dose of 200 mg every 23 wk</td>
<td>Begin replacement therapy in boys with suspected hypogonadotropic hypogonadism by bone age 14 yr</td>
</tr>
<tr>
<td>To induce fertility at appropriate time: pulsatile LHRH or FSH and hCG therapy</td>
<td>gonadal steroids by age 14 in boys and age 13 in girls regardless of the definitive diagnosis of gonadotropin deficiency, thus, these children with GH deficiency would be treated similarly to those with isolated delayed puberty.</td>
</tr>
</tbody>
</table>

Patients with isolated GH deficiency may have a delayed onset of puberty; with GH administration, puberty usually occurs at an appropriate age but may progress faster than in normal individuals. A study of over 200 children with GH deficiency treated with hGH showed a correlation between the age of onset of induced puberty and final height in patients who were also gonadotropin-deficient, whereas those who underwent spontaneous puberty, which occurred earlier than the age of hormone-induced puberty in the gonadotropin-deficient children, had a lower final height; this supports the advisability of waiting to initiate puberty in GH- and gonadotropin-deficient subjects. Height at the onset of puberty is also correlated with final height in GH-deficient children. Clinical trials are in progress to determine the effects of artificially delaying puberty with an LHRH analogue to attempt to achieve a greater final height in patients with isolated GH deficiency treated with hGH (see earlier).

---

**TABLE 24-31 — Hormonal Substitution Therapy in Girls with Hypogonadism**

Micropenis resulting from fetal androgen deficiency caused by a primary testicular defect or gonadotropin deficiency can be successfully treated with small doses of testosterone enanthate (25 to 50 mg/month intramuscularly) administered for short periods during infancy (also see Chapter 22). Patients with isolated congenital GH deficiency occasionally have micropenis that may be successfully treated with GH replacement alone.

As discussed earlier, episodic administration of LHRH can elicit pulsatile LH and FSH release and gonadal stimulation in prepubertal children or hypogonadotropic patients. Portable pumps have been used to administer LHRH in episodic fashion over prolonged periods. Pulsatile LHRH therapy can induce puberty and promote the development of secondary sexual characteristics and spermatogenesis in men and ovulation in women; pregnancy has been achieved with this regimen in women with hypogonadotropic hypogonadism. A lower frequency of LHRH administration favors FSH secretion and a higher frequency favors LH secretion and, ultimately, has been associated with a PCOS-like picture. A comparison of two different frequencies of LHRH administration did not reveal a difference between an LH pulse given subcutaneously every 3 hours or every 45 minutes in the rapidity of onset of pubertal development or serum LH, FSH, or sex steroid concentrations, this indicates that the hypothalamic-pituitary-gonadal axis is sufficiently robust to accommodate various frequencies of LHRH secretion. The use of pulsatile LHRH administration is not practical for the routine induction of puberty in adolescent boys and girls with gonadotropin deficiency. Both hCG and human menopausal gonadotropin can be used as effective substitutes for recombinant human pituitary LH and FSH to produce full gonadal...
When diagnosis of hypogonadism is firmly established (e.g., girls with 45,X gonadal dysgenesis), begin hormonal substitution therapy at 1213 yr of age.

**Goal:** to approximate normal adolescent development

**Initial therapy:** ethinyl estradiol 5 µg by mouth or conjugated estrogen 0.3 mg (or less) by mouth daily for 46 mo

After 6 mo of therapy (or sooner if “breakthrough” bleeding occurs) begin cyclic therapy:

**Ethinyl estradiol:**

- **Progesterone:**
  - **Progestagen (e.g., medroxyprogesterone acetate 5 mg by mouth)** 12th to 21st day of month
  - Gradually increase dose of estrogen over next 23 yr to conjugated estrogen 0.6125 mg or ethinyl estradiol 1020 µg daily for first 21 days of month

In hypogonadotropic hypogonadism: to induce ovulation at appropriate time: pulsatile LHRH or FSH and hCG therapy

Maturation, especially in those with pituitary pathology. But, again, this regimen is cumbersome and expensive. Thus, long-term gonadal steroid replacement therapy is the treatment of choice for hypothyroidic or pituitary gonadotropin deficiency until fertility is achieved.

Hypergonadotropic hypogonadism is treated by replacement of testosterone in boys and estradiol in girls. For treatment of gonadal dysgenesis, estrogen therapy should be initiated when the patient is age 13 (bone age > 11 years) to allow secondary sexual development at an appropriate chronologic age. The Kinleffel syndrome is compatible with varying degrees of masculinization at puberty, but some patients require testosterone replacement. The concentrations of plasma testosterone and LH should be monitored every 6 months during puberty and yearly thereafter. If the LH level rises more than 2.5 SD above the mean value or the testosterone level decreases below the normal range for age, testosterone replacement therapy is indicated.

Patients receiving gonadal steroid replacement follow the same treatment regimen whether the diagnosis is hypogonadotropic hypogonadism or hypergonadotropic hypogonadism (see Table 24-30 and Table 24-31). Various testosterone preparations with several routes of administration are available. Alkylated testosterone preparations are to be avoided because of the risk of peliosis hepatis (hemorrhagic liver cysts), which is not related to dose or duration of treatment; although regression is possible with discontinuation of testosterone treatment, progression to liver failure can occur. Males may receive testosterone enanthate, propionate, or cypionate. 50 to 100 mg every 4 weeks intramuscularly at the start; later the dosage is gradually increased to 200 to 300 mg every 2 to 3 weeks.

Low-dose replacement therapy is appropriate until well into the pubertal growth spurt.

Testosterone may be administered by cutaneous patch on scrotal skin or nonsexual skin to cause secondary sexual development in hypogonadal adolescents: patches may be given at night to increase the diurnal variation of testosterone seen in early puberty. Physiologic values of serum testosterone are possible along with secondary sexual development. A teenage boy may be less likely to apply a patch daily, and biweekly or monthly injections may allow better compliance; nonetheless, we and others find that 2.5- and 5-mg dermal testosterone patches may be useful in motivated teenagers. New testosterone gel preparations, usually rubbed onto the forearms, are approved for adults and may be used in a similar manner. Testosterone ointment may be used as therapy for micropenis to enlarge the size of the phallus intentionally, but an infant coming in contact with the skin of an individual with testosterone gel (before it is absorbed into the intended subject’s skin) runs the risk of unplanned testosterone effects.

Initially, girls 12 to 13 years of age are given ethinyl estradiol, 5 µg/day orally, or conjugated estrogens, 0.3 mg/day by mouth, on the first 21 days of the month. The dose is gradually increased over the next 2 to 3 years to 10 µg of ethinylestradiol or 0.6 to 1.25 mg of conjugated estrogen for the first 21 days of the month. The maintenance dose should be the minimal amount to maintain secondary sexual characteristics, sustain withdrawal bleeding, and prevent osteoporosis. After breakthrough bleeding occurs, or no later than 6 months after the start of cyclic therapy, a progestagen (e.g., medroxyprogesterone acetate, 5 mg/day) is added on days 12 through 21 of the month. Undesirable effects are uncommon but may include weight gain, headache, nausea, peripheral edema, and mild hypertension. Application of portions of transdermal 17-estradiol patches at night was shown to mimic levels of estrogen produced in early puberty and to bring about breast development slowly; other therapeutic schedules are possible. As with testosterone, there must be care that the preparation is not placed in contact with young children or untoward estrogen effects may occur.

There is concern about the increased risk of endometrial and breast carcinoma in patients receiving chronic estrogen replacement therapy including patients with Turner’s syndrome. This is not an issue in adolescents or young children but is a consideration in older women. The use of postmenopausal agents to antagonize the effect of estrogens reduces the risk of endometrial cancer, but knowledge of the optimal dose of estrogen and progesterone to enhance development without unduly increasing the risk of cancer must come from future studies. Estrogen replacement is important for its antosteoporotic action on bone. Surprisingly, we lack controlled studies on optional sex steroid replacement regimens in adolescent women.

Patients with hypopituitarism may complain of sparse pubic hair growth or, in girls, total absence of public hair. Pubic hair thickens further in affected males with hCG treatment that adds the testicular contribution of testosterone to the exogenous testosterone therapy. GH therapy in males with GH and gonadotropin deficiency enhances the steroidogenic response of the testes to hCG administration. Further, adolescent or young adult women have been given low doses (25 mg) of long-acting intramuscular testosterone every 4 weeks to stimulate the growth of pubic hair without virilization.

The result of treatment with testosterone in boys with radiation-induced primary testicular failure is normal final height, although in a group of patients with concomitant spinal radiation, the upper/lower segment ratio was much reduced, indicating impaired spinal growth. The results of clinical trials of biosynthetic hGH therapy in Turner’s syndrome indicate that an increase in growth rate with a substantial increase in final height into the lower range of the normal growth curves is possible, especially with a dose higher than used in GH deficiency (see more detailed earlier discussion in the section on Turner’s syndrome). There is some degree of improvement of the abnormal body proportions of Turner’s syndrome with hGH treatment, but the disproportionate growth of the foot may dissuade some girls from continuing treatment to maximal benefit for height.

The addition of estrogen therapy at low doses has been reported either to exert no effect on adult height or to reduce the adult height obtained with GH therapy administered alone. Indeed, the length of time of exposure to GH before estrogen treatment is said to be the major determinant of whether GH and estrogen treatment increased final height. However, it is postulated that if GH is started early enough (e.g., 2 to 8 years of age), estrogen therapy may be added at an age (13 years) appropriate for the institution of puberty (see discussion in Turner’s syndrome section). Counseling and a peer support group are exceedingly important components of the long-term management.

The bone density is decreased in Turner’s syndrome, at least in part, because of hypogonadism at puberty, and this tendency becomes more severe with age in patients who discontinue or do not receive estrogen replacement therapy. Transdermal estrogen was shown to increase bone density in subjects with Turner’s syndrome who have finished statural growth.
DISORDERS OF PUBERTY (Continued)

Sexual Precocity

Sexual precocity (Table 24-32) is defined as the appearance of any sign of secondary sexual maturation at an age more than 2.0 SD below the mean; in the past, the ages of 8 years in girls and 9 years in boys were considered the lower limits of the normal onset of puberty. Present data detailed previously indicate that the limits in normal boys remain at 9 years but that the lower limit for white girls is 7 years and for black girls is 6 years, assuming that there is no sign or symptom of CNS disorders or other serious or chronic disease that might cause sexual precocious puberty. These guidelines are similar to those proposed by the Drug and Therapeutics and Executive Committees of the Lawson Wilkins Pediatric Endocrine Society. The new data noted in the first section of this chapter show that breast development and pubic hair development may occur in girls as young as 6 years in substantial numbers, especially in black girls, leading to a need for careful evaluation and conservatism, even in these young years, in evaluating and treating girls with only minimal, relatively non-progressive signs of sexual precocity.

If the sexual precocity results from premature reactivation of the hypothalamic LHRH pulse generator (or LHRH-independent sexual precocity), the condition is called complete isosexual precocity or true or central precocious puberty and is LHRH-dependent. Pulsatile LH release has a pubertal pattern in this form, and the rise in the concentration of LH after LHRH administration is indistinguishable from the normal pubertal pattern of serum LH. If extrapituitary secretion of gonadotropins or secretion of gonadal steroids independent of pulsatile LH release leads to virilization in boys or feminization in girls, the condition is termed incomplete isosexual precocity, pseudoprecocious puberty, or LHRH-independent sexual precocity. The production of excessive estrogens in males leads to inappropriate feminization, and the production of increased androgen levels in females leads to inappropriate virilization; these conditions are termed contrasexual precocity (also termed heterosexual precocity). Hence, the disorders that cause sexual precocity can be separated into those in which the increased secretion of gonadal steroids depends on LHRH stimulation of pituitary gonadotropins and those in which it is unrelated to activation of the hypothalamic LHRH pulse generator.

In all forms of sexual precocity, the increased gonadal steroid secretion increases height velocity, somatic development, and the rate of skeletal maturation and, because of premature epiphyseal fusion, can lead to the paradox of tall stature in childhood but short adult height. Data on the final height in true precocious puberty are scarce (see Table 24-34), but several studies of untreated females with idiopathic central precocious puberty demonstrated a mean final height of 151 to 155 cm (Table 24-34). In the boys followed to adult stature by Thamdrup the mean height was 155.4 cm ± 8.3 (SD) and all were well below midparental height and far below the fathers' height. Blood pressure matches that of height- and weight-related

<table>
<thead>
<tr>
<th>TABLE 24-32 – Classification of Sexual Precocity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>True Precocious Puberty or Complete Isosexual Precocity (LHRH-Dependent Sexual Precocity or Premature Activation of the Hypothalamic LHRH Pulse Generator)</strong></td>
</tr>
<tr>
<td>Idiopathic true precocious puberty</td>
</tr>
<tr>
<td>CNS tumors</td>
</tr>
<tr>
<td>Optic glioma associated with neurofibromatosis type 1</td>
</tr>
<tr>
<td>Hypothalamic astrocytoma</td>
</tr>
<tr>
<td>Other CNS disorders</td>
</tr>
<tr>
<td>Developmental abnormalities including hypothalamic hamartoma of the tuber cinereum</td>
</tr>
<tr>
<td>Encephalitis</td>
</tr>
<tr>
<td>Static encephalopathy</td>
</tr>
<tr>
<td>Brain abscess</td>
</tr>
<tr>
<td>Sarcoid or tubercular granuloma</td>
</tr>
<tr>
<td>Head trauma</td>
</tr>
<tr>
<td>Hydrocephalus</td>
</tr>
<tr>
<td>Arachnoid cyst</td>
</tr>
<tr>
<td>Myelomingingocele</td>
</tr>
<tr>
<td>Vascular lesion</td>
</tr>
<tr>
<td>Cranial irradiation</td>
</tr>
</tbody>
</table>

True precocious puberty after late treatment of congenital virilizing adrenal hyperplasia or other previous chronic exposure to sex steroids

Incomplete Isosexual Precocity (Hypothalamic LHRH-Independent)

**Males**

Gonadotropin-secreting tumors

hCG-secreting CNS tumors (e.g., chorioepitheliomas, germinoma, teratoma)

hCG-secreting tumors located outside the CNS (hepatoma, teratoma, chorionicarcinoma)

Increased androgen secretion by adrenal or testis

Congenital adrenal hyperplasia (CYP21 and CYP11B1 deficiencies)

Virilizing adrenal neoplasm
Leydig cell adenoma
Familial testotoxicosis (sex-limited autosomal dominant pituitary gonadotropin-independent precocious Leydig cell and germ cell maturation)
Cortisol resistance syndrome

**Females**
- Ovarian cyst
- Estrogen-secreting ovarian or adrenal neoplasm
- Peutz-Jeghers syndrome

**In Both Sexes**
- McCune-Albright syndrome
- Hypothyroidism
- iatrogenic or exogenous sexual precocity (including inadvertent exposure to estrogens in food, drugs, or cosmetics)

**Variations of Pubertal Development**
- Premature thelarche
- Premature isolated menarche
- Premature adrenarche
- Adolescent gynecomastia in boys
- Macro-orchidism

**Feminization in Males**
- Adrenal neoplasm
- Chorioepithelioma
- CYP11B1 deficiency
- Late-onset adrenal hyperplasia
- Testicular neoplasm (Peutz-Jeghers syndrome)
- Increased extraglandular conversion of circulating adrenal androgens to estrogen
- iatrogenic (exposure to estrogens)

**Virilization in Females**
- Congenital adrenal hyperplasia
- CYP21 deficiency
- CYP11B1 deficiency
- 3-HSD deficiency
- Virilizing adrenal neoplasm (Cushing's syndrome)
- Virilizing ovarian neoplasm (e.g., arhenoblastoma)
- iatrogenic (exposure to androgens)
- Cortisol resistance syndrome
- Aromatase deficiency

LHHR, luteinizing hormone-releasing factor (GnRH); CNS, central nervous system; CYP21, 21-hydroxylase; CYP11B1, 11-hydroxylase; 3-HSD, 3-hydroxysteroid dehydrogenase 4,5-isomerase.


**Long-Term Follow-up of True Precocious Puberty**

Pregnancy has occurred in patients with true or central precocious puberty as early as 5 years of age. Of course, such pregnancies are in fact the result of childhood sexual abuse of a child with true precocious puberty, a fact rarely reported by the sensational press. Fertility in later life is less well documented, but in our experience as well as that of others, normal pregnancies have occurred in women who had idiopathic true precocious puberty, a CNS abnormality triggering true precocious puberty, or premature menarche. In the isosexual precocity of the McCune-Albright syndrome, there are also reports of adult fertility. 

**Idiopathic True or Central Precocious Puberty**

By common definition, 2.5% of normal children develop signs of puberty before the lower limits of normal. The Hermann-Giddens study found that 27% of black and 7% of white girls have some manifestation of secondary sex development by 7 years of age. Thus, as stated previously, a useful definition of sexual precocity is onset before 6 years in black girls, before 7 years in white girls, and before 9 years in boys, assuming there

<table>
<thead>
<tr>
<th>Reference</th>
<th>No. of Patients (Women/Men)</th>
<th>Women (cm)</th>
<th>Men (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thamdrup</td>
<td>26/4</td>
<td>151.3 ±8.8</td>
<td>155.4 ±8.3</td>
</tr>
<tr>
<td>Sigurfonsdottir and</td>
<td>40/11</td>
<td>152.7 ± 8.0</td>
<td>156.0 ±7.3</td>
</tr>
<tr>
<td>Hayles</td>
<td>4/0</td>
<td>150.9 ±5.0</td>
<td></td>
</tr>
<tr>
<td>Werder</td>
<td>15/0</td>
<td>155.3 ± 9.6</td>
<td></td>
</tr>
<tr>
<td>UCSF</td>
<td>8/4</td>
<td>153.8 ± 6.8</td>
<td>159.6 ± 8.7</td>
</tr>
</tbody>
</table>
is no sign or symptom of CNS or other serious disease. Although this definition is a useful guideline, a significant proportion of girls 6 to 8 years of age with idiopathic precocious puberty represent one end of the bell-shaped curve for normal puberty onset, as described at the beginning of the chapter, and are examples of early normal puberty, just as those with constitutional delay in growth and adolescence are healthy but late maturers who fall in the older age segment of the normal distribution.

The nature of the striking sex difference in the prevalence of idiopathic true precocious puberty (females >> males) in contrast to constitutional delay in growth and puberty (males >> females) is poorly understood. There may be a history of early maturation in the family; rarely, true precocious puberty is transmitted as an autosomal recessive trait in boys and girls. A larger group of children, however, develop true precocious puberty with no familial tendency toward early maturation and no signs of organic disease; these children have idiopathic true precocious puberty. This condition, which may be manifest in infancy, is about nine times more common in girls than in boys (see Table 24-35) and is commonly associated with electroencephalographic abnormalities. The age at onset in girls in about 50% of cases is 6 to 7 years, in about 25% is 2 to 6 years, and in 18% is in infancy (see Fig. 24-57). Organic forms of true precocious puberty, especially if associated with hypothalamic hamartoma, have an earlier mean age of onset than the idiopathic form.

In boys (see Fig. 24-58) the testes usually enlarge under gonadotropin stimulation before any other signs of puberty are seen; in girls (see Fig. 24-57) an increase in the rate of growth, the appearance of breast development, enlargement of the labia minora, and maturational changes in the vaginal mucosa are the usual presenting signs, with variable manifestations of public hair depending on the age at onset. Progression of secondary sexual maturation is often more rapid than the normal pattern of pubertal maturation. A waxing and waning course of development may be encountered. The rapid growth is associated with the rise in estrogen synthesis and secretion and the increased GH secretion and elevation of serum IGF-I levels because of stimulation by estradiol. The ratio of bone age to chronologic age and the rise of IGF-I above normal values for age are predictive of outcome; more mildly affected children progress less rapidly and tend to maintain their target height. Patients with slowly progressive

### TABLE 24-34 – Distribution by Sex of Children with Idiopathic and Neurogenic Precocious Puberty

<table>
<thead>
<tr>
<th>Series</th>
<th>Idiopathic Female</th>
<th>Neurogenic Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thamdrup (1961)</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>Wilkins (1965)</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>Sigurjonsdottir and Hayles (1968)</td>
<td>8</td>
<td>18</td>
</tr>
<tr>
<td>University of California, San Francisco (1981)</td>
<td>13</td>
<td>26</td>
</tr>
<tr>
<td>Unpublished</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### TABLE 24-35 – Etiology of True Precocious Puberty

<table>
<thead>
<tr>
<th>Etiology</th>
<th>Number and Sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Idiopathic</td>
<td>121F, 13M</td>
</tr>
<tr>
<td>Other causes</td>
<td></td>
</tr>
<tr>
<td>CNS-hypothalamic tumors including hamartomas</td>
<td>11F, 15M</td>
</tr>
<tr>
<td>Arachnoid cyst</td>
<td>2F, 1M</td>
</tr>
<tr>
<td>Hydrocephalus</td>
<td>6F, 1M</td>
</tr>
<tr>
<td>Head trauma (child abuse)</td>
<td>1F</td>
</tr>
<tr>
<td>Perinatal asphyxia, cerebral palsy</td>
<td>3F, 1M</td>
</tr>
<tr>
<td>Encephalitis or meningitis</td>
<td>3F, 1M</td>
</tr>
<tr>
<td>Sex chromosome abnormalities (47,XXY; 48,XXXXY)</td>
<td>2M</td>
</tr>
<tr>
<td>Nonspecific seizure disorder or mental retardation</td>
<td>26F, 16M</td>
</tr>
<tr>
<td>Degenerative CNS disease</td>
<td>3M</td>
</tr>
<tr>
<td>Congenital virilizing adrenal hyperplasia with secondary true precocious puberty</td>
<td>3M</td>
</tr>
</tbody>
</table>

*Data from University of California, San Francisco, Pediatric Endocrine Clinic.


From Paul D, Conte FA, Grumbach MM, Kaplan SL. Long-term effect of gonadotropin-releasing hormone agonist therapy on final and near-final height in 26 children with true precocious puberty treated at a median age of less than 5 years. J Clin Endocrinol Metab, 80:546551.
When photographed, the patient had been treated with medroxyprogesterone acetate for 1.5 years. His height was 95.2 cm (+1 SD), the phallicus was 6 × 3 cm, and the testes were 2.4 × 1.3 cm. Basal concentrations of LH (LER-960) were 0.9 ng/mL; FSH (LER-869) 0.8 ng/mL; and testosterone 7 ng/dL. After 100 µg of LHRH, LH concentrations rose to 2.3 ng/mL, whereas FSH concentrations did not change. We have been treating with medroxyprogesterone acetate since 1982; during this time, he has had no treatment change except for the testosterone loss. His height was 150 cm at the age of 8 years 1671 and 168 cm at the age of 10 years 1672.

Figure 24-59 Left, Mean basal plasma lutinizing hormone (LH) level (LER-960) and mean peak and increment after intravenous LH-releasing hormone (LHRH) (100 µg) in normal prepubertal and pubertal females and in females with idiopathic true precocious puberty. The mean peak and increments of plasma LH are higher in true precocious puberty than in normal puberty. Right, Basal follicle-stimulating hormone (FSH) level (LER-1364) and mean peak and increment after intravenous LH (100 µg) in normal prepubertal and pubertal females with true precocious puberty. The concentration of FSH and the response to LH were greater in females with true precocious puberty and normal puberty than in prepubertal females. (From Kaplan SL, Grumbach MM. Pathogenesis of sexual precocity. In Grumbach MM, Sizonenko PC, Aubert ML [eds]. Control of the Onset of Puberty. Baltimore, Williams & Wilkins, 1990, pp 620–660. © 1990, the Williams & Wilkins Co., Baltimore.)

The pituitary gland undergoes hypertrophy in early infancy, puberty, and pregnancy and is increased in size on MRI in patients with central precocious puberty. Growth hormone, thyroid-stimulating hormone, and, rarely, prolactin and gonadotropins' concentrations are increased. In our experience, a CNS tumor was present in at least half of this group. Central Nervous System Tumors Causing True Precocious Puberty

The pituitary gland undergoes hypertrophy in early infancy, puberty, and pregnancy and is increased in size on MRI in patients with central precocious puberty. Growth hormone, thyroid-stimulating hormone, and, rarely, prolactin and gonadotropins' concentrations are increased. In our experience, a CNS tumor was present in at least half of this group. T1 images indicate a convex upper border of the pituitary gland in patients in normal or central precocious puberty, indicating the similarity in the physiologic changes of both conditions. Two sisters have been reported with the rare finding of pituitary gland hypertrophy (height greater than 1 cm) in central precocious puberty. Although empty sella may be associated with central precocious puberty, the empty sella syndrome is less frequently observed in patients with central precocious puberty than in patients with pituitary hypofunction. Although empty sella was found in 10% of children imaged for suspected hypothalamic-pituitary disorders including hypogonadotropic hypogonadism, the incidence in the general population is not known. 1673 1674

The dependent and gonadotropin steroid concentrations in plasma, the LH response to LHRH administration, and the amplitude and frequency of LH pulses are in the normal pubertal range. (Fig. 24-59 and Fig. 24-60; see Fig. 24-41.) The new third-generation gonadotropin assays may allow the diagnosis of true precocious puberty by determination in a single serum sample for LH in the basal state or 40 minutes after a single subcutaneous dose of LHRH. 1675 1676 1677 1678 1679 1680 1681 1682 1683

The notable improvement in the discriminatory power and specificity of the new commercial LH and FSH assays in the diagnosis of true precocious puberty is well illustrated by the observations of Brito and co-workers. 1684 The mean basal concentration of LH in 100 prepubertal children (60 males, 40 females) was 0.6 µIU/mL, the minimal detectable concentration. In Tanner stage 2 (breast development in girls, testes size in boys), the LH levels ranged from less than 0.6 to 2.7 µIU/mL in boys and less than 0.6 to 1.2 µIU/mL in girls. In 28 children with true precocious puberty, the mean basal LH concentration was 1.6 µIU/mL in boys, with a sensitivity of 71% and a specificity of 100%. In girls the sensitivity was 62% and the specificity 100%. In 10 children with LHRH-independent sexual precocity, the LH level was undetectable (<0.6 µIU/mL).

After the administration of an intravenous bolus of LHRH, the mean peak LH value in 16 prepubertal boys and 11 prepubertal girls was 3.1 µIU/mL, whereas in the children with true precocious puberty the mean peak LH value was greater than 21.6 µIU/mL, in contrast to LHRH-independent sexual precocity, in which the value was 1.5 µIU/mL. Peak LH levels above 9.6 µIU/mL in boys and 6.9 µIU/mL in girls were set as a pubertal response to LH stimulation. In boys and girls with true precocious puberty the combination of an elevated basal LH value and LHRH-stimulated peak value had a specificity and a positive predictive value of 100%.

Adrenarche usually does not accompany gonadarche in girls with true precocious puberty younger than 5 or 6 years. 1685 Pubic hair is sparse or absent initially in girls of this age. 1686 When the onset of true precocious puberty occurs after age 6, it is usually associated with early adrenarche for chronologic age but not for bone age. 1687

Girls with true precocious puberty have a tendency toward obesity (a mean BMI greater than that of normal girls of the same chronologic age), unrelated to treatment with LHRH agonist. 1688 1689

A small proportion of patients with central precocious puberty, including a pubertal response of LH to LHRH and increased pulsatile LH secretion at night, may either revert spontaneously to a more immature pubertal state, persist without further progression, or fluctuate between progression and regression. In some patients the course is inexorably progressive and, indeed, clinical evidence suggests a continuum in girls from premature thelarche through unsustained or slowly progressive precocious puberty to the relatively rapid progression of sexual maturation, once begun, in true precocious puberty. 1690

Central Nervous System Tumors Causing True Precocious Puberty

True precocious puberty resulting from CNS tumors (see Table 24-33 and Table 24-34) has about the same prevalence in boys and girls; however, in boys, neurologic abnormalities account for two thirds of those with true precocious puberty, and in our experience a CNS tumor was present in at least half of this group. A CNS neoplasm must be considered in the differential diagnosis for any patient with true precocious puberty. Optic and hypothalamic glioma (often associated with neurofibromatosis) and craniopharyngioma may cause true precocious puberty, either by impinging on the neural pathways that inhibit the LHRH pulse generator in childhood or as a consequence of cranial radiation for therapy of a brain tumor. The prevalence of true precocious puberty is increased after cranial radiation for local tumors or leukemia. Even radiotherapy targeting the pituitary gland may result in true precocious puberty. 1691

The unusual combination of GH deficiency and central precocious puberty can occur in children previously subjected to

### Table 24-36

<table>
<thead>
<tr>
<th>CNS Tumors Associated with Isosexual Sexual Precocity at University of California, San Francisco</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males/IPF: Isosexual precocious puberty = 13/15 = 0.9/1</td>
</tr>
</tbody>
</table>
neurons; it may be sessile or pedunculated and is usually attached to the posterior hypothalamus between the tuber cinereum and the mammillary bodies. These

<table>
<thead>
<tr>
<th>TABLE 24-37 — Clinical and Laboratory Characteristics of Children with True Precocious Puberty Caused by Hypothalamic Hamartoma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Characteristic</td>
</tr>
<tr>
<td>Age at onset of pubertal signs</td>
</tr>
<tr>
<td>Birth to 1 yr</td>
</tr>
<tr>
<td>1 to 2 yr</td>
</tr>
<tr>
<td>2 to 4 yr</td>
</tr>
<tr>
<td>7 yr</td>
</tr>
<tr>
<td>Neurologic signs</td>
</tr>
<tr>
<td>Seizures including gelastic type</td>
</tr>
<tr>
<td>Headache and visual symptoms</td>
</tr>
<tr>
<td>None</td>
</tr>
</tbody>
</table>

*Literature review.

therapeutic radiation of the CNS in association with a CNS neoplasm or, because of a variety of CNS abnormalities, including developmental malformations and head trauma. The lack of GH may not be apparent because of the increased growth resulting from the elevated gonadotropin levels. Nonetheless, GH-deficient children with central precocious puberty grow more slowly than GH-sufficient children with central precocious puberty but faster than GH-deficient children without sexual precocity. Further, in GH-deficient children with central precocious puberty the IGF-I concentrations are intermediate between the higher levels found in GH-sufficient children with sexual precocity and the lower levels found in the prepubertal GH-deficient children. 

GH deficiency and true precocious puberty can occur with CNS radiation doses of only 18 Gy, and gonadotropin deficiency, thyrotropin deficiency, and adrenocorticotropic deficiency occur with doses greater than 40 Gy, as does hypoprolactinemia. Boys and girls are both subject to an earlier onset of puberty in combination with GH deficiency with CNS irradiation of 25 to 47 Gy for tumors outside the hypothalamic-pituitary area. Treatment with a combination of GH and LHRH agonist is indicated in these patients and results in better growth and improved height compared with the use of LHRH agonist alone. Because GH secretion is related to BMI, it is important to rule out a decrease of GH secretion because of increased BMI in true precocious puberty on this basis before interpreting the decrease as evidence of GH deficiency.

Hamartomas of the tuber cinereum include malformations composed of a heterotopic mass of neurosecretory neurons, fiber bundles, and glial cells. This is frequently associated with true precocious puberty. The occurrence of seizure is uncommon when the mass diameter of the hamartoma is less than 10 mm, whereas a larger mass is associated with a high risk. 

With CT and MRI brain scans, hamartomas of the tuber cinereum are now being detected in young boys and girls previously thought to have idiopathic true precocious puberty and are now considered the most common known cause of true precocious puberty (Fig. 24-61). Before 1980 there were 37 patients in the literature with hamartomas of the tuber cinereum, and since 1980 another 80 have been reported (see Table 24-37). This increase is attributable to use of CT and MRI brain scans. For example, Pescevitz and colleagues, in reviewing the experience at the National Institutes of Health, reported that of 87 girls with true precocious puberty, 16% had a hypothalamic hamartoma, 40% had other CNS abnormalities, and 60% had idiopathic true precocious puberty. Among 20 boys with true precocious puberty, 2 had idiopathic true precocious puberty, 10 had a hypothalamic hamartoma, and 8 had other CNS abnormalities including hypothalamic neoplasms. The LHRH-secreting hypothalamic hamartoma consists of a heterotopic mass of nervous tissue containing LHRH neurosecretory neurons; it may be sessile or pedunculated and is usually attached to the posterior hypothalamus between the tuber cinereum and the mammillary bodies. These masses project into the suprasellar cistern, and the pedunculated hamartoma has a distinct stalk; hamartomas have a characteristic appearance that does not change with time. They appear on a CT or MRI scan as an isodense, abnormal fullness of the interpeduncular, preoptine, and posterior suprasellar cisterns, occasionally with distortion of the anterior third ventricle. This there is no enhancement with contrast material. MRI gives the best visualization of the lesion (see Fig. 24-62). Hamartomas of the tuber cinereum commonly cause true precocious puberty but are also associated with seizures, often gelastic (laughing) in type, developmental disorders, and behavior abnormalities. The etiology of the development of the hypothalamic hamartoma may be the converse of the lack of migration of LHRH neurons in Kallmann's syndrome because of the lack of production of adhesion molecules encoded by the KAL gene. We may postulate that in the hypothalamic

1208

**TABLE 24-37** — Clinical and Laboratory Characteristics of Children with True Precocious Puberty Caused by Hypothalamic Hamartoma

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>University of California, San Francisco (n = 12: 6M, 6F)</th>
<th>Hochman et al. (n = 27: 18M, 9F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at onset of pubertal signs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Birth to 1 yr</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>1 to 2 yr</td>
<td>4</td>
<td>17</td>
</tr>
<tr>
<td>2 to 4 yr</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>7 yr</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Neurologic signs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seizures including gelastic type</td>
<td>3/12</td>
<td>11/24</td>
</tr>
<tr>
<td>Headache and visual symptoms</td>
<td>1/12</td>
<td>5/24</td>
</tr>
<tr>
<td>None</td>
<td>7/12</td>
<td>7/24</td>
</tr>
</tbody>
</table>

*Literature review.

of Health, reported that of 87 girls with true precocious puberty, 16% had a hypothalamic hamartoma, 40% had other CNS abnormalities, and 60% had idiopathic true precocious puberty. Among 20 boys with true precocious puberty, 2 had idiopathic true precocious puberty, 10 had a hypothalamic hamartoma, and 8 had other CNS abnormalities including hypothalamic neoplasms. The LHRH-secreting hypothalamic hamartoma consists of a heterotopic mass of nervous tissue containing LHRH neurosecretory neurons; it may be sessile or pedunculated and is usually attached to the posterior hypothalamus between the tuber cinereum and the mammillary bodies. These masses project into the suprasellar cistern, and the pedunculated hamartoma has a distinct stalk; hamartomas have a characteristic appearance that does not change with time. They appear on a CT or MRI scan as an isodense, abnormal fullness of the interpeduncular, preoptine, and posterior suprasellar cisterns, occasionally with distortion of the anterior third ventricle. There is no enhancement with contrast material. MRI gives the best visualization of the lesion (see Fig. 24-62). Hamartomas of the tuber cinereum commonly cause true precocious puberty but are also associated with seizures, often gelastic (laughing) in type, developmental disorders, and behavior abnormalities. The etiology of the development of the hypothalamic hamartoma may be the converse of the lack of migration of LHRH neurons in Kallmann's syndrome because of the lack of production of adhesion molecules encoded by the KAL gene. We may postulate that in the hypothalamic

1208
Hamartomas of the tuber cinereum are not true neoplasms and do not progress or enlarge. Analysis of hamartomas associated with true precocious puberty has revealed the presence of ectopic LH-releasing neurosecretory cells similar to the LH-releasing-neurosecretory neurons in the median basal hypothalamus. This developmental abnormality exerts its endocrine effects by the elaboration and pulsatile release of LH-RH.

Indeed, LH-RH-containing fibers have been identified passing from the hamartoma toward the median eminence. We have suggested that the LH-RH-containing neurosecretory neurons in the tumor are unrestrained by the intrinsic CNS mechanism that inhibits the normal LH-RH pulse generator and act as an ectopic LH-RH pulse generator, either independently or in synchrony with the LH-RH neurosecretory neurons in the medial basal hypothalamus, to produce intermittent secretory bursts of LH-RH (see Fig. 24-38). The LH-RH is transported to the pituitary by way of the portal circulation and elicits pulsatile release of LH.

If the hamartoma secreted LH-RH in a continuous fashion, true precocious puberty should not occur as the LH-RH receptors would be desensitized. (About 10% of hypothalamic hamartomas are not associated with true precocious puberty.)

The hypothalamic hamartomas of two young girls with rapidly progressive true precocious puberty did not contain immunoreactive LH-RH neurons on biopsy, but with use of a variety of techniques, including light microscopy, immunohistochemistry, and in situ hybridization for cells containing TGF-α protein or mRNA, respectively, the mass showed a network of TGF-α-containing astroglial cells. As discussed earlier, the hypothesis is that some hypothalamic hamartomas, by virtue of the increased production of TGF-α, a member of the epidermal growth factor family, and neuregulins synthesized by hypothalamic and astroglial cells through paracrine mechanisms, effect the release of bioactive factors including prostaglandin E₂ that act on LH-RH neurons to increase LH-RH secretion. This is a novel hypothesis to explain the occurrence of true precocious puberty with hypothalamic hamartomas that apparently lack LH-RH neurons. Several unusual features of the hamartomas in both patients, however, confound this interpretation. These hamartomas were much larger than the typical hypothalamic hamartomas associated with true precocious puberty; in one patient the mass bulged into the third ventricle, and in the other the pituitary gland was hypertrophic and bulged through the diaphragm sellae.

Hamartomas of the tuber cinereum should not be approached surgically except in unusual circumstances. Although there are cases in which removal of a hypothalamic hamartoma has led to reversal of the pubertal process, deaths have been reported after attempted operative removal. Long-term follow-up of the hypothalamic hamartomas associated with true precocious puberty demonstrated lack of growth on monitoring with periodic CT or MRI scans. Accordingly, although some have advocated neurosurgical removal of these hamartomas, we do not recommend neurosurgical extirpation in the absence of strong evidence of growth of the mass or of an associated complication such as intractable seizures or hydrocephalus.

Central Nervous System Neoplasms

Sexual precocity may be the first manifestation of a hypothalamic tumor of any cell type when it arises in or impinges on the posterior hypothalamus. However, neurologic symptoms such as headaches and visual disturbances may develop, and children may have diabetes insipidus, hydrocephalus, or optic atrophy caused by an enlarging tumor.

The location of CNS tumors causing true precocious puberty makes surgical removal difficult. A conservative approach calls for biopsy of the neoplasm and radiation therapy or chemotherapy or both, depending on the pathologic findings.

Other Central Nervous System Conditions

True precocious puberty may be secondary to encephalitis, static cerebral encephalopathy, hydrocephalus, brain abscess, or sarcoid granulomas or tuberous granulomas of the hypothalamus, with or without tuberculous meningitis. Central precocious puberty can occur after severe head trauma (usually in girls), and it has been associated with the cerebral atrophy or focal encephalomalacia following cerebral edema complicating the treatment of severe diabetic ketoacidosis.

Children with nontumor hydrocephalus, even if shunted, experience earlier pubertal development, and those who have not been adequately treated may develop true precocious puberty. Delayed puberty is an alternative outcome in a minority of affected children. The growth pattern of children with severe hydrocephalus often includes poor prepubertal growth and an early pubertal growth spurt leading to decreased final height.

Arachnoid cysts arising de novo, after infection, or after surgery can cause premature sexual development, possibly with associated GH deficiency.
Neurofibromatosis type 1 (von Recklinghausen’s disease) is associated with a propensity to develop the optic chiasmal tumors that are the most common but not the only cause for development of true precocious puberty in a child with neurofibromatosis. The prevalence of optic gliomas in neurofibromatosis type 1 is about 17%, most occurring in the first decade, but only about 20% to 30% become symptomatic; these tumors rarely progress in the years after diagnosis. The tumor suppressor NF1 gene located on the long arm of chromosome 17 (q11.2), which has a high mutation rate, encodes a 327-kd protein, neurofibromin, which is widely expressed, even though neurofibromatosis type 1 involves mainly tissues derived from the neural crest. A wide variety of mutations of the gene have been reported, especially deletions and nonsense and truncating mutations distributed over the coding region of the NF1 gene. In sporadic cases, the new mutation originates in the paternally derived NF1 allele in the vast majority of instances, which suggests a role for genomic imprinting.

Neurofibromatosis type 1 is characterized by multiple pigmented areas and overgrowth of nerve sheaths and fibrous tissue elements. Multiple café-au-lait spots are frequent and are smoother in outline than those of those of the McCune-Albright syndrome. The diagnosis is made if two or more of the following are present: (1) six or more café-au-lait macules, the greatest diameter of which is more than 5 mm in prepubertal and more than 12.5 mm in postpubertal subjects; (2) two or more neurofibromas of any type or one plexiform neurofibroma; (3) freckling in the axilla or inguinal region; (4) optic glioma; (5) two or more iris Lisch nodules (ophthalmic hamartomas); (6) a distinctive osseous lesion such as sphenoid dysplasia or pseudoarthrosis; and (7) a first-degree relative with neurofibromatosis type 1 according to the preceding criteria.

Neurofibromas of the skin in neurofibromatosis may be subcutaneous sessile or deep plexiform masses in children; pedunculated lesions develop in later childhood. Internal neurofibromas cause most of the complications. Bone abnormalities such as cysts and pseudarthrosis, hemihypertrophy, bowing, scoliosis, and skull and facial defects are common (20%); dumbbell-shaped tumors of spinal nerve roots may cause pain, sensory and motor dysfunction, and bone erosions; gliomas or neurofibromas of any part of the CNS, including the optic nerves and hypothalamus, may calcify. Lisch nodules of the iris are frequent, particularly in adults. Sarcomatous degeneration occurs in 5% to 15% of patients. Other neoplasms include CNS astrocytomas often involving the visual pathways, ependymomas, meningiomas, neurofibrosarcomas, rhabdomyosarcomas, and nonlymphocytic leukemias. Neurofibromatosis may develop in affected adults.

The clinical manifestations of neurofibromatosis include seizures, visual defects, and either delayed or true precocious puberty. Mental retardation occurs more often in this population but is usually not severe; there is also an increased incidence of psychiatric disease. Most affected children have some manifestations of the disease by 1 year of age.

In addition, both medroxyprogesterone acetate and cyproterone acetate have undesirable effects in high doses. A hypothesis involving the primary role of recovery from nutritional deprivation has been challenged. In a retrospective study of true precocious puberty from Belgium, 28% (39 families in Sweden and in children referred after kwashiorkor prior to 3 years of age, 1055 screening MRI scans are recommended for early detection of CNS tumors.

Other CNS abnormalities associated with true precocious puberty but without demonstrable lesions on imaging study include epilepsy, laughing seizures, mental retardation, and the post-traumatic state. Septo-optic dysplasia (described earlier) may be associated not only with multiple pituitary hormone deficiencies and delayed puberty but also rarely with true precocious puberty. True precocious puberty may be consisting of some pituitary hormones and excessive secretion of others, including prolactin. Patients with myelomeningocele (myelodysplasia) have an increased prevalence of endocrine abnormalities, including hypothalamic hypothyroidism, hyperprolactinemia, and elevated gonadotropin concentrations, that in some patients are associated with true precocious puberty.

Neurofibromatosis type 1 (von Recklinghausen’s disease) is widely expressed, even though neurofibromatosis type 1 involves mainly tissues derived from the neural crest.

Neurofibromatosis type 1 is characterized by multiple pigmented areas and overgrowth of nerve sheaths and fibrous tissue elements. Multiple café-au-lait spots are frequent and are smoother in outline than those of the McCune-Albright syndrome. The diagnosis is made if two or more of the following are present: (1) six or more café-au-lait macules, the greatest diameter of which is more than 5 mm in prepubertal and more than 12.5 mm in postpubertal subjects; (2) two or more neurofibromas of any type or one plexiform neurofibroma; (3) freckling in the axilla or inguinal region; (4) optic glioma; (5) two or more iris Lisch nodules (ophthalmic hamartomas); (6) a distinctive osseous lesion such as sphenoid dysplasia or pseudoarthrosis; and (7) a first-degree relative with neurofibromatosis type 1 according to the preceding criteria.

Neurofibromas of the skin in neurofibromatosis may be subcutaneous sessile or deep plexiform masses in children; pedunculated lesions develop in later childhood. Internal neurofibromas cause most of the complications. Bone abnormalities such as cysts and pseudarthrosis, hemihypertrophy, bowing, scoliosis, and skull and facial defects are common (20%); dumbbell-shaped tumors of spinal nerve roots may cause pain, sensory and motor dysfunction, and bone erosions; gliomas or neurofibromas of any part of the CNS, including the optic nerves and hypothalamus, may calcify. Lisch nodules of the iris are frequent, particularly in adults. Sarcomatous degeneration occurs in 5% to 15% of patients. Other neoplasms include CNS astrocytomas often involving the visual pathways, ependymomas, meningiomas, neurofibrosarcomas, rhabdomyosarcomas, and nonlymphocytic leukemias. Neurofibromatosis may develop in affected adults.

The clinical manifestations of neurofibromatosis include seizures, visual defects, and either delayed or true precocious puberty. Mental retardation occurs more often in this population but is usually not severe; there is also an increased incidence of psychiatric disease. Most affected children have some manifestations of the disease by 1 year of age.

In addition, both medroxyprogesterone acetate and cyproterone acetate have undesirable effects in high doses. A hypothesis involving the primary role of recovery from nutritional deprivation has been challenged. In a retrospective study of true precocious puberty from Belgium, 28% (39 girls, 1 boy) were foreign children who had emigrated from developing countries. A toxicologic screen found a greatly elevated mean concentration of the p,p’-DDE), which raised the possibility of a role of organochlorine pesticide dichlorodiphenyltrichloroethane (DDT) derivative. The etiology is not established but may be related to the effects of undernutrition or its rapid repair during a sensitive time in development. In Sweden, the adopted Indian children had pubertal growth spurts similar to those of Swedish children, but the loss of height in childhood and the early puberty after virilizing disorders.

Sarcomatous degeneration occurs in 5% to 15% of patients. Other neoplasms include CNS astrocytomas often involving the visual pathways, ependymomas, meningiomas, neurofibrosarcomas, rhabdomyosarcomas, and nonlymphocytic leukemias. Neurofibromatosis may develop in affected adults. 

True Precocious Puberty in Children Adopted from Developing Countries

An increased prevalence of true precocious puberty occurred in children (with established birth dates) from developing countries adopted after 3 years of age into families in Sweden and in children referred after kwashiorkor prior to 3 years of age, as well as in The Netherlands, France, and Italy. A hypothesis involving the primary role of recovery from nutritional deprivation has been challenged. In a retrospective study of true precocious puberty from Belgium, 28% (39 girls, 1 boy) were foreign children who had emigrated from developing countries. A toxicologic screen found a greatly elevated mean concentration of the organochlorine pesticide dichlorodiphenyltrichloroethane (DDT) derivative. This phenomenon, secondary true precocious puberty, occurred in boys and girls with congenital virilizing adrenal hyperplasia in whom glucocorticoid replacement therapy was begun after age 4 to 8 and who had an advanced bone age. True precocious puberty has also been documented in children who received or were exposure to androgens or estrogens for long periods during early childhood for a variety of conditions.

Management of True Precocious Puberty

The objectives in the management and therapy of true precocious puberty are summarized in Table 24-38, which addresses the major psychosocial and clinical goals. Three principal agents have been used in the treatment of true precocious puberty whether idiopathic or neurologic: medroxyprogesterone acetate, cyproterone acetate, and superactive LHRH agonists.

Medroxyprogesterone and cyproterone reversed or arrested the progression of secondary sexual characteristics but had no apparent effect or a small effect on final height, especially in affected girls. This failure may be due in part to the disproportionate action of the small amount of circulating estradiol on skeletal growth relative to its effect on secondary sexual characteristics. In any event, in none of the early studies with medroxyprogesterone or cyproterone was the concentration of plasma estradiol in girls and of testosterone in boys systematically monitored and the dosage of these agents adjusted accordingly. More encouraging effects of long-term treatment with medroxyprogesterone acetate on final height were reported in the latest study. In addition, both medroxyprogesterone acetate and cyproterone acetate have undesirable effects in high doses.

### TABLE 24-38 — Objectives for the Management and Treatment of True Precocious Puberty

<table>
<thead>
<tr>
<th>Objective</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detection and treatment of an expanding intracranial lesion</td>
<td></td>
</tr>
<tr>
<td>Arrest of premature sexual maturation until the normal age at onset puberty</td>
<td></td>
</tr>
<tr>
<td>Regression of secondary sexual characteristics already present</td>
<td></td>
</tr>
<tr>
<td>Attainment of normal mature height; suppression of the rapid rate of skeletal maturation</td>
<td></td>
</tr>
<tr>
<td>Prevention of emotional disorders and handicaps and alleviation of parental anxiety; promotion of understanding by counseling, early sex education, and acceleration of social age</td>
<td></td>
</tr>
<tr>
<td>Reduction of risk of sexual abuse and early sexual debut</td>
<td></td>
</tr>
<tr>
<td>Prevention of pregnancy in girls</td>
<td></td>
</tr>
<tr>
<td>Preservation of future fertility</td>
<td></td>
</tr>
<tr>
<td>Diminish the increased risk of breast cancer associated with early menarche</td>
<td></td>
</tr>
</tbody>
</table>

The dosage of medroxyprogesterone acetate is 5 to 20 mg twice a day orally or 100 to 200 mg/m² surface area intramuscularly every 1 or 2 weeks. We prefer the oral route.[2138] This agent inhibits gonadotropin secretion by its action on the hypothalamic LHRH pulse generator/pituitary gonadotropin unit and has a direct suppressive effect on gonadal steroidogenesis through type II 3-hydroxysteroid dehydrogenase/isomerase. Medroxyprogesterone acetate has glucocorticoid action and can suppress ACTH and cortisol secretion, increase appetite and lead to excessive weight gain, and induce hypertension and a cushingoid facies and appearance.[2138]

Cypreterone acetate has antiandrogenic, antigonadotropic, and gestational properties. It has been used outside the United States for the treatment of true precocious puberty.[195] Its advantages and disadvantages are similar to those of medroxyprogesterone acetate. The usual oral dose is 70 to 100 mg/m² surface area daily, given in two divided doses; the intramuscular dose is 100 to 200 mg/m² every 14 to 28 days. Cypreterone acetate suppresses the secretion of ACTH and the plasma concentration of cortisol. Fatigue and weakness are common side effects, probably as a consequence of secondary adrenal insufficiency. This agent lacks glucocorticoid activity and does not appear to produce cushingoid features.

The long-term effect of either of these agents is not known. For the treatment of true precocious puberty, medroxyprogesterone and cypreterone acetate have been replaced by the much more effective LHRH agonists; however, these are useful backup agents for the occasional patients who experience untoward effects of LHRH agonist therapy.

**Supraventricular Luteinizing Hormone-Releasing Hormone Agonists**

The LHRH agonists, synthetic analogues of the amino acid sequence of the natural LHRH decapetide, are the treatment of choice for true precocious puberty, whether idiopathic or not.[1213] (Table 24-39 and Table 24-40).

**TABLE 24-39 – Action of Luteinizing Hormone-Releasing Hormone Agonists in True Precocious Puberty**

<table>
<thead>
<tr>
<th>Chronic administration induces desensitization of the pituitary gonadotrope to the action of endogenous LHRH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inhibition of pulsatile secretion of LH and FSH</td>
</tr>
<tr>
<td>Inhibition of gonadotropin secretion results in a striking decrease in gonadal steroid output by testes or ovaries and reduction in gonadal size</td>
</tr>
</tbody>
</table>

Paradoxically, these pharmacologic agents, when administered chronically, suppress pulsatile LH and FSH release, gonadotroidal output, and gametogenesis. In contrast to the effects of pulsatile administration at physiologic doses and frequency, continuous administration of natural LHRH suppresses gonadotropin secretion[726] after an initial, brief stimulation of gonadotropin release. Suppression results from binding of the agonist to the LHRH receptor on gonadotrophs and subsequent desensitization of the gonadotroph to LHRH. Initially, down-regulation and loss of receptors occur. When receptor levels return to normal, desensitization persists as a result of uncoupling of the receptors from the intracellular signaling effectors pathway.[711] Administration of a potent LHRH agonist subcutaneously once a day, although it initially stimulates gonadotropin release, induces desensitization of the gonadotrophs to LHRH within a few days. In children with true precocious puberty, this regimen blocks the effect of endogenous LHRH and functions as a selective, highly specific pharmacologic clamp on the secretion of gonadotropins without interfering directly with release of the other pituitary hormones. In essence, the regimen produces a reversible medical gonadectomy (see Table 24-39). The superactive agonist analogues of LHRH have about 15 to 200 times the potency of the natural LHRH decapetide, prolonged action, and low toxicity (see Table 24-40).

Replacing the glycine-amide terminus of LHRH with alkyl amines as in (Pro9-ethylamide(NEt))LHRH, substituting certain -amino acids at position 6 as in (o-Trp6)LHRH, and making bulky hydrophobic alterations at position 6 as in (-Trp3)6)LHRH, increase the potency and duration of action.

The suppressive effects of the agonists on gonadotropin secretion make them useful in the treatment of true precocious puberty. Various agonists are available, some for subcutaneous administration, some for intranasal treatment, and some for injection intramuscularly in depot formulations (see Table 24-40). A comparison of intranasal treatment with buserelin to monthly intramuscular treatment with triptorelin demonstrated the superiority of the latter in men and the plaater final height, although intranasal preparations given in adequate amounts are able to improve final height in affected children.[1718] At present, the depot formulation of leuprolein (leuprolide acetate) is the only depot preparation that may be given every 4 weeks that is approved by the FDA; long-term studies have established its efficacy and safety.[1765] The bioavailability of agonists given intranasally is much reduced, as reflected in the need to use a high dose at more frequent intervals. The effectiveness of LHRH agonist in the treatment of true precocious puberty varies with the potency of the analogue, dose, route of administration, and compliance.[1766] The addition of GH treatment to the LHRH regimen is a consideration when growth velocity is reduced sufficiently over a 6-month period to compromise predicted final height.[1767]

In both the idiopathic and organic forms of true precocious puberty, treatment with a potent LHRH agonist initially results in 1 to 3 days of increased FSH and LH release and a rise in circulating gonadotroidal levels; chronic therapy suppresses the pulsatile secretion of LH and FSH and blocks the pulsatile LH response to the administration of native LHRH after 7 to 14 days of treatment (Fig 24-65 and Fig 24-66). There is, as well, a change in the isofoms of gonadotropins toward a more basic charge, suggesting a differential effect of the LHRH agonist on pituitary secretion of gonadotropins. Within 2 to 4 weeks in girls and 6 weeks in boys, gonadal steroid secretion is reduced to prepubertal levels and maintained in the prepubertal state by chronic treatment (Fig 24-67; see Fig 24-65 and see Fig 24-66). A plasma estradiol concentration less than 18 pmol/L (5 pg/mL) in girls and a plasma testosterone level less than 0.7 nmol/L (20 ng/dL) in boys indicate adequate supression of gonadal steroidogenesis through type II 3-hydroxysteroid dehydrogenase/isomerase. Careful monitoring of serum gonadotropins and sex steroid is necessary for evaluation of the effectiveness of LHRH agonist treatment, and the same laboratory should be used for these studies to allow consistency of interpretation of the results.

Changes in secondary sexual characteristics occur within the first 6 months of therapy (see Fig 24-67). In girls, these effects include reduction in breast size and decrease in pubic hair, cessation of menses if present before treatment, and decreased size of the uterus and ovaries as assessed by pelvic ultrasonography.

**TABLE 24-40 – Luteinizing Hormone-Releasing Hormone Agonists: Pharmacologic Treatment of True Precocious Puberty**

<table>
<thead>
<tr>
<th>Structure of Natural LHRH and Substitutions in LHRH Analogue</th>
<th>Relative Potency</th>
<th>Formula</th>
<th>Dosage Form</th>
<th>Dose</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glu-His-Pro-Ser-Trp-Gly-Leu-Arg-Pro-Gly-NH</td>
<td>1</td>
<td>LHRH</td>
<td>Subcutaneous</td>
<td>48 µg/kg/day</td>
<td>1717</td>
</tr>
<tr>
<td>1 2 3 4 5 6 7 8 9 10</td>
<td>Deslorelin</td>
<td>-Trp</td>
<td>-NET</td>
<td>150</td>
<td>(o-Trp6)6NEt) LHRH</td>
</tr>
</tbody>
</table>

**Medroxyprogesterone Acetate and Cyproterone Acetate**

This agent inhibits gonadotropin secretion by its action on the hypothalamic LHRH pulse generator/pituitary gonadotropin unit and has a direct suppressive effect on gonadal steroidogenesis through type II 3-hydroxysteroid dehydrogenase/isomerase. Medroxyprogesterone acetate has glucocorticoid action and can suppress ACTH and cortisol secretion, increase appetite and lead to excessive weight gain, and induce hypertension and a cushingoid facies and appearance.
sonography. A small proportion of girls have recurrent episodes of hot flushes and moodiness. In boys, pubic hair thickens, the testes decrease in size, acne and seborrhea regress, genitile erections and masturbation become much less frequent, the high energy level and aggressive behavior diminish, and self-esteem improves.

Height velocity, expressed in centimeters per year or as standard deviations above the mean height velocity for chronological and bone age, decreases about 60% during the first year of therapy. Skeletal maturation slows dramatically during the first 3 years, to a rate often less than the progression in chronological age. From the second year on, height velocity for bone age is usually appropriate. At present, the growth data for compliant patients strongly suggest improved adult height potential as a result of therapy in young children (6 years of age) with true precocious puberty, especially when treatment is begun soon after the onset of precocity and when the bone age is advanced only a few years. Even with the limitations of estimates of predicted height taken into account, effective therapy improves final height predictions or maintains the normal target height in young children.

The accumulating data on final height in children treated with LHRH agonists support this contention. A study of 26 children treated with Leuprolide demonstrated a striking benefit in the children treated before 5 years of age (girls, height 164.3 ± 7.7 cm) compared with those treated after age 5 years (157.6 ± 6.6 cm) and with untreated patients (152.7 ± 8.6 cm). We recommend the treatment of all affected children in whom puberty began before 6 years of age to ensure an optimal prognosis for adult height. European studies confirmed that final height is improved more with therapy starting before 6 years of age than after 6 years of age: the age that is optimal for discontinuation of therapy is still open to question as the post-treatment growth spurt is important in determining adult height. The remaining challenge is the rapid diagnosis of central precocious puberty and early initiation of GnRH therapy if final height is to be preserved.

Children with true precocious puberty have higher mean concentrations of circulating IGF-I for chronological age, consistent with the increased secretion of gonadal steroids. Treatment with LHRH agonists reduces the level of IGF-I to the normal range for bone age but not for chronological age. This indicates that gonadal steroids increase plasma IGF-I concentrations in true precocious puberty as well as in normal puberty. Secretion of GH is increased in true precocious puberty to a level comparable to that in normal puberty.
compliance. However, irregular or inadequate treatment or poor compliance results in persistent or intermittent increases in the concentrations of plasma gonadal steroids. Regular assessment is essential, initially at intervals of 1 to 3 months. Such assessment involves periodic determinations of plasma testosterone levels in boys and estradiol levels in girls; the change in basal concentrations of LH and FSH in third-generation assays or the LH and FSH response to exogenous LHRI; measurement of growth, bone age, and secondary sexual characteristics; and in girls serial evaluations of ovarian morphology and uterine size by pelvic sonography. Although the urinary excretion of LH correlates with the stage of pubertal development in normal individuals, urinary gonadotropin determinations are not sufficiently sensitive to be used for monitoring purposes.

**Figure 24-67** A 2 5/12-year-old girl with true precocious puberty after 6 weeks of desloralin therapy (4 µg/day subcutaneously). Note the regression in the size of the breasts; however, the rapid rate of growth had not decreased. At the end of 1 year of therapy, growth rate was suppressed to 4 cm/year, and bone age advanced only 1 year. CA, chronologic age; HT, height; WT, weight; BA, bone age. (From Syne DM, Grumbach MM. Puberty in the male and female: its physiology and disorders. In Yen SCC, Jaffe RB [eds]. Reproductive Endocrinology, 2nd ed. Philadelphia, WB Saunders, 1986; pp 313384.)

**Figure 24-68** Effect of luteinizing hormone-releasing hormone (LHRH) agonist therapy in true precocious puberty on growth. Left, Changes in mean height velocity (cm/year ± 1 SE) after the initiation of LHRH agonist therapy with DTrp6 Pro2 Net (LHRH) (filled bars) or with nafarelin (hatched bars). A sharp decrease in height velocity occurred within 1 year. Right, Mean (±1 SE) height for bone age before and during LHRH agonist treatment. The discrepancy between height and the more advanced bone age decreases (reverts to normal) with chronic LHRH agonist treatment. (From Kaplan SL, Grumbach MM. True precocious puberty: treatment with GnRH agonists. In Delemarre-Van de Waal H, Plant TM, van Rees GP, et al [eds]. Control of the Onset of PubertyIII. Amsterdam, Elsevier, 1989; pp 357373.)

Regularly scheduled visits also provide the opportunity for continued counseling. Increasing numbers of patients are now completing treatment after years of therapy, and the return of gonadal function has been assessed. In 46 girls treated for at least 2 years who completed therapy at a mean age of 11 years, menarche occurred at a mean age of 12.1 years, which represents an average of 1.2 years after discontinuing therapy. Ovulation occurred in 50% of girls 1 year after menarche and in 90% of those studied 2 or more years after menarche. Spontaneous menstrual cycles occurred at a mean of 18 months in another follow-up study after LHRH agonist therapy ceased with a range of 0 to 60 months. This pattern is quite similar to that of normal pubertal maturation. There is no delay in the development of the hypothalamo-pituitary axis in girls with true precocious puberty treated with LHRH agonist therapy. Reports on boys with true precocious puberty after LHRH therapy was discontinued confirm the reversible nature of the therapy. Gonadotropins in the basal or LHRH-stimulated state return to normal pubertal values by 1 year after cessation of therapy. However, testicular size may take longer to reach normal values.

The criteria for treatment of patients with true precocious puberty are listed in Table 24-42. Therapy is not indicated if a pubertal pattern of pulsatile LH secretion during sleep is not present or if the LH response to exogenous LHRH is prepubertal. Before beginning treatment, it is essential to establish the rapidly progressive nature of the sexual precocity. Girls without a reduced height potential do not require LHRH agonist therapy to ensure an appropriate final height outcome; these girls tend to have lower serum IGF-I and estradiol concentrations and may have fewer signs of estrogenization. The most severely affected girls are the ones who respond best to LHRH agonist therapy. In a subset of girls, the tempo is relatively slow and the sexual precocity may not be sustained. The growth rate slows to normal for age, skeletal maturation progresses in accordance with chronologic age, and there is little to no risk of impairment of final height. In some girls, we have observed within a 1- to 2-month period the return of a pubertal pattern of LH pulsatility during sleep, of a pubertal LH response to LHRH, and of the concentration of plasma estradiol to a pubertal state; unlike the typical patients, such girls do not exhibit the initial hyperresponse of plasma estradiol and LH to the LHRH agonist. Many girls in this subset have clinical and hormonal features that fall between those of premature thelarche and true precocious puberty and are typical of neither condition, so-called exaggerated thelarche.

**Figure 24-69** A 2 5/12-year-old girl with true precocious puberty after 6 weeks of desloralin therapy (4 µg/day subcutaneously). Note the regression in the size of the breasts; however, the rapid rate of growth had not decreased. At the end of 1 year of therapy, growth rate was suppressed to 4 cm/year, and bone age advanced only 1 year. CA, chronologic age; HT, height; WT, weight; BA, bone age. (From Syne DM, Grumbach MM. Puberty in the male and female: its physiology and disorders. In Yen SCC, Jaffe RB [eds]. Reproductive Endocrinology, 2nd ed. Philadelphia, WB Saunders, 1986; pp 313384.)

Regularly scheduled visits also provide the opportunity for continued counseling. Increasing numbers of patients are now completing treatment after years of therapy, and the return of gonadal function has been assessed. In 46 girls treated for at least 2 years who completed therapy at a mean age of 11 years, menarche occurred at a mean age of 12.1 years, which represents an average of 1.2 years after discontinuing therapy. Ovulation occurred in 50% of girls 1 year after menarche and in 90% of those studied 2 or more years after menarche. Spontaneous menstrual cycles occurred at a mean of 18 months in another follow-up study after LHRH agonist therapy ceased with a range of 0 to 60 months. This pattern is quite similar to that of normal pubertal maturation. There is no delay in the development of the hypothalamo-pituitary axis in girls with true precocious puberty treated with LHRH agonist therapy. Reports on boys with true precocious puberty after LHRH therapy was discontinued confirm the reversible nature of the therapy. Gonadotropins in the basal or LHRH-stimulated state return to normal pubertal values by 1 year after cessation of therapy. However, testicular size may take longer to reach normal values.

The criteria for treatment of patients with true precocious puberty are listed in Table 24-42. Therapy is not indicated if a pubertal pattern of pulsatile LH secretion during sleep is not present or if the LH response to exogenous LHRH is prepubertal. Before beginning treatment, it is essential to establish the rapidly progressive nature of the sexual precocity. Girls without a reduced height potential do not require LHRH agonist therapy to ensure an appropriate final height outcome; these girls tend to have lower serum IGF-I and estradiol concentrations and may have fewer signs of estrogenization. The most severely affected girls are the ones who respond best to LHRH agonist therapy. In a subset of girls, the tempo is relatively slow and the sexual precocity may not be sustained. The growth rate slows to normal for age, skeletal maturation progresses in accordance with chronologic age, and there is little to no risk of impairment of final height. In some girls, we have observed within a 1- to 2-month period the return of a pubertal pattern of LH pulsatility during sleep, of a pubertal LH response to LHRH, and of the concentration of plasma estradiol to a pubertal state; unlike the typical patients, such girls do not exhibit the initial hyperresponse of plasma estradiol and LH to the LHRH agonist. Many girls in this subset have clinical and hormonal features that fall between those of premature thelarche and true precocious puberty and are typical of neither condition, so-called exaggerated thelarche.

About 10% of girls with apparently classical premature thelarche progress to definite true precocious puberty with no signs at the time of first presentation to differentiate them from girls who continued with the pattern of premature thelarche; in the majority of this situation, the onset of breast development is noted after 2

**TABLE 24-41** Comparison of Current Height (Final or Near Final) and Height Gain of Luteinizing Hormone-Releasing Hormone Agonist-Treated Patients

<table>
<thead>
<tr>
<th>Unretrated</th>
<th>No. of Patients</th>
<th>Female</th>
<th>Male</th>
<th>Mean Ht Gain (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>116</td>
<td>152.7 ± 8.6</td>
<td>155.6 ± 7.7</td>
<td></td>
</tr>
<tr>
<td>&lt;5 yr</td>
<td>41</td>
<td>150.2 ± 7.6</td>
<td>153.3 ± 7.1</td>
<td></td>
</tr>
<tr>
<td>&gt;5 yr</td>
<td>75</td>
<td>153.4 ± 8.4</td>
<td>161.3 ± 6.0</td>
<td></td>
</tr>
<tr>
<td>LHRH-treated</td>
<td>28</td>
<td>160.5 ± 6.6</td>
<td>166.3 ± 12.2</td>
<td></td>
</tr>
<tr>
<td>UCSF</td>
<td>11</td>
<td>164.3 ± 7.7</td>
<td>172.1</td>
<td></td>
</tr>
<tr>
<td>&lt;5 yr</td>
<td>15</td>
<td>157.6 ± 6.6</td>
<td>163.3 ± 13.0</td>
<td></td>
</tr>
<tr>
<td>&gt;5 yr</td>
<td>15</td>
<td>157.6 ± 5.9</td>
<td>168.8 ± 8.3</td>
<td></td>
</tr>
<tr>
<td>Oertel</td>
<td>8</td>
<td>151.2 ± 5.9</td>
<td>154.4</td>
<td></td>
</tr>
<tr>
<td>Kaufl</td>
<td>26</td>
<td>151.2 ± 5.9</td>
<td>154.4</td>
<td></td>
</tr>
</tbody>
</table>


| a | Final predicted height initial predicted height (Bayley-Pinneau method). |
| b | Final height. |
| c | Final or nearly final height. |
| d | CA at start of therapy. |
years of age. Psychosocial factors and parental anxiety that adversely affect the well-being of the child need to be assessed in the decision to initiate LHRH agonist treatment.

**Adverse Effects**

Unwanted reactions to LHRH agonists in the treatment of true precocious puberty have so far been minimal but include local and systemic allergic reactions in a few patients, including asthmatic episodes when the agent is given intranasally. However, the prevalence of a sterile abscess at the site of intramuscular injection of long-acting repository preparations, including leuprolein and triptorelin, is clearly increased (5% to 10%), unpredictable, and intermittent and in most instances is related to the polyactic and polyglycolic polymer and not to the LHRH agonist itself. Switching to daily subcutaneous injections or to intranasal administration of nondepot preparations is rarely associated with a recurrence.

When treatment is discontinued, even after 8 years, the gonadal suppression is reversed within a few weeks to months with a rise in the concentration of plasma gonadal steroids, progression of sexual maturation, and return of menses. A small increase in serum prolactin above normal limits, but not galactorrhea, has been described in girls following treatment with LHRH agonist. Areal BMD, but not volumetric bone, is increased in children with central precocious puberty. Although a decrease in bone density has been reported in LHRH agonist-treated patients, later studies indicated that volumetric BMD and peak bone mass are normal during and after discontinuation of LHRH therapy. Despite these encouraging results, one must be alert to the possible emergence of unforeseen long-term side effects. Four patients developed slipped capital femoral epiphyses during or just after treatment of central precocious puberty with LHRH agonist.

LHRH agonists are effective in both boys and girls with idiopathic true precocious puberty, the androgen-induced form of secondary true precocious puberty after therapy for virilizing congenital adrenal hyperplasia with glucocorticoids, and organic or neurogenic forms of true precocious puberty associated with hamartomas of the tuber cinereum, hypothalamic neoplasms, and other CNS lesions. Although there are reports in the literature of surgical removal of hamartomas of the tuber cinereum, the ease of medical treatment of the sexual precocity associated with this congenital malformation, the finding that the mass does not enlarge on MRI or CT brain scans, and the risks of an adverse outcome of surgical intervention in this region of the CNS, the use of LHRH agonists in idiopathic true precocious puberty has increased in recent years. Psychosocial factors and parental anxiety, including evidence that the child's psychosocial well-being is adversely affected, are important factors in the decision to initiate treatment with LHRH agonists.

**TABLE 24-42 -- Indications for Therapy with Luteinizing Hormone-Releasing Hormone Agonists in True or Central Precocious Puberty**

<table>
<thead>
<tr>
<th>Indication</th>
<th>LHRH Agonists</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rapid advancement over a period of 6 mo to 1 yr of secondary sex characteristics, height, height velocity, and bone age (increased &gt;2.5 SD for chronologic age) in affected boys and girls</td>
<td></td>
</tr>
<tr>
<td>A plasma testosterone concentration sustained &gt;2.5 nmol/L (&gt;75 ng/dL) in boys younger than 8 yr of age determined by sensitive, specific immunoassay</td>
<td></td>
</tr>
<tr>
<td>A plasma estradiol, recurrently 36 pmol/L (10 pg/mL) determined by a sensitive, specific assay capable of quantifying low concentrations of estradiol</td>
<td></td>
</tr>
<tr>
<td>Onset of menarche (and recurrent menses) in girls younger than 9 yr of age</td>
<td></td>
</tr>
<tr>
<td>Psychosocial factors and parental anxiety, including evidence that the child's psychosocial well-being is adversely affected</td>
<td></td>
</tr>
</tbody>
</table>

In children with neurogenic or organic true precocious puberty, especially those with associated GH deficiency, the course is almost invariably progressive and LHRH treatment should not be delayed.

**TABLE 24-43 -- Potential Use of Aromatase Inhibitors or Estrogen Receptor Antagonists to Restrain Skeletal Maturation in Disorders of Growth and Sexual Maturation**

<table>
<thead>
<tr>
<th>Disorder</th>
<th>LHRH Agonists</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth disorders or variants of normal growth (to restrain epiphyseal maturation)</td>
<td></td>
</tr>
<tr>
<td>Isolated growth hormone deficiency</td>
<td></td>
</tr>
<tr>
<td>Genetic short stature/constitutional delay in growth</td>
<td></td>
</tr>
<tr>
<td>Sexual precocity</td>
<td></td>
</tr>
<tr>
<td>Congenital virilizing adrenal hyperplasia in male and female</td>
<td></td>
</tr>
<tr>
<td>To reduce dose of glucocorticoid</td>
<td></td>
</tr>
<tr>
<td>To inhibit conversion of C19 steroids to estrogens (or estrogen action)</td>
<td></td>
</tr>
<tr>
<td>With/without use of C17/20 yase inhibitor or anti-androgen</td>
<td></td>
</tr>
<tr>
<td>Testotoxicosis</td>
<td></td>
</tr>
<tr>
<td>To inhibit conversion of C19 steroids to estrogens</td>
<td></td>
</tr>
<tr>
<td>McCune-Albright syndrome</td>
<td></td>
</tr>
<tr>
<td>To inhibit conversion of C19 steroids to estrogens (or estrogen action)</td>
<td></td>
</tr>
<tr>
<td>Adolescent gynecomastia</td>
<td></td>
</tr>
<tr>
<td>To inhibit estrogen synthesis (or estrogen action)</td>
<td></td>
</tr>
</tbody>
</table>


**Psychosocial Aspects**

Psychological management is an important aspect of the care of children with true precocious puberty. With the advanced physical maturation for chronologic age, they tend to seek friends closer to their size, strength, and physical development. Difficulties may arise because they lack the social skills of older children. Sex education of the child and the family is essential and must be given in a skilful, sensitive, and explicit manner; the risks of sexual abuse in both sexes and of pregnancy in girls need to be discussed. The parents need to be informed about the management of menses. The onset of sexual activity may be earlier than average but generally remains within the normal range.

It is imperative to provide support in handling the increased height, advanced sexual maturation, and effects of gonadal steroids on behavior, activity, and emotional
In this group of disorders, the secretion of testosterone in boys and of estrogen in girls is independent of the hypothalamic LHRH pulse generator (see Table 24-32). Affected individuals do not exhibit a pubertal-type LH response to administration of LHRH or a pubertal pattern of pulsatile LH secretion, nor do they respond to chronic administration of an LHRH agonist with suppression of gonadal steroid output. Incomplete isosexual precocity or precocious pseudopuberty is a consequence of both gonadal and adrenal steroid secretion independent of LHRH, of iatrogenic exposure to gonadal steroids, or, in boys, of rare hCG- or LH-secreting tumors.

Boys

Choricontin Gonadotropin-Secreting Tumors

Several types of germ cell tumors can secrete a glycoprotein hormone that has the bioactivity of LH or hCG and can cross-react in some polyclonal LH radioimmunoassay systems. Studies using highly specific antisera to the subunit of hCG, however, confirm that the gonadotropin is hCG. Boys with hCG-secreting neoplasms may have slightly enlarged testes (although not usually to a size consonant with the size of the phallus and other male secondary sex characteristics) and may be difficult to differentiate from boys with accelerated puberty on the basis of physical examination alone. However, plasma hCG levels are elevated without an increase in the concentration of FSH or LH. Hepatomas and hepatoblastomas are among the most serious of these tumors and cause firm, irregular nodular or smooth hepatic enlargement (Fig. 24-69). The hCG has been localized to the multicellular tumor giant cells; in one case, -fetoprotein was found in the embryonal-type tumor cells spread throughout the hepatoblastoma. The average survival is only 10.7 months after diagnosis; the mean age at onset is 2 years, 8 months.

Some teratomas, chorioepitheliomas, or mixed germ cell tumors in the hypothalamic region (or in the mediastinum, the lungs, or the retroperitoneum) and certain hypothalamic pineal tumors (usually a germ cell tumor or mixed germ cell tumor, less commonly a chorioneplithelioma or its variants) cause sexual precocity in boys by secreting hCG rather than by activating the hypothalamic LHRH pulse generator and the pituitary gonadotropin gonadotropin axis. The prevalence of hCG-secreting embryonal neoplasms, especially of the mediastinum, is increased in boys with 47,XXY or mosaic Klinefelter's syndrome. About 20% of mediastinal germ cell tumors occur in boys with Klinefelter's syndrome, a prevalence 30 to 50 times that in unaffected boys. Plasma -fetoprotein is a useful additional marker for yolk sac (endodermal sinus) or mixed germ cell tumors; the cells in the tumor that secrete -fetoprotein appear to differ from those that secrete hCG. Intracranial germ cell tumors are 2.6 times more common in males than females; in females, germ cell tumors do not cause gonadotropin-induced isosexual precocity because of the paucity of effects of hCG in prepubertal females. However, they can cause true precocious puberty through disinhibition of the hypothalamic LHRH pulse generator by local effects; rarely, the germ cells contain sufficient aromatase activity to convert circulating C19 precursors (of adrenal origin after adrenarche) to estradiol, which in some instances is sufficient to induce breast development. In mixed germ cell tumors, on the other hand, hCG is commonly present in the blood as well as in cerebrospinal fluid.

"True" pure CNS germ cell tumors (germinomas) secrete insufficient hCG to be readily detectable in the circulation, but in some patients hCG can be detected in the cerebrospinal fluid. In mixed germ cell tumors, on the other hand, hCG is commonly present in the blood as well as in cerebrospinal fluid.

In children, germ cell tumors in the suprasellar-hypothalamic region do not exhibit a sex predominance and are generally associated with pituitary hormone deficiencies including diabetes insipidus. Germ cell tumors that secrete hCG are rarely located in the thalamus and basal ganglia. Intracranial germ cell tumors account for 3% to 11% of malignant CNS tumors in children and adolescents, with a predominance in the Far East. Germ cell tumors of the hypothalumus or pineal region constitute less than 1% of primary CNS tumors in Western countries but account for 4.5% of such tumors in Japan. Mixed germ cell tumors and especially pure germinomas are radiosensitive, and regression of sexual precocity may occur if the bone age is less than 11 years, only to progress later into normal puberty. Long-term survival was reported to be 88% after appropriate therapy. Calcification of the pineal is found in 8% to 11% of 8- to 11-year-old children and by itself is not indicative of a tumor. Gonadotropin-secreting pituitary adenomas are exceedingly rare in children. An LH- and prolactin-secreting pituitary adenoma caused sexual precocity in two boys. The concentration of serum LH was strikingly elevated (900 IU/L) and did not rise further after the administration of LHRH. The elevated serum testosterone (7 mmol/L, 200 ng/dL) and prolactin (215 μg/L) levels and the high concentration of LH fell to prepupal values after removal of a "chromophobe" adenoma with suprasellar extension.

Precocious Androgen Secretion Caused by Congenital Adrenal Hyperplasia, Virilizing Adrenal Tumor, or Leydig Cell Tumor

Virilizing congenital adrenal hyperplasia caused by a defect in 21-hydroxylase (CYP21, cytochrome P450 21α deficiency) leads to elevated androgen concentrations and masculinization and is a common cause of LH- and hCG-independent sexual precocity in boys (see Chapter 22). Approximately 75% of patients with P450 21α deficiency have salt loss resulting from impaired aldosterone secretion and have low serum sodium and high serum potassium concentrations. Increased plasma concentrations of 17-hydroxyprogesterone, increased levels of urinary 17-ketosteroids and pregnanetriol, and advanced bone age and rapid growth are characteristic. Treatment with glucocorticoids suppresses the abnormal androgen secretion and virilization; treatment with mineralocorticoids, when necessary, corrects the electrolyte imbalance. A rarer form of virilizing adrenal hyperplasia is usually accompanied by hypertension and is caused by 11-hydroxylase deficiency; the plasma hCG levels are elevated without an increase in the concentration of FSH or LH. Calcium and phosphorus and ketogenic hyperplasia are among the most serious of these tumors and cause firm, irregular nodular or smooth hepatic enlargement (Fig. 24-69). The hCG has been localized to the multicellular tumor giant cells; in one case, -fetoprotein was found in the embryonal-type tumor cells spread throughout the hepatoblastoma. The average survival is only 10.7 months after diagnosis; the mean age at onset is 2 years, 8 months.

Some teratomas, chorioepitheliomas, or mixed germ cell tumors in the hypothalamic region (or in the mediastinum, the lungs, or the retroperitoneum) and certain hypothalamic pineal tumors (usually a germ cell tumor or mixed germ cell tumor, less commonly a chorioneplithelioma or its variants) cause sexual precocity in boys by secreting hCG rather than by activating the hypothalamic LHRH pulse generator and the pituitary gonadotropin gonadotropin axis. The prevalence of hCG-secreting embryonal neoplasms, especially of the mediastinum, is increased in boys with 47,XXY or mosaic Klinefelter's syndrome. About 20% of mediastinal germ cell tumors occur in boys with Klinefelter's syndrome, a prevalence 30 to 50 times that in unaffected boys. Plasma -fetoprotein is a useful additional marker for yolk sac (endodermal sinus) or mixed germ cell tumors; the cells in the tumor that secrete -fetoprotein appear to differ from those that secrete hCG. Intracranial germ cell tumors are 2.6 times more common in males than females; in females, germ cell tumors do not cause gonadotropin-induced isosexual precocity because of the paucity of effects of hCG in prepubertal females. However, they can cause true precocious puberty through disinhibition of the hypothalamic LHRH pulse generator by local effects; rarely, the germ cells contain sufficient aromatase activity to convert circulating C19 precursors (of adrenal origin after adrenarche) to estradiol, which in some instances is sufficient to induce breast development. In mixed germ cell tumors, on the other hand, hCG is commonly present in the blood as well as in cerebrospinal fluid.

"True" pure CNS germ cell tumors (germinomas) secrete insufficient hCG to be readily detectable in the circulation, but in some patients hCG can be detected in the cerebrospinal fluid. In mixed germ cell tumors, on the other hand, hCG is commonly present in the blood as well as in cerebrospinal fluid.

In children, germ cell tumors in the suprasellar-hypothalamic region do not exhibit a sex predominance and are generally associated with pituitary hormone deficiencies including diabetes insipidus. Germ cell tumors that secrete hCG are rarely located in the thalamus and basal ganglia. Intracranial germ cell tumors account for 3% to 11% of malignant CNS tumors in children and adolescents, with a predominance in the Far East. Germ cell tumors of the hypothalumus or pineal region constitute less than 1% of primary CNS tumors in Western countries but account for 4.5% of such tumors in Japan. Mixed germ cell tumors and especially pure germinomas are radiosensitive, and regression of sexual precocity may occur if the bone age is less than 11 years, only to progress later into normal puberty. Long-term survival was reported to be 88% after appropriate therapy. Calcification of the pineal is found in 8% to 11% of 8- to 11-year-old children and by itself is not indicative of a tumor. Gonadotropin-secreting pituitary adenomas are exceedingly rare in children. An LH- and prolactin-secreting pituitary adenoma caused sexual precocity in two boys. The concentration of serum LH was strikingly elevated (900 IU/L) and did not rise further after the administration of LHRH. The elevated serum testosterone (7 mmol/L, 200 ng/dL) and prolactin (215 μg/L) levels and the high concentration of LH fell to prepupal values after removal of a "chromophobe" adenoma with suprasellar extension.

Precocious Androgen Secretion Caused by Congenital Adrenal Hyperplasia, Virilizing Adrenal Tumor, or Leydig Cell Tumor

Virilizing congenital adrenal hyperplasia caused by a defect in 21-hydroxylase (CYP21, cytochrome P450 21α deficiency) leads to elevated androgen concentrations and masculinization and is a common cause of LH- and hCG-independent sexual precocity in boys (see Chapter 22). Approximately 75% of patients with P450 21α deficiency have salt loss resulting from impaired aldosterone secretion and have low serum sodium and high serum potassium concentrations. Increased plasma concentrations of 17-hydroxyprogesterone, increased levels of urinary 17-ketosteroids and pregnanetriol, and advanced bone age and rapid growth are characteristic. Treatment with glucocorticoids suppresses the abnormal androgen secretion and virilization; treatment with mineralocorticoids, when necessary, corrects the electrolyte imbalance. A rarer form of virilizing adrenal hyperplasia is usually accompanied by hypertension and is caused by 11-hydroxylase deficiency; the progressive virilization ceases and the blood pressure falls to normal with glucocorticoid therapy.

All forms of congenital adrenal hyperplasia are inherited as autosomal recessive traits. Virilizing congenital adrenal hyperplasia, if untreated, can cause anovulatory amenorrhea in females and oligospermia in males; with treatment, the infertility is usually corrected (see Chapter 22). Treatment of virilizing congenital adrenal hyperplasia may unmask LH- and hCG-dependent sexual precocity (secondary true precocious puberty) as a consequence of the advanced somatic and presumably hypothalamic maturation because of exposure to androgen before glucocorticoid therapy is initiated.
canceroma may cause isosexual precocity and growth failure in boys. Rarely, an adrenal adenoma may produce both testosterone and aldosterone, leading to sexual precocity and hypertension with hypokalemia.  

Adrenal rests, or heterotopic adrenal tissue in the testes, may enlarge with endogenous ACTH stimulation in boys with untreated or inadequately treated virilizing congenital adrenal hyperplasia and may mimic bilateral or unilateral interstitial cell tumors. Rarely, the adrenal rests may lead to massive enlargement of the testes (see Chapter 22). MRI sonography including Doppler flow studies of the testes is useful in defining the extent and nature of the testicular masses. In boys in whom the testicular tumors are unresponsive to glucocorticoid therapy or improved compliance, surgical management including enucleation of the tumor has been useful in preventing further damage to the testes and improving the potential for fertility. LH receptors have been detected on adrenal or cortical cells. Although some of the testicular masses may become autonomous, it seems possible that LH stimulation is a factor in some patients.

Infrequently, a Leydig cell tumor in boys is the cause of sexual precocity; unilateral enlargement (often nodular) of the testis usually occurs in this neoplasm (although 5% to 10% are bilateral), in contrast to the usually normal size of both testes for chronologic age in boys with congenital adrenal hyperplasia or a virilizing adrenal tumor. Of interest, an LH receptoractivating mutation was detected in three boys with a sporadic Leydig cell adenoma (see later).  

Women with a previous history of congenital adrenal hyperplasia may exhibit ovarian hyperandrogenism associated with persistent elevation of LH in spite of successful treatment of their initial virilizing condition in childhood: this is not usually the case in women who have late-onset congenital adrenal hyperplasia.  

A unique form of sexual precocity in males is pituitary gonadotropin-independent familial premature Leydig cell and germ cell maturation, or testotoxicosis.  

Although it has been recognized as an LHRH-independent form of male isosexual precocity only since 1981, this disorder was described more than 50 years ago. Indeed, Andrew Shenker brought to our attention a report by Stone in 1852. Affected boys have secondary sexual development with penile enlargement and bilateral enlargement of testes to the early or midpubertal range, although the testes are often smaller than expected in relation to penile growth and pubertal maturation (Fig. 24-70). The tests show premature Leydig and Sertoli cell maturation and spermatogenesis; in some instances, Leydig cell hyperplasia is present. The rate of linear growth is rapid, skeletal maturation is advanced, and muscular development is prominent. Serum hormone determinations reveal prepubeal basal and LHRH-stimulated gonadotropin concentrations and lack of a pubertal pattern of LH pulsatility, with LH being stimulated by immunologic or biologic techniques (Table 24-44). In affected boys, plasma testosterone values are in the normal pubertal or adult range with normal clearance of testosterone. The onset of adrenarche and its biochemical marker, serum DHEAS, correlates with bone age rather than chronologic age.

Treatment with an LHRH agonist does not suppress the testicular function or maturation. When most untreated affected individuals reach late childhood or early adolescence, fertility is achieved and an adult pattern of LH secretion and response to LHRH is demonstrable; secondary LHRH-dependent true precocious puberty is superimposed on the substrate of testotoxicosis. In some adults, impaired spermatogenic function is associated with elevated concentrations of plasma FSH. This disorder, although it occurs sporadically, quite likely as a consequence of a germ line mutation or even a postzygotic one, is inherited as a sex-linked autosomal dominant trait and probably accounts for the earlier descriptions of "true" precocious puberty in families in which only males were affected. A kindred with nine generations of affected males has been reported; obligatory female carriers of the trait were unaffected as constitutional activation of the LH receptor on the ovary causes no ill effects.

In 1993, Shenker and Kremers and their colleagues independently described heterozygous activating mutations of the heterotrimeric G protein-coupled LH/CG receptor that in concert transduce the LH/CGR signal to the main effector, adenylyl cyclase (Fig. 24-71). The receptors for pituitary and placental gonadotropins and TSH belong to a subfamily of the seven-transmembrane-spanning, G protein-coupled receptors. The LH receptor, first cloned from the rat  and pig  and later the human, is a glycoprotein of 80 to 90 kDa. It is encoded by a gene localized to chromosome 2q21 (the same as the FSH receptor) that spans at least 70 kb and contains 11 exons separated by 10 introns. The large glycosylated amino-terminal extracellular hormone-binding domain of the 701-amino-acid LH/CG receptor is encoded by exons 1 to 10. A single exon, the large exon 11, encodes the entire G-linked transmembrane domain with its 7-helical segments connected by alternating extracellular and intracellular loops, the intracellular domain, and the 3 untranslated region. Most two-thirds of the receptor  and (see Fig. 24-71).

Thirteen constitutively activating heterozygous missense mutations (in over 60 reported patients) all residing within exon 11 have been reported (see Fig. 24-71): six involve the transmembrane helix VI, two involve the flanking third cytoplasmic loop, there is one each in helix V and helix II, and less commonly there are mutations in the first transmembrane helix . Thus, nine mutations are between amino acid residues 542 and 561, suggesting a hot spot. There appears to be a limited repertoire of mutations in American boys, consistent with a founder effect; European pedigrees are more diverse. A model of the transmembrane domain of the receptor provides novel suggestions concerning the structural and functional effects of these activating mutations. Transfected cultured cells with these mutations exhibited increased basal cAMP production in the absence of agonist, observations consistent with a constitutive activating mutation. The conformational changes in the LH receptor that lead to its constitutive activation have yet to be established, but various possibilities have been considered. The LHRH-dependent puberty usually ensues in adolescence. Infertility related to testicular damage can occur in adult men. Inactivating mutations of the LH receptor and their clinical consequences are discussed in Chapter 22.

In one Polish family, the disorder, a mutation of M298T in the second transmembrane domain of the LH receptor, led to sexual precocity in one boy but not in the mother, who carried the same mutation, or in her father or his son, the male uncle, suggesting the involvement of epigenetic factors. Three boys with sexual precocity related to a sporadic Leydig cell adenoma had an Asp 578 His mutation in the tumor (Fig. 24-72) (Fig. Not Available). The remarkable association of testotoxicosis and pseudohermaphroditism type Ia due to a mutation in the subunit of Gs is considered in the following. Boys with LHRH-independent, pituitary gonadotropin-independent maturation of the testes do not respond to chronic administration of an LHRH.
agonist with suppression of testosterone secretion, in contrast to the characteristic response in patients with true precocious puberty. However, testosterone secretion, height velocity and rate of bone maturation, and aggressive and hyperactive behavior have been decreased by treatment with oral medroxyprogesterone acetate.

Two other therapies have been used (Table 24-45). Ketoconazole, an orally active substituted imidazole derivative, suppresses gonadal and adrenal biosynthesis at several steps. At the dosage used in testotoxicosis (200 mg every 8 to 12 hours orally), ketoconazole mainly inhibits the enzyme cytochrome P450, which regulates both 17β-hydroxylation and the scission (17,20-lyase) of hydroxypregnenolone to dehydroepiandrosterone (see adrenarche in this chapter). However, even at the recommended dose, the agent produces a mild transient decrease in cortisol secretion and interferes with binding of testosterone to TeBG. Secondary true precocious puberty often occurs when the bone age advances to or has already reached the pubertal range (usually > 11.5 years), at which time addition of an LHRH agonist is appropriate. Ketoconazole can cause hepatic injury, which is usually mild and reversible, but hepatotoxicity is rarely severe. Further, reversible renal injury, rash, and interstitial pneumonia have been reported in a patient who tolerated lower doses, suggesting a dose-response effect.

Another therapeutic approach has been the use of the antiandrogen (and antiminalarcolcorticoid) spironolactone combined with testolactone, an inhibitor of cytochrome P450 aromatase.

![Figure 24-71](https://example.com/figure24-71.png) A, The serpentine seven transmembrane Gq protein coupled LH/hCG receptor with its large extracellular domain and the intracellular domain. The seven helical transmembrane domains are indicated by Roman numerals. E, The two-dimensional seven-transmembrane topology of the LH/hCG receptor with positions of constitutively activating mutations causing testotoxicosis (male-limit ed autosomal dominant sexual precocity). The mutations are indicated by solid circles and the residue number. Note the cluster of mutations in the VI transmembrane helix and third cytoplasmic loop. The aspartine 578 glycine mutation is the most common. (Redrawn from Yano K, Kohn LD, Saij M, et al. A case of male limited precocious puberty caused by a point mutation in the second transmembrane domain of the luteinizing hormone chorionadotropin receptor gene. Biochim Biophys Acta 1996; 220:10381042.)

(CYP19), the key enzyme in the conversion of androgens to estrogens. Because these boys often experience secondary central precocious puberty after control with spironolactone and testolactone, the addition of an LHRH agonist is a useful step to suppress pituitary gonadotropin secretion.

Girls

Incomplete isosexual precocity in girls (see Table 24-32) is caused by conditions in which estrogen is secreted autonomously by an ovarian cyst or tumor, by an adrenal neoplasm, or because of inadvertent exposure to estrogen. In a pure hCG-secreting tumor in girls, signs of isosexual precocity are absent. Girls harboring a teratoma or teratocarcinoma (or a cyst germ cell tumor) that secretes hCG have had sexual precocity caused by concurrent estrogen secretion by the tumor; these girls may also have galactorrhea, especially if chorionic somatomammotropin (human chorionic somatomammotropin, human placental lactogen) is also secreted.

**Autonomous Ovarian Follicular Cysts**

The most common estrogen-secreting ovarian mass and ovarian cause of sexual precocity is the follicular cyst. Antral follicles up to about 8 mm in diameter are common in the ovaries of normal prepubertal girls and may be seen in third-trimester fetuses and newborn infants. They may appear and regress spontaneously. Large follicular cysts may be discovered because of the presence of an abdominal mass or abdominal pain, especially after torsion or as an unexpected finding on pelvic sonography for other reasons. Occasionally, the antral follicles secrete estrogen and may enlarge to form large masses, or the follicular cysts may recur and cause recurrent signs of sexual precocity and acyclic vaginal bleeding.

Enlarged antral follicles or cysts occur in premature thelarche, true precocious puberty, and transient or incomplete sexual precocity. With some ovarian follicular cysts, the transient or recurrent sexual precocity is LH/FSH-independent. The concentration of estradiol fluctuates, usually correlating with changes in the size of the follicular cyst when monitored by pelvic sonography, and may increase to levels found in a granulosa cell tumor. The concentration of LH is suppressed, a pubertal pattern of pulsatile LH secretion is absent, and the LH rise induced by LH/FSH is prepubertal.

It is curious that a constitutive activating mutation of the FSH receptor is undescribed in a female, especially because a heterozygous mutation, Asp 567 Gly, has been detected in the third intracellular loop of the FSH receptor in a hypophysectomized man who, despite the gonadotropin deficiency, was fertile and had normal-sized testes. A constitutive activating mutation of the FSH receptor seems worthy of study. The McCune-Albright syndrome needs to be considered in any girl with recurrent ovarian cysts, even at the recommended dose, the agent produces a mild transient decrease in cortisol secretion and interferes with binding of testosterone to TeBG. Secondary true precocious puberty often occurs when the bone age advances to or has already reached the pubertal range (usually > 11.5 years), at which time addition of an LHRH agonist is appropriate. Ketoconazole can cause hepatic injury, which is usually mild and reversible, but hepatotoxicity is rarely severe. Further, reversible renal injury, rash, and interstitial pneumonia have been reported in a patient who tolerated lower doses, suggesting a dose-response effect.

Another therapeutic approach has been the use of the antiandrogen (and antiminalarcolcorticoid) spironolactone combined with testolactone, an inhibitor of cytochrome P450 aromatase.

**Table 24-45** Pharmacologic Therapy for Sexual Precocity

<table>
<thead>
<tr>
<th>Disorder</th>
<th>Treatment</th>
<th>Action and Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>LH/FSH dependent</td>
<td>LHRH agonists</td>
<td>Desensitization of gonadotropes; blocks action of endogenous LHRH</td>
</tr>
<tr>
<td>True or central precocious puberty</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LH/FSH independent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incomplete sexual precocity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Girls</td>
<td>Medroxyprogesterone acetate</td>
<td>Inhibition of ovarian steroidogenesis; regression of cyst (inhibition of FSH release)</td>
</tr>
</tbody>
</table>

**Figure 24-72** (Figure Not Available) Mutations in the luteinizing hormone (LH) receptor protein. Schematic structure of the LH receptor protein and localization of the inactivating (open squares) and activating (filled circles) mutations currently known in the human LH receptor. The short lines across the amino acid chain separate the 11 exons. (From Themmen APN, Kifor A, Faiman C. Mutations of gonadotropins and gonadotropin receptors. Endocr Rev 2000;21(5):551583.)

An unusual syndrome of estradol-secreting ovarian cysts in preterm infants born before 30 weeks of gestation is associated with edema of the labia majora and, in some instances, of the lower abdominal wall. In four preterm neonates the syndrome appeared weeks after birth and 1 to 4 weeks before the putative date of a full-term gestation. The follicular cysts, which may be unilateral or bilateral, were detected by abdominal
and pelvic sonography. The LH and FSH response to LHRH suggested that the cysts were LHRH-dependent. Treatment with medroxyprogesterone acetate was associated with regression of the cysts. LHRH agonists are useful in the treatment of ovarian follicular cysts associated with true precocious puberty (LHRH-dependent) but not so-called autonomous cysts. However, girls with autonomously functioning ovarian follicular cysts, whether recurrent or an isolated episode, often respond to treatment with oral medroxyprogesterone-one acetate but not to LHRH agonists. Medroxyprogesterone acetate also seems to prevent recurrence and accelerate involution of the follicular cysts and reduce the risk of torsion. The use of one of the new, potent estrogen inhibitors such as letrozole to reduce estradiol secretion is another potential approach to treatment. Surgical intervention is rarely indicated; a large or persistent cyst can be reduced by puncture at laparoscopy. The size of the cyst can be monitored readily by pelvic sonography.

Plasma estradiol concentrations in girls with recurrent cysts (>7 cm) may increase to high levels indistinguishable from those in granulosa cell tumors of the ovary, but they do not have increased plasma granulosa cell tumor markers such as AMH and inhibin. Alternatively, the levels of estrogen in blood and urine may be in the early pubertal range. A characteristic feature in girls with recurrent cysts is waxing and waning of estrogen levels that correlate with changes in the appearance of the ovary on pelvic sonography. Pelvic sonography is useful for visualization of ovarian cysts and estimation of functional activity. Sonograms of the ovary facilitate diagnosis. After surgical removal, measurements of plasma estradiol and inhibin levels are a useful screen for metastases. If the patient is younger than 9 years an elevated estradiol and at any age an abnormal rise in concentration of plasma AMH or inhibin suggests recurrence or metastasis.

Occasionally, gonadoblastomas in streak gonads, rare lipid tumors, cystadenomas, and ovarian carcinomas secrete estrogens, androgens, or both hormones. Even with successful resection of a gonadal sex steroidsecreting neoplasm, the child is at risk for secondary central precocious puberty developing in the future. Gonadal tumors composed of a mixture of germ cells and sex cord stromal cells that are distinct from gonadoblastoma are usually benign when discovered in female infants or children with 46,XX karyotypes, although neoplastic transformation is a risk. Two cases of metastasizing malignant mixed germ cell sex cordstromal tumors have been described in prepubertal girls with isosexual precocity. Some of these neoplasms secrete -fetoprotein and other tumor markers. Ovarian tumors are rare in the prepubertal period, accounting for about 1% of all tumors in girls younger than 17 years, and most are benign. The majority of ovarian tumors arise from germ cells or sex cord stromal cells in childhood and less than 20% are of epithelial origin, whereas in adults the majority of tumors are of epithelial origin. Early diagnosis of most childhood tumors of the ovary, unlike ovarian cancer in women, allows successful cure.

McCune-Albright syndrome  Medroxyprogesterone acetate: Inhibition of ovarian steroidogenesis; regression of cyst (inhibition of FSH release)

<table>
<thead>
<tr>
<th>Boys</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Family testotoxicosis</td>
<td>Ketoconazole. Inhibition of P-450-c17 (CYP17) (mainly 17,20-lyase activity)</td>
</tr>
<tr>
<td>Familial testotoxicosis</td>
<td>Spironolactone: or flutamide and letrozole Antiandrogen</td>
</tr>
<tr>
<td>Familial testotoxicosis</td>
<td>Medroxyprogesterone acetate: Inhibition of testicular steroidogenesis</td>
</tr>
</tbody>
</table>

LHRH, luteinizing hormone-releasing hormone.


*If true precocious puberty develops, an LHRH agonist can be added.

---

Granulosa Cell Tumor of the Ovary

This tumor is rare in childhood, and theca cell tumors are even less common. Juvenile granulosa cell tumors have distinctive features that differentiate them from the tumors in adults. Characteristic histologic features include nodular architecture, follicle formation, abundant interstitial and intrafollicular acid mucopolysaccharidich fluid, irregular microcysts, individual cell necrosis, and high mitotic activity (mean activity, 11 mitotic figures per 10 high-power fields). Size can vary from 2.5 to 25 cm with a mean diameter of 12 cm. The interstitial mucinous fluid consists predominantly of hyaluronic acid. The prognosis is good as only about 3% of patients die of the disease. Approximately 80% of granulosa cell tumors can be palpated on bimanual examination. Less than 5% are bilateral or clinically malignant. The concentration of plasma estradiol may increase to high levels; FSH and LH concentrations are usually suppressed. The tumor secretes AMH and inhibin, which are sensitive tumor markers. Sonograms of the ovary facilitate diagnosis. After surgical removal, measurements of plasma estradiol and AMH levels are a useful screen for metastases. If the patient is younger than 9 years an elevated estradiol and at any age an abnormal rise in concentration of plasma AMH or inhibit suggests recurrence or metastasis.

This autosomal dominant syndrome of mucocutaneous pigmentation of the lips, buccal mucosa, fingers, and toes; gastrointestinal hamartomatous polyposis; and a predisposition to malignancy is associated with a rare, distinctive sex cord tumor with annular tubules in both boys and girls. Estrogen secretion by the tumor may lead to feminization and incomplete sexual precocity in boys as well as girls. Less frequently, an epithelial tumor of the ovary, dysgerminoma, or a feminizing Sertoli-Leydig cell tumor has been found in patients with Peutz-Jeghers syndrome. Children with this disorder should be examined at regular intervals for the presence of gonadal tumors by pelvic sonography. The syndrome is due to mutations in the gene on 9p13.3 encoding a serine/threonine protein kinase STK11 leading to haplosufficiency of this novel tumor-suppressing gene.

Sex cord-stromal tumors derive from the coelomic epithelium or mesenchymal cells of the embryonic gonads and are composed of granulose, theca, Leydig, and Sertoli cells. Estrogen secretion from these tumors can cause pseudoprecocious puberty, and androgen secretion can cause virilization. Both inhibin A and B activin are produced as well as antimüllerian factor; all serve as useful tumor markers. Sex cord-stromal tumors not associated with Peutz-Jeghers syndrome are malignant in 25% of patients; these tumors may grow quite large, whereas those associated with Peutz-Jeghers syndrome are often small and multiple, and contain

---

1223

1224

Peutz-Jeghers Syndrome

This autosomal dominant syndrome of mucocutaneous pigmentation of the lips, buccal mucosa, fingers, and toes; gastrointestinal hamartomatous polyposis; and a predisposition to malignancy is associated with a rare, distinctive sex cord tumor with annular tubules in both boys and girls. Estrogen secretion by the tumor may lead to feminization and incomplete sexual precocity in boys as well as girls. Less frequently, an epithelial tumor of the ovary, dysgerminoma, or a feminizing Sertoli-Leydig cell tumor has been found in patients with Peutz-Jeghers syndrome. Children with this disorder should be examined at regular intervals for the presence of gonadal tumors by pelvic sonography. The syndrome is due to mutations in the gene on 9p13.3 encoding a serine/threonine protein kinase STK11 leading to haplosufficiency of this novel tumor-suppressing gene.

Sex cord-stromal tumors derive from the coelomic epithelium or mesenchymal cells of the embryonic gonads and are composed of granulose, theca, Leydig, and Sertoli cells. Estrogen secretion from these tumors can cause pseudoprecocious puberty, and androgen secretion can cause virilization. Both inhibin A and B activin are produced as well as antimüllerian factor; all serve as useful tumor markers. Sex cord-stromal tumors not associated with Peutz-Jeghers syndrome are malignant in 25% of patients; these tumors may grow quite large, whereas those associated with Peutz-Jeghers syndrome are often small and multiple, and contain
Adrenocortical tumors are rare in childhood (reported to be 0.6% of all childhood tumors and 0.3% of all malignant childhood tumors), but most produce steroid hormones in children whereas those in adults usually do not. The median age of diagnosis is 4 years, with various studies stating mean ages of 2, 4.3, and 5 years. Forty-one percent appear before 2 years and 71% before 5 years of age. Most cause virilization or Cushing's syndrome, but adrenal tumors may produce estrogen as well and androgens and cause sexual precocity in a girl or gynecomastia in a boy. One adrenal adenoma found in a 7-year-old girl expressed the gene for aromatase, demonstrating that the tumor could directly produce estrogen. There was substantial production of estrogen leading to a serum estradiol concentration of 145 pg/mL, a value in the range that is found in adrenal carcinomas as well as adrenomas.

McCune-Albright Syndrome

This sporadic syndrome, which occurs about twice as often in girls as in boys, is due to somatic activating mutations in the gene (GNAS1) encoding the subunit of the trimeric guanosine triphosphate (GTP)-binding protein (Gαs) that stimulates adenyl cyclase. It is characterized by the triad of irregularly edged hyperpigmented macules (café-au-lait spots); a slowly progressive bone disorder, polyostotic fibrous dysplasia, that can involve any bone and is frequently associated with facial asymmetry and hyperostosis of the base of the skull; and, more commonly in girls, LHRH-independent sexual precocity. Autonomous hyperfunction most commonly involves the ovary, but other endocrine involvement includes thyroid (nodular hyperplasia with thyrotoxicosis or, remarkably, with euthyroid status).

![Image](https://example.com/image.png)

**Figure 24-74** A 7 4/12-year-old girl with luteinizing hormone-releasing hormone (LHRH)-independent sexual precocity associated with McCune-Albright syndrome. She had breast development since infancy, and it increased noticeably at about 5 years of age; 6 months later episodes of recurrent vaginal bleeding began. Growth of pubic hair was noted at about 4 to 5 years of age. At age 5 1/2 years the bone age was 6 1/2 years; height was +1 SD above the mean value for age. By 6 1/2 years of age, when she was seen at the University of California, San Francisco, the bone age had advanced to 9 years, and height was +1 SD. Breasts were at Tanner stage 4; pubic hair at stage 3. Extensive irregular café-au-lait macules cover the right side of the face, left lower abdomen and thigh, and both buttocks. A bone survey showed widespread involvement of the long bones with typical polyostotic fibrous dysplasia, and the floor of the anterior fossa of the skull was sclerotic and the diploic space widened. She has had two pathologic fractures through bone cysts in the right upper femur. Note the osseous deformities. Plasma estradiol concentrations were consistently in the pubertal range; LH response to LHRH was prepubertal. Results of thyroid function studies were normal, including the thyrotropin response to thyrotropin-releasing hormone administration and antithyroid antibodies were not detected. Treatment with oral medroxyprogesterone acetate suppressed menses and arrested pubertal development but did not slow skeletal maturation. Her final height is 142 cm (+2.5 SD). Menstrual cycles are regular.

Adrenal (multiple hyperplastic nodules with Cushing's syndrome), pituitary (adenoma or mammosomatotroph hyperplasia with gigantism and acromegaly and hyperprolactinemia), and parathyroids (adenoma or hyperplasia with hyperparathyroidism). In addition, hypophosphatemic vitamin D-resistant rickets or osteomalacia can occur in this syndrome because of either overproduction of a phosphaturic factor, phosphatonin, or secreted by the bone lesions or an

<table>
<thead>
<tr>
<th>Manifestation</th>
<th>Patients (%) (n = 158)</th>
<th>Male (n = 53)</th>
<th>Female (n = 105)</th>
<th>Age at Diagnosis (yr) (range)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fibrous dysplasia</td>
<td>97 (51)</td>
<td>49</td>
<td>86</td>
<td>7.7 (0.052)</td>
<td>Polyostotic more common than monostotic</td>
</tr>
<tr>
<td>Cafe-au-lait lesion</td>
<td>85 (49)</td>
<td>43</td>
<td>86</td>
<td>7.7 (0.052)</td>
<td>Variable size and number of lesions, irregular border (“coast of Maine”)</td>
</tr>
<tr>
<td>Sexual precocity</td>
<td>52 (36)</td>
<td>25</td>
<td>77</td>
<td>4.9 (0.39)</td>
<td>Common initial manifestation</td>
</tr>
<tr>
<td>Acromegaly/gigantism</td>
<td>27 (16)</td>
<td>17</td>
<td>20</td>
<td>14.8 (0.242)</td>
<td>17/26 with adenoma on MRI/CT</td>
</tr>
<tr>
<td>Hyperprolactinemia</td>
<td>15 (9)</td>
<td>9</td>
<td>6</td>
<td>16.0 (0.242)</td>
<td>23/42 of acromegalic with PRL</td>
</tr>
<tr>
<td>Hyperthyroidism</td>
<td>19 (12)</td>
<td>7</td>
<td>23</td>
<td>14.4 (0.537)</td>
<td>Euthyroid goiter is common</td>
</tr>
<tr>
<td>Hypercortisolism</td>
<td>5 (4)</td>
<td>3</td>
<td>5</td>
<td>4.4 (0.217)</td>
<td>All primary adrenal</td>
</tr>
<tr>
<td>Myxomas</td>
<td>5 (3)</td>
<td>3</td>
<td>5</td>
<td>34 (1750)</td>
<td>Extremity myxomas</td>
</tr>
<tr>
<td>Osteosarcoma</td>
<td>2 (1)</td>
<td>1</td>
<td>2</td>
<td>36 (3437)</td>
<td>At site of fibrous dysplasia, not related to prior radiation therapy</td>
</tr>
<tr>
<td>Rickets/osteomalacia</td>
<td>3 (1)</td>
<td>1</td>
<td>3</td>
<td>27.3 (852)</td>
<td>Responsive to phosphorus plus calcirol</td>
</tr>
<tr>
<td>Cardiac abnormalities</td>
<td>11 (8)</td>
<td>8</td>
<td>9</td>
<td>1.9 (0.166)</td>
<td>Arrhythmias and CHF reported</td>
</tr>
<tr>
<td>Hepatic abnormalities</td>
<td>10 (6)</td>
<td>6</td>
<td>10</td>
<td>1.9 (0.34)</td>
<td>Neonatal icterus is most common</td>
</tr>
</tbody>
</table>

MR Imaging. Magnetic resonance imaging; CT, computed tomography; PRL, prolactin; CHF, congestive heart failure.


*Excludes all clinical and biochemical data; other rarely described manifestations include metabolic acidosis, nephrocalcinosis, developmental delay, thymic and splenic hyperplasia, and colonic polyps.

**TABLE 24-46** -- Clinical Manifestations of McCune-Albright Syndrome in 158 Reported Patients

The skin manifestations may not be conspicuous in infancy, although the majority of patients have pigmented skin lesions in infancy that usually increase in size along with body growth. The irregular-bordered cafe-au-lait macules usually do not cross the midline, are often located on the same side as the main bone lesions, and have a segmented distribution.
manifested bone abnormalities by 8 years of age. Increased serum GH levels have an adverse effect on the skull deformities.

The sexual precocity, the onset of which is often in the first 2 years of life and is frequently heralded by menstral bleeding, is due to an autonomously functioning luteinized follicular cyst of the ovary (see Table 24-47). The ovaries contain multiple follicular cysts but not corpora lutea and commonly exhibit asymmetrical enlargement as a result of a large solitary cyst that characteristically enlarges and spontaneously regresses only to recur. Serum estradiol is elevated (at times to extraordinarily high levels); in contrast, the LH response to LHRH is prepubertal and the pubertal pattern of nighttime LH pulses is absent at the onset and during the initial years. Later in the course of the sexual precocity, when the bone age approaches 12 years, the LHRH pulse generator becomes operative and ovulatory cycles ensue. Thus, an affected girl may progress from LHRH-independent puberty to LHRH-dependent puberty (see Table 24-47). LHRH agonists are not effective for treatment in the LHRH-independent stage. Testolactone (40 mg/kg per day orally), a relatively weak aromatase inhibitor, has been of equivocal useffulness and some patients become resistant to the drug. The new, highly potent, specific, third-generation aromatase inhibitors, for example, letrozole, should be more effective. Antiestrogens offer another method of control of the estrogenic effects of the disorder. A single case of treatment with tamoxifen showed decreases in the age advancement, growth rate, menses, and pubertal development. Multicenter trials of antiestrogen and antiaromatase therapy in this disorder are in progress.

Sexual precocity is rare in boys with McCune-Albright syndrome. Affected boys may have asymmetrical enlargement of the testes in addition to signs of sexual precocity. The histologic changes are reminiscent of those in testotoxicosis; the seminiferous tubules are enlarged and exhibit spermatogenesis, and Leydig cells may be hyperplastic. The LH response to LHRH was prepubertal in two cases. The hormonal data (although scant) and the testicular findings appear similar to those in boys with familial testotoxicosis. A 3.8-year-old boy with McCune-Albright syndrome (Arg 201 His mutation detected in bone and testis tissue) had the unusual feature of macro-orchidism (right testis 9 mL, left testis 7 mL) and absence of sexual precocity. He had several cafe-au-lait lesions on the back and a radiograph of the skeleton showed polyostotic fibrous dysplasia. Gonadotropins, the LHRH stimulation test, and sex steroid levels were prepubertal but serum inhibin B and AMH concentrations were strikingly elevated. The tests on histology showed that most seminiferous tubules were "slightly" increased in diameter and filled with Sertoli cells but lacked a lumen. The tubules stained intensively for inhibin B subunit; mature Leydig cells were absent.

The pathogenesis of the sporadic McCune-Albright syndrome was uncertain since its first description. It may occur concordantly or discordantly in monozygotic twins; familial cases have not been described. In 1986, Happle posited that the disorder is caused by an autosomal dominant lethal gene that results in loss of the zygote in utero and that cells bearing this mutation survive only in embryos mosaic for the lethal gene. The early somatic mutation would lead to a mosaic pattern of the distribution of cells containing the mutation. The severity of the disorder would depend on the proportion of mutant cells in various embryonic tissues. The description of somatic mutations in human endocrine tumors that convert the peptide chain of the Gs protein into a putative oncogene (referred to as a gsp mutation) raised the possibility of a similar defect in the McCune-Albright syndrome that both affects a differentiated function such as a signaling pathway and mediates the regulation of proliferation. These hypotheses have now been established. Mutations in the gene encoding the subunit of the stimulatory G protein for adenyl cyclase were identified in the tissues of children with the McCune-Albright syndrome.

The heterotrimeric guanine nucleotide binding proteins (G proteins) are a subfamily within the large superfamily of GTP-binding proteins and serve to transduce signals from a large number of cell-surface receptors with a common structural motif of seven membrane-spanning domains to their intracellular effector molecules, including enzymes and ion channels; in essence, they couple serpentine cell-surface receptors to effectors. For Gs, the stimulatory G protein, the effector is adenyl cyclase, which is controlled by Gs and an inhibitory (Gi) G protein. The heterotrimer is composed of (1) an alpha subunit (39 to 45 kDa) that binds GTP and has intrinsic GTPase activity, which converts GTP to guanosine diphosphate (GDP), and (2) a subunit (35 to 36 kDa) and a smaller subunit (7 to 8 kDa) that are tightly but noncovalently associated with each other. A distinct gene encodes each of the subunits.

The G proteins function as conformational switches. The GDP-activated subunit is bound to the (i) subunits and is in an inactivated state. When its ligand or agonists activate the cell-surface receptor, the GTP is catalytically released from the subunit and enables GTP to bind. This leads to dissociation of the GTP-activated subunit, its dissociation from the bound (i) subunits, and activation of the effector, adenyl cyclase. When GTP is hydrolyzed by the intrinsic GTPase activity of G alpha subunit, the (i) subunits reassociate and the subunit is now in the off or inactive conformation. The three-dimensional structure of the heterotrimeric G proteins has been determined. Activating heterozygous mutations in the G subunit that occurred as an early postzygotic event have now been described in the McCune-Albright syndrome. The somatic constitutive activating mutation, which leads to excess cAMP production and in some tissues cAMP-induced hyperplasia, has a mosaic pattern and the proportion of the hyperactive mutant to normal cells varies in different tissues, contributing, at least in part, to the varied clinical findings, its severity, its sporadic nature, and the discordant occurrence in monozygotic twins. A germ line mutation is presumed to be lethal to the embryo. Two gain-of-function somatic missense mutations have been described in this disorder, both of which involve the arginine 201 residue of the subunit. The site of covalent modification by cholera toxin: arginine 201 with either a cysteine or a histidine substitution (see Fig. 24-77). The arginine

<table>
<thead>
<tr>
<th>Chronologic Age (yr)</th>
<th>Bone Age (yr)</th>
<th>Height (cm)</th>
<th>Physical Signs.</th>
<th>Basal and Post-LHRH</th>
<th>Plasma Estradiol, pmol/L (pg/mL)</th>
<th>Radiograph, Long Bones</th>
</tr>
</thead>
<tbody>
<tr>
<td>1951</td>
<td>1951</td>
<td>1662</td>
<td>A Patient with McCune-Albright Syndrome and Recurrent Ovarian Cysts</td>
<td>![Figure 24-75a](Bone lesions in McCune-Albright syndrome. A, The skull with severe thickening primarily at the base due to fibrous dysplasia. The auditory and optic nerves could be caught in narrowed foramina but that is not the case in these patients. B and C, distortions of the long bones, which can develop into a &quot;shepherd's crook&quot; appearance; note the multiple bone cysts. This mutation survive only in embryos mosaic for the lethal gene. The early somatic mutation would lead to a mosaic pattern of the distribution of cells containing the mutation. The severity of the disorder would depend on the proportion of mutant cells in various embryonic tissues. The description of somatic mutations in human endocrine tumors that convert the peptide chain of the Gs protein into a putative oncogene (referred to as a gsp mutation) raised the possibility of a similar defect in the McCune-Albright syndrome that both affects a differentiated function such as a signaling pathway and mediates the regulation of proliferation. These hypotheses have now been established. Mutations in the gene encoding the subunit of the stimulatory G protein for adenyl cyclase were identified in the tissues of children with the McCune-Albright syndrome.)</td>
<td>![Figure 24-75b](Bone scan showing the areas of remodeling that “light up” depending upon the area affected in individual patients. There are examples of patients primarily affected in the cranial area, in the appendicular area, and in both areas as well as the axial skeleton. (Courtesy of Michael T. Collins, M.D., National Institutes of Health, Bethesda, Maryland and Sandra Gorges, M.D., University of California, Davis.) The heterotrimeric guanine nucleotide binding proteins (G proteins) are a subfamily within the large superfamily of GTP-binding proteins and serve to transduce signals from a large number of cell-surface receptors with a common structural motif of seven membrane-spanning domains to their intracellular effector molecules, including enzymes and ion channels; in essence, they couple serpentine cell-surface receptors to effectors. (Fig. 24-77). For Gs, the stimulatory G protein, the effector is adenyl cyclase, which is controlled by Gs and an inhibitory (Gi) G protein. The heterotrimer is composed of (1) an alpha subunit (39 to 45 kDa) that binds GTP and has intrinsic GTPase activity, which converts GTP to guanosine diphosphate (GDP), and (2) a subunit (35 to 36 kDa) and a smaller subunit (7 to 8 kDa) that are tightly but noncovalently associated with each other. A distinct gene encodes each of the subunits. The G proteins function as conformational switches. The GDP-activated subunit is bound to the (i) subunits and is in an inactivated state. When its ligand or agonists activate the cell-surface receptor, the GTP is catalytically released from the subunit and enables GTP to bind. This leads to dissociation of the GTP-activated subunit, its dissociation from the bound (i) subunits, and activation of the effector, adenyl cyclase. When GTP is hydrolyzed by the intrinsic GTPase activity of G alpha subunit, the (i) subunits reassociate and the subunit is now in the off or inactive conformation. The three-dimensional structure of the heterotrimeric G proteins has been determined. Activating heterozygous mutations in the G subunit that occurred as an early postzygotic event have now been described in the McCune-Albright syndrome. The somatic constitutive activating mutation, which leads to excess cAMP production and in some tissues cAMP-induced hyperplasia, has a mosaic pattern and the proportion of the hyperactive mutant to normal cells varies in different tissues, contributing, at least in part, to the varied clinical findings, its severity, its sporadic nature, and the discordant occurrence in monozygotic twins. A germ line mutation is presumed to be lethal to the embryo. Two gain-of-function somatic missense mutations have been described in this disorder, both of which involve the arginine 201 residue of the subunit. The site of covalent modification by cholera toxin: arginine 201 with either a cysteine or a histidine substitution (see Fig. 24-77). The arginine...</td>
<td></td>
</tr>
</tbody>
</table>
201 residue is critical for subunit GTPase activity, and each of the two mutations decreases the GTPase activity of the \( \alpha \) subunit and leads to constitutive activation. These activating mutations have been found in all tissues affected in the syndrome, including bone lesions.

Manifestations of the McCune-Albright syndrome in infancy include Cushings syndrome related to macronodular adrenal cortical hyperplasia, hyperthyroidism caused by thyroid adenomas, jaundice associated with hepatobilary disease, and pancreatitis. Another nonendocrine manifestation is cardiac disease that carries the risk of cardiac arrhythmia and sudden death.

** Gonadotropin-Independent Sexual Precocity and Pseudohypoparathyroidism Type Ia Caused by a \( \alpha \) Mutation **

Mutations in \( \alpha \) can either constitutively activate or inactivate adenyl cyclase. Two boys who presented in infancy with classical pseudohypoparathyroidism type Ia, a disorder characterized by resistance to hormones whose action is mediated by cAMP, developed signs of sexual precocity with the hormonal characteristics of testotoxicosis (gonadotropin-independent sexual precocity) at about 24 months of age. They both had a unique alanine 366 to serine mutation in one allele of the \( \alpha \) gene; the alanine residue is absolutely conserved in all heterotrimeric G proteins. Pseudohypoparathyroidism type Ia is due to a wide variety of inactivating mutations in \( \alpha \) that lead to about a 50% reduction in GTPase activity in functional assays.

The paradox of a \( \alpha \) mutation causing both inactivation and pseudohypoparathyroidism and constitutive activation and testotoxicosis was resolved by in vitro studies. In cultured cells, the \( \alpha \) Ala 366 Ser mutant protein was rapidly degraded at 37°C but constitutively activated adenyl cyclase at 33°C. Unlike other activating mutations of \( \alpha \), which involve mutations inhibiting its intrinsic GTPase activity and decreasing the rate of hydrolysis of GTP to GDP, the mutation in the two boys caused accelerated dissociation of GDP at 33°C in transfected Leydig cells but rapid degradation at 37°C in a lymphoma cell line and in skin fibroblasts at both 33 and 37°C. The observations explain the clinical consequences of increased \( \alpha \) activity in the testes, which are 3 to 5°C cooler than the body, and the tissue specificity and temperature dependence of the mutation. The mother of one patient appeared to be a mosaic for the \( \alpha \) mutation, whereas a germ line mutation is likely in the other boy.

** Juvenile Hypothyroidism **

Long-standing untreated primary hypothyroidism, usually a consequence of Hashimotos thyroiditis, is an uncommon cause of incomplete sexual precocity in both girls and boys and occurs in association with impaired growth and delayed skeletal maturation. If the concentration of plasma prolactin is elevated, galactorrhea may be demonstrable, more commonly in affected girls than boys (Fig. 24-78). The signs of sexual maturation are not accompanied by a pubertal growth spurt; instead, growth is impaired. Girls have breast development, enlarged labia minora, and estrogenic changes in the vaginal smear, usually without the appearance of pubic hair, and solitary or multiple ovarian cysts may be demonstrable by pelvic sonography or by physical examination. In about 80% of boys with juvenile hypothyroidism, the testes are enlarged because of an increase in the size of the seminiferous tubules, but signs of virilization and Leydig cell maturation are absent and the plasma concentration of testosterone is prepubertal. Enlargement of the sella turcica and the pituitary gland (see Fig. 24-79) has led to the misdiagnosis of a pituitary neoplasm. The hypothyroidism, incomplete sexual maturation, galactorrhea, and pituitary enlargement are reversed or corrected by levothyroxine therapy within a few months.

In 1960, Van Wyk and Grumbach suggested that the syndrome resulted from hormonal “overlap” in negative feedback regulation with increased secretion of gonadotropins, prolactin, and TSH as a consequence of the chronic hypothyroidism. With the advent of radioimmunoassays for pituitary hormones, increased prolactin secretion was documented in children and adults with primary hypothyroidism and in affected girls with the syndrome.

Hyperprolactinemia correlated with the increased production of TSH. GH release is usually decreased as in uncomplicated primary hypothyroidism. Hypothalamic TRH stimulates the release of both prolactin and TSH, and the increased TRH concentration in children with primary hypothyroidism seems to account
for the rise in serum prolactin and TSH levels.

However, the explanation for the sexual maturation remains uncertain. Pubertal development in primary hypothyroidism is usually delayed and is only rarely advanced for chronologic age. By using radioimmunoassays for FSH and LH in which the cross-reaction with TSH is negligible, an increased (pubertal) concentration of plasma immunoreactive and bioactive FSH but not LH has been detected. 

Bioactive LH activity is also low. In addition, increased FSH pulsatility, mainly at night, but not LH release was demonstrated in patients with the syndrome and in some children with primary hypothyroidism who did not exhibit premature sexual maturation. 

The increased FSH release and the high FSH/LH ratio (in contrast to that in normal puberty) seem to account for the increased ovarian estrogen secretion in girls and for the enlarged testes without signs of virilization in affected boys; the suggestion here is that FSH-induced Sertoli cell proliferation is an important determinant of pubertal testicular size, whereas LH has only minor effects on testicular growth (it is not true that LH does not play a role in testicular development). 

An LH-induced increase in the interstitial cell mass is quite unlikely as LH reactivity of these cells is markedly reduced in primary hypothyroidism. 

Pulsatile TSH release is increased at night, and administration of TRH appears to increase FSH release in normal children (but not adults). Moreover, the FSH response to TRH, but not LHRH, is augmented in primary hypothyroidism and this FSH response to TRH can occur in gonadotropin-secreting pituitary adenomas.

If the latter observations are confirmed, it is likely that the incomplete sexual precocity and the increased prolactin secretion and galactorrhea are a consequence of the increased release of TRH, the increased sensitivity of the mammotrophs and gonadotrophs to TRH, or both. This mechanism, which has gained support, would explain the relatively rapid and complete reversal of the syndrome by levotiroxine treatment. 

Recombinant TSH at a dose about 1000-fold greater than hFSH evoked a dose-dependent cAMP response in COS-7 cells transfected with the human FSH receptor, which suggests another possible but less likely mechanism for the FSH-dependent (or FSH-like-dependent) but LHRH-independent sexual precocity. A direct effect of severe hypothyroidism on the prepubertal tests that leads to overproliferation of Sertoli cells has also been advanced to explain the macro-orchidism.

Diagnosis of Sexual Precocity

See Table 24-48 and Figure 24-80 Figure 24-82 . The separation of patients with self-limited benign disorders, such as premature adrenarche or premature thelarche or normal but early puberty, from those with serious or even potentially fatal disorders is the first step in evaluation. The history may reveal symptoms suggesting perinatal abnormalities or injuries, previous infections, adventitious ingestion of or exposure to gonadal steroids, or the presence of similar conditions in family members. In addition, previous measurements should be plotted on a growth chart to determine height velocity and the age of onset of any increase in the rate of growth.

Important aspects of the physical examination include description of the secondary sexual development according to

![Figure 24-77](https://example.com/figure2477)

Tanner stages; measurement of the penis and the testes in boys and breast tissue in girls; and examination for acne, oily skin, facial and body hair, pubic and axillary hair development, apocrine gland odor, muscular development, and galactorrhea. A careful examination of the external genitalia should be done with a nonrelated chaperone present because the performance of such an examination has been interpreted by patients as sexual abuse in several cases. A thorough neurologic examination is indicated, with emphasis on assessment of the visual fields and optic discs and search for signs of increased intracranial pressure; evaluation for skin lesions of the McCune-Albright syndrome or neurofibromatosis; and examination for abdominal, gonadal, or adnexal masses and for coexisting endocrine disease.

Radiographs should be obtained for determination of bone age and an MRI with contrast (or, if necessary, the less sensitive CT brain scan) for investigation of CNS abnormalities in all boys with evidence of LHRH-dependent sexual precocity as well as all girls 6 years of age or younger and selected girls between 8 and 8 years of age. Ultrasonography of the ovary and uterus is exceedingly useful in evaluation of affected girls. Standards are available for the shape and volume of the uterus and the ovaries. 

The largest measurement of uterine size in infants and children by ultrasonography is found at puberty and in the neonatal period. 

The upper limit of uterine length in the pubertal state is 3.5 cm. Further, a uterine volume greater than 1.8 mL is quite specific for the onset of puberty and an increased ovarian volume is less specific. Patients with premature thelarche were indistinguishable from age-matched controls when this sonographic standard was used. The presence of microcysts and macrocysts of the ovary can be detected by ultrasonography as well. Cysts may be found in the ovaries in true precocious puberty or LHRH-independent isosexual precocity, but the cysts are usually smaller than 9 mm in the former and larger than 9 mm in the latter.

Measurements of plasma gonadotropin concentrations using third-generation assays, the plasma concentration of testosterone in boys and of estradiol in girls, and the LH response to administration of LHRH (or the amplitude and frequency of LH pulses, especially at night) are of primary importance in diagnosis. Girls early in the course of true precocious puberty

**Figure 24-78 Left and center, Steven'a, chronic hypothyroidism of Hashimoto's thyroiditis in a 7 1/2-year-old girl with sexual precocity (without pubic or axillary hair), episodic vaginal bleeding, and galactorrhea. She had symptoms of hypothyroidism and a sharply decreased rate of growth over the previous 2 years (height, 1' 11" SD; bone age, 5 3/12 years). Breast development was Tanner stage 3; the labia minora were enlarged, and the vaginal mucosa was pink, thickened, and rugated with evidence of an estrogenic effect. No acne, seborrhea, or hirsutism was present. The uterus was of adolescent size, and the endometrial mucosa was in a proliferative phase. Urinary gonadotropins were barely detectable by biosay. Right, Striking change in appearance after 8 months of thyroid hormone treatment. She had grown 7 cm in height and lost 8.1 kg in weight; the breasts had decreased in size, galactorrhea was no longer demonstrable, the labia minora had regressed, and the vaginal mucosa was pink and glistening (no estrogen effect). Ten weeks after the initiation of thyroid hormone replacement therapy, she developed a right slipped capital femoral epiphysis that was repaired surgically; recovery was uneventful.**

An elevated hCG level with a prepubertal LH response indicates an ectopic, autonomous, gonadotropin-secreting tumor. If this tumor is in the CNS, abnormalities are present on MRI or CT brain scans. Enlargement of the liver or a mediastinal, retroperitoneal mass in boys with sexual precocity suggests an hCG-producing hepatic or
germ cell tumor; the possibility of Klinefelter's syndrome needs to be considered with the latter. Pubertal concentrations of LH and FSH, a pubertal mode of pulsatile LH secretion (initially during sleep), or a pubertal LH response in the LHRH test confirms the diagnosis of true precocious puberty (and in boys differentiates true precocity from familial testotoxicosis). A CNS tumor must be considered as a potential cause of this premature activation of the hypothalamic LHRH pulse generator.

The evaluation for a CNS tumor as a cause of true precocious puberty is similar to the investigation of an hCG-secreting tumor of the CNS. Although CT scanning is now a well-established procedure for determining the presence of a CNS abnormality, MRI with contrast is more sensitive for the detection of small tumors in the hypothalamus, such as a hamartoma of the tuber cinereum (see Fig. 24-63). The use of contrast adds to diagnostic certainty and is recommended for MRI of the CNS. The height of the pituitary gland on MRI correlates with advancing age and with pubertal development; patients with true precocious puberty and higher peak LH/FSH ratios had pituitary heights exceeding 6 mm on the average whereas those with a lower LH/FSH ratio or precocious thelarche had lower heights of approximately 5 mm. The shape of the pituitary gland is also of importance; a convex appearance rather than a flat top is associated with true precocious puberty of all etiologies. The size and shape of the pituitary gland do not decrease with successful LHRH therapy.

The premature appearance of pubic hair, phallic enlargement, and other signs of virilization in a male without enlargement of the testes or the liver suggests the diagnosis of congenital virilizing adrenal hyperplasia, virilizing adrenal tumor, or, rarely, Cushings syndrome. Measurement of plasma 17-hydroxyprogesterone and DHEAS concentrations and the excretion of urinary 17-ketosteroids and their suppressibility with glucocorticoids distinguishes adrenal hyperplasia from a virilizing adrenal tumor. If growth rate is suppressed, the possibility of primary hypothyroidism or of Cushing's syndrome is raised.

![Figure 24-79](image)

The appearance of pubic hair without other signs of puberty in boys or girls is usually a result of premature adrenarche but may be the first sign of sexual precocity or of adrenal virilism of other causes.

In a girl, breast development associated with dulling and thickening of the vaginal mucosa and enlargement of the labia minora indicates significant estrogen secretion in iatrogenic exposure to estrogen. The differential diagnosis includes true precocious puberty, an estrogen-secreting neoplasm, and a cyst of the ovary. If the plasma concentrations of gonadotropins in girls in early or true precocious puberty are in the pubertal range much of the day, if LH pulses of pubertal amplitude are detected, or if a pubertal LH response to LHRH is elicited, true precocious puberty is present. Estrogen concentrations in girls early in normal or true precocious puberty are in the prepubertal range much of the day, and a single determination may be inadequate to reflect ovarian function (see Fig. 24-60). A CNS tumor is less likely in girls than in boys to be the cause of this premature reactivation of the hypothalamic LHRH pulse generator in a gonadal system. However, studies using MRI or CT brain scans indicate that the hypothalamic hamartoma is more prevalent in both boys and girls with so-called idiopathic true precocious puberty than was previously suspected.

If the concentration of plasma estradiol is elevated but gonadotropin levels are low, an estrogen-secreting cyst or neoplasm is present. Ovarian tumors of moderate size can be palpated by bimanual examination. Advances in pelvic sonography allow the delineation of ovarian cysts or tumors and the determination of uterine size, and this procedure has become an essential component of the diagnostic evaluation. An estrogen-secreting neoplasm of the ovary is usually accompanied by high estradiol concentrations. However, some ovarian cysts are associated with concentrations of estradiol as high as those in granulosa cell tumors; the differential diagnosis between these cysts and ovarian neoplasms rarely requires exploratory laparotomy or laparoscopy and can usually be made by pelvic sonography and by the use of tumor markers. Breast development in the absence of other estrogen effects is almost always a result of premature thelarche.

**Intragenic Sexual Precocity**

Prepubertal children are remarkably sensitive to exogenous gonadal steroids and may show signs of sexual maturation resulting from overlooked sources of androgens or estrogens, such as ingested or absorbed tonics, lotions, or creams that contain or are inadvertently contaminated with an estrogen. Interest in the effects of environmental estrogens or disruptors on reproductive development in girls and boys is mounting. Estrogen exposure may come from cosmetics. Hair creams and straighteners may contain estrogenic substances; this source of exogenous estrogens is more potent than low levels of endogenous estrogen secreted by the thelarche of the ovaries.

**TABLE 24-48 -- Differential Diagnosis of Sexual Precocity**

<table>
<thead>
<tr>
<th>Plasma Gonadotropins</th>
<th>LH Response to LHRH</th>
<th>Serum Sex Steroid Concentration</th>
<th>Gonadal Size</th>
<th>Miscellaneous</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Precocious Puberty (premature reactivation of LHRH pulse generator)</td>
<td>Prominent LH pulses, initially during sleep</td>
<td>Pubertal LH response</td>
<td>Pubertal values of testosterone or estradiol</td>
<td>Normal pubertal testicular enlargement or ovarian and uterine enlargement (by ultrasonography)</td>
</tr>
</tbody>
</table>

**Incomplete Sexual Precocity (pluripotential gonadotropin-independent)**

**Males**

<table>
<thead>
<tr>
<th>Chorionic gonado-tropin-secreting tumor in males</th>
<th>High hCG, low LH response</th>
<th>Pubertal value of testosterone</th>
<th>Slight to moderate uniform enlargement of testes</th>
<th>Hepatomegaly suggests hepatoblastoma; CT scan of brain if chorionic gonadotropin-secreting CNS tumor suspected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leydig cell tumor in males</td>
<td>Suppressed</td>
<td>No LH response</td>
<td>Very high testosterone</td>
<td>Irregular asymmetrical enlargement of testes</td>
</tr>
</tbody>
</table>

**TABLE 24-48 -- Differential Diagnosis of Sexual Precocity**

<table>
<thead>
<tr>
<th>Plasma Gonadotropins</th>
<th>LH Response to LHRH</th>
<th>Serum Sex Steroid Concentration</th>
<th>Gonadal Size</th>
<th>Miscellaneous</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Precocious Puberty (premature reactivation of LHRH pulse generator)</td>
<td>Prominent LH pulses, initially during sleep</td>
<td>Pubertal LH response</td>
<td>Pubertal values of testosterone or estradiol</td>
<td>Normal pubertal testicular enlargement or ovarian and uterine enlargement (by ultrasonography)</td>
</tr>
</tbody>
</table>

**Incomplete Sexual Precocity (pluripotential gonadotropin-independent)**

**Males**

<table>
<thead>
<tr>
<th>Chorionic gonado-tropin-secreting tumor in males</th>
<th>High hCG, low LH response</th>
<th>Pubertal value of testosterone</th>
<th>Slight to moderate uniform enlargement of testes</th>
<th>Hepatomegaly suggests hepatoblastoma; CT scan of brain if chorionic gonadotropin-secreting CNS tumor suspected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leydig cell tumor in males</td>
<td>Suppressed</td>
<td>No LH response</td>
<td>Very high testosterone</td>
<td>Irregular asymmetrical enlargement of testes</td>
</tr>
</tbody>
</table>
### Females

<table>
<thead>
<tr>
<th>Condition</th>
<th>LH Response</th>
<th>Estradiol Values</th>
<th>Other Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Granulosa cell tumor (follicular cysts may present similarly)</td>
<td>Prepubertal</td>
<td>Very high estradiol</td>
<td>Ovarian enlargement on physical examination, CT, or ultrasonography</td>
</tr>
<tr>
<td>Follicular cyst</td>
<td>Prepubertal</td>
<td>Prepubertal to very high estradiol</td>
<td>Ovarian enlargement on physical examination, CT, or ultrasonography</td>
</tr>
<tr>
<td>Feminizing adrenal tumor</td>
<td>Prepubertal</td>
<td>High estradiol and DHEAS</td>
<td>Ovaries prepubertal</td>
</tr>
<tr>
<td>Premature thelarche</td>
<td>Prepubertal</td>
<td>Prepubertal or early sex steroids pubertal or higher</td>
<td>Ovaries prepubertal</td>
</tr>
<tr>
<td>Premature adrenarche</td>
<td>Prepubertal</td>
<td>Prepubertal estradiol</td>
<td>Ovaries prepubertal</td>
</tr>
<tr>
<td>Late-onset virilizing congenital adrenal hyperplasia</td>
<td>Prepubertal</td>
<td>Elevated 17-OHP in basal or corticotropin-stimulated state</td>
<td>Ovaries prepubertal</td>
</tr>
</tbody>
</table>

### In Both Sexes

<table>
<thead>
<tr>
<th>Syndrome</th>
<th>LH Response</th>
<th>Estradiol Values</th>
<th>Other Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>McCune-Albright syndrome</td>
<td>Suppressed</td>
<td>Suppressed sex steroids puberty or higher</td>
<td>Ovarian (on ultrasound); slight testicular enlargement</td>
</tr>
<tr>
<td>Primary hypo-roidism</td>
<td>LH prepubertal; FSH may be slightly elevated</td>
<td>Estradiol may be pubertal</td>
<td>Testicular enlargement; ovaries cystic T and prolactin elevated; T low</td>
</tr>
</tbody>
</table>

CNS, central nervous system; CT, computed tomography; DHEAS, dehydroepiandrosterone sulfate; hCG, human chorionic gonadotropin; LH, luteinizing hormone; MRI, magnetic resonance imaging; 17-OHP, 17-hydroxyprogesterone; T, thyroxine; TSH, thyrotropin.

who inhaled estrogen dust have developed sexual precocity. A short course of application of estrogen cream is used to treat labial adhesions, but long courses may lead to breast development or even withdrawal bleeding. In addition to breast development, pigmentation of the areolae and the linea alba and the appearance of public hair may be seen in children exposed to estrogen.

Food is another potential source of estrogen. Epidemics of gynecomastia in boys and thelarche in girls have occurred in schoolchildren in Italy; meat contaminated by estrogens was suspected in some cases. However, no etiology was uncovered in Milan and environs, where 21.1% of 1- to 2-year-old girls and 36.6% of 1- to 2-year-old boys were reported with premature thelarche or gynecomastia. During a 10-year period more than 600 cases of gynecomastia in boys and premature thelarche or incomplete sexual precocity in girls were discovered in Puerto Rico, the highest prevalence noted in the world, about 10 to 15 times higher than the prevalence in a survey in Olmsted, Minnesota. Maternal ovarian cysts were demonstrated in two thirds of affected Puerto Rican girls.

The wide publicity given these observations and the questions raised about contamination of the food supply by the clandestine use of estrogens as growth-promoting agents for meat production caused anxiety among parents, cattle raisers, and farmers. It was suggested that the use of estrogen preparations in animals to stimulate weight gain led to ingestion of estrogen-contaminated meat. Although this idea has not been confirmed by selected analyses of meat, poultry, and milk in Puerto Rico by the U.S. Department of Agriculture, it has not been excluded. Guidelines from the FDA define a limit of not more than 1% of normal daily estrogen production of prepubertal children as a safe intake of estrogen, which translates into 0.43 ng/day for boys and 3.24 ng/day for girls using the latest data from extremely sensitive estrogen assays. A possible association has been advanced between plasticizers with documented estrogenic and antiandrogenic activity and premature
thelarche in Puerto Rico. Significantly elevated concentrations of phthalates and their major metabolites were found in 28 girls (68%) of a cohort with premature breast development.  

Girls breast-fed following an accidental exposure of their mothers to polybrominated biphenyls experienced early menarche (by about 1 year) and early appearance of pubic hair but not breast development compared with girls who were not exposed and girls who were not breast-fed.  

The administration of hCG to boys with undescended testes may induce secretion of testosterone sufficient to cause incomplete sexual precocity. From these examples it is clear that careful investigation of sources of possible exposure to exogenous hormones is mandatory in every case of sexual precocity.

Feminization in Boys and Virilization in Girls (Corresponding Precocity)

Boys

Feminization in a boy before the age of puberty is rare. Rarely, an estrogen-secreting adrenal adenoma or a chorioneplithelioma may cause gynecomastia. Gynecomastia has been reported in a 1-year-old boy with 11-hydroxylase deficiency and in boys with late-onset congenital adrenal hyperplasia.

Aromatase Excess Syndrome

Gynecomastia in prepubertal boys can also be caused by increased extraglandular aromatization of C19 steroids of adrenal origin such as androstenedione and hence increased extraglandular estrogen production in sporadic or familial cases. This autosomal dominant disorder leading to excess estrogen synthesis from C19 precursors because of aromatase overexpression, especially in fat and skin, is a consequence of novel gain-of-function mutations of CYP19, the gene encoding aromatase, related to a chromosome arrangement that give rise to a cryptic promoter.

Feminizing Testicular Tumors

Feminizing testicular tumors may cause gynecomastia in boys younger than age 6 with the Peutz-Jeghers syndrome. Both testes may be enlarged, and the histology indicates sex cord or Sertoli cell tumors that form annular tubules and often have areas of calcification; increased estradiol secretion is noted in the basal state, and a further rise occurs after hCG administration. Otherwise, feminizing Sertoli cell tumors are rare in boys. Sonography or MRI scans of the testes may be useful in the diagnosis.

Girls

Virilization in a girl indicates organic disease except for premature adrenarche. Congenital adrenal hyperplasia resulting from 21-hydroxylase or 11-hydroxylase deficiency and androgen-producing tumors of the adrenal can cause virilization, and these were discussed earlier as occurring in males. 3-Hydroxydehydrogenase deficiency is a rare type of congenital adrenal hyperplasia characterized by elevated 5,17-hydroxyepiandrosterone, DHEA, and DHEAS levels and, in the severe form, decreased secretion of aldosterone and cortisol. Severely affected patients have mineralocorticoid and glucocorticoid deficiency and may die in infancy. Excess adrenal androgens lead to virilization in utero and to ambiguous external genitalia, including clitormal enlargement in females with continued virilization after birth (see Chapter 22). Milder forms of this disorder can cause hirsutism in women. 46,XY phenotypic women with incomplete forms of androgen resistance syndrome or with 17-hydroxydehydrogenase deficiency (17-HSD) type 3 deficiency may have virilization as well as breast development at the time of expected puberty. Mutations in the CYP19 gene that encodes aromatase are associated not only with intrauterine masculinization of the external genitalia in affected XX individuals but also with progressive virilization, lack of female secondary sex characteristics, multicystic ovaries at the age of puberty, tall stature, and osteopenia (see Chapter 22).

Cushing's syndrome resulting from adrenal carcinoma is usually manifest as growth failure with or without virilization, obesity, and moon facies; striae may not appear until months to years later.

The syndrome of glucocorticoid resistance is manifest in various degrees. Some patients demonstrate hyperandrogenic signs such as acne, hirsutism, male-type baldness, menstrual irregularities, and oligomenorrhea and infertility. Dexamethasone decreased the excessive adrenal androgen secretion, virilization, and advancing bone age found in a prepubertal boy with general glucocorticoid resistance.

Anthemoblastoma, the most common virilizing ovarian tumor, is rare in children. Lipoid cell tumor of the ovary and gonadoblastoma are even more unusual sources of androgens.

Variations of Pubertal Development

Premature Thelarche

Unilateral or bilateral breast enlargement without other signs of sexual maturation (e.g., sexual hair and growth of the labia minora and the uterus) is not uncommon in infancy and childhood and has been referred to as premature thelarche. The disorder usually occurs by age 2 (in over 80%) and rarely after age 4. In a retrospective study in Minnesota, premature thelarche occurred with an incidence of 21.2 per 100,000 patient-years, 40% of cases were noted during 6 months and 2 years of age, and most regressed in 6 months to 6 years after diagnosis, although a few persisted until puberty. A 10- to 35-year follow-up was available in 25 cases, and no untoward effects on later health, growth, or fertility were evident. The breast enlargement usually regresses after a few months but occasionally persists for years or lasts until the onset of normal puberty; in about half of affected girls the breast development, which is characteristically cyclic, lasts 3 to 5 years. Usually, significant nipple and areola development is absent and estrogen-induced thickening and dulling of the vaginal mucosa are uncommon. Enlargement of the uterus on ultrasonography (>1.8 mL volume and length > 36 mm) is rare. Measurement of the ellipsoid volume of the uterus (V = longitudinal diameter × anteroposterior diameter × transverse diameter × 0.523) is the most sensitive and specific discriminator between premature thelarche and early true precocious puberty and provides better early discrimination than the LH response to LHRH. Growth in stature is normal.

Premature thelarche is a benign, self-limited disorder compatible with normal pubertal development at an appropriate age; only reassurance and follow-up are usually necessary. The appearance of premature thelarche can, however, be the harbinger of further sexual maturation in a minority of cases as discussed earlier. Because the development may be unilateral, it is important to consider the condition in girls with unilateral breast development so that needless worry about a breast neoplasm is not stimulated in the parents and no unnecessary surgical procedure is carried out. Indeed, the removal of tissue in premature thelarche may leave the child with no possibility of future breast development. In selected instances, sonography of the breast is useful in distinguishing unilateral premature...
Although premature adrenarche is usually considered a benign condition with no substantial long-term risk, accumulating observations indicate that girls with hyperandrogenism, hirsutism, anovulation, amenorrhea or oligomenorrhea, and insulin resistance and compensatory hyperinsulinemia with its attendant risk of major

The concept of exaggerated adrenarche was first advanced in relation to a postulated childhood antecedent of PCOS, the hallmarks of which are hyperandrogenism, hirsutism, anovulation, amenorrhea or oligomenorrhea, and insulin resistance and compensatory hyperinsulinemia with its attendant risk of major metabolic sequelae including type 2 diabetes mellitus, dyslipidemia, an increased propensity for coronary artery disease, and in about 50% of affected women obesity. It has been extended to include rare instances of premature adrenarche associated with excessive responses of 17-hydroxyprogrenolone, DHEAS, and androstenedione to ACTH found in women with functional adrenal hyperandrogenism.

Premature adrenarche is commonly slowly progressive and does not have an untoward effect on either the onset or normal progression of gonadarche or final adult height.
premature adrenarche are at increased risk of developing functional ovarian hyperandrogenism and the polycystic ovarian syndrome, hyperinsulinism, acanthosis nigricans, and dyslipidemia in adolescence and adult life, especially if fetal growth was reduced and the birth weight was low. Premature adrenarche is a risk factor for the later development of PCOS and functional ovarian hyperandrogenism in adolescent and adult women; the magnitude of this risk is unknown, but it appears to vary with the exception of girls with a history of decreased fetal growth. Plasma androstenedione levels above the 90th percentile for age may be helpful in predicting hyperandrogenism in girls with premature adrenarche. The risk of PCOS or clinical hyperandrogenism in these subjects may be as high as 50%. 

Girls with reduced fetal growth are at risk for a reduced number of ovarian primordial follicles at birth, small ovaries and uterus at puberty, and an increased serum FSH level and decreased estradiol concentrations, suggesting relative ovarian resistance to FSH. Girls who are short, especially black and Hispanic girls, have a higher risk of the association of premature adrenarche with syndrome X (obesity, hyperinsulinism, dyslipidemia, and later coronary heart disease) and the development of PCOS in late adolescence and early adulthood. Low birth weight, low birth length, and low birth head circumference were independently associated with hyperandrogenism. If women were exposed to androgens during fetal life, they also have increased risk of development of PCOS and hyperandrogenism in adult life.

Of interest, the adrenal steroid pattern in the black and Hispanic patients in the latter study did not differ from that in children with uncomplicated premature adrenarche.

As discussed previously, hyperinsulinism is associated with many metabolic and endocrine conditions and functional ovarian hyperandrogenism that in some cases is heralded by premature adrenarche. When the role of insulin resistance and hyperinsulinism was recognized in the pathogenesis of PCOS, therapeutic approaches to reduce insulin resistance were introduced, especially the use of insulin sensitizers. Among the latter, the most widely used is metformin because of its low prevalence of adverse effects and therapeutic efficacy. In early studies, this agent was shown to decrease insulin resistance and levels of fasting and postprandial glucose. However, in subsequent studies, metformin achieved its effects on insulin resistance and hyperinsulinism, independent of its effects on adiposity, cholesterol, and serum triglycerides in obese adolescents. In this regard, metformin has been shown to decrease fasting serum insulin, improve insulin sensitivity, and reduce circulating levels of the proinflammatory cytokines such as interleukin-6 and tumor necrosis factor-alpha. Although metformin is effective in reducing insulin resistance and hyperinsulinism, the long-term efficacy and safety of metformin in reducing cardiovascular disease risk in adolescents and young adults with PCOS and premature adrenarche remain to be established.

Premature adrenarche can be associated with nonclassical congenital adrenal hyperplasia caused by homozygous or compound heterozygous missense mutations in the CYP21 gene encoding the 21-hydroxylase enzyme. Premature adrenarche is a risk factor for the later development of PCOS and functional ovarian hyperandrogenism in girls with premature adrenarche. Therefore, it is important to identify those with 21-hydroxylase deficiency prior to development of hyperandrogenism and PCOS. The 21-hydroxylase enzyme is necessary for the biosynthesis of cortisol from 11-deoxycortisol and 18-hydroxycorticosterone. Patients with 21-hydroxylase deficiency will present in early childhood with a salt-losing crisis, hypotension, dehydration, and hyperkalemia._cardiovascular disease including women with PCOS, was increased in girls with premature adrenarche, especially those with low birth weights, and may be useful in the identification of those with a greater risk of developing PCOS. In girls with reduced fetal growth are at risk for a reduced number of ovarian primordial follicles at birth, small ovaries and uterus at puberty, and an increased serum FSH level and decreased estradiol concentrations, suggesting relative ovarian resistance to FSH. Girls who are short, especially black and Hispanic girls, have a higher risk of the association of premature adrenarche with syndrome X (obesity, hyperinsulinism, dyslipidemia, and later coronary heart disease) and the development of PCOS in late adolescence and early adulthood. Low birth weight, low birth length, and low birth head circumference were independently associated with hyperandrogenism. If women were exposed to androgens during fetal life, they also have increased risk of development of PCOS and hyperandrogenism in adult life.

Of interest, the adrenal steroid pattern in the black and Hispanic patients in the latter study did not differ from that in children with uncomplicated premature adrenarche.

As discussed previously, hyperinsulinism is associated with many metabolic and endocrine conditions and functional ovarian hyperandrogenism that in some cases is heralded by premature adrenarche. When the role of insulin resistance and hyperinsulinism was recognized in the pathogenesis of PCOS, therapeutic approaches to reduce insulin resistance were introduced, especially the use of insulin sensitizers. Among the latter, the most widely used is metformin because of its low prevalence of adverse effects and therapeutic efficacy. In early studies, this agent was shown to decrease insulin resistance and levels of fasting and postprandial glucose. However, in subsequent studies, metformin achieved its effects on insulin resistance and hyperinsulinism, independent of its effects on adiposity, cholesterol, and serum triglycerides in obese adolescents. In this regard, metformin has been shown to decrease fasting serum insulin, improve insulin sensitivity, and reduce circulating levels of the proinflammatory cytokines such as interleukin-6 and tumor necrosis factor-alpha. Although metformin is effective in reducing insulin resistance and hyperinsulinism, the long-term efficacy and safety of metformin in reducing cardiovascular disease risk in adolescents and young adults with PCOS and premature adrenarche remain to be established.

Premature adrenarche can be associated with nonclassical congenital adrenal hyperplasia caused by homozygous or compound heterozygous missense mutations in the CYP21 gene encoding the 21-hydroxylase enzyme. Premature adrenarche is a risk factor for the later development of PCOS and functional ovarian hyperandrogenism in girls with premature adrenarche. Therefore, it is important to identify those with 21-hydroxylase deficiency prior to development of hyperandrogenism and PCOS. The 21-hydroxylase enzyme is necessary for the biosynthesis of cortisol from 11-deoxycortisol and 18-hydroxycorticosterone. Patients with 21-hydroxylase deficiency will present in early childhood with a salt-losing crisis, hypotension, dehydration, and hyperkalemia. cardiov...
Fragile X syndrome is the most common inherited cause of mental retardation. This condition is due to multiple repeats of a CCC expansion that leads to hypermethylation of the \textit{FMR1} gene (Xq,27fra) that prevents transcription and translation of the \textit{FMRP} protein. Affected individuals have mental retardation in association with multiple physical anomalies including large testes. Eighty percent to 95% of adolescents or adults with fragile X syndrome have testicular volume greater than 30 mL with an average of 45 mL, although the 95th percentile for the syndrome is 70 mL. In most cases the enlargement begins at 8 to 9 years of age, prior to the appearance of pubic hair, although some prepubertal children already have a testicular volume greater than 4 mL.

The enlarged testes are due to increased interstitial volume and excessive connective tissue, including increased peri-tubular collagen fibers, rather than to increase in the seminiferous tubules. Testicular biopsy has demonstrated normal Leydig and Sertoli cells, normal to slightly decreased spermatogenic cells, and an increase in testicular interstitial fluid. Although there may be subtle elevation of serum gonadotropins that might be part of the etiology of macro-orchidism, affected males are fertile, although most are not sexually active. Other associated anomalies include increased birth weight, high forehead, large ears, prognathism, pale irises, and an increased head circumference.

Macro-orchidism without androgenization is a rare manifestation of the McCune-Albright syndrome. Macro-orchidism is an occasional finding in prepubertal boys with long-standing primary hypothyroidism. This form of testicular enlargement appears to result from increased FSH secretion independent of a pubertal increase in LH secretion or a pubertal LH response to LHRH (see above). Testicular adrenal rests in congenital adrenal hyperplasia (see Chapter 22) and a lymphoma can cause bilateral macro-orchidism. It was a feature of severe aromatase deficiency in a young male adult and in men with an FSH-secreting pituitary macroadenoma. Bilateral megalotestis (testicular volume 26 mL) in adults can occur as a normal variant. One may speculate that some instances of bilateral macro-orchidism are due to a heterozygous constitutive activating mutation of the FSH receptor.
Disorders of Sexual Differentiation with Both Virilization and Feminization at Puberty

Virilization as well as feminization at puberty may occur in a phenotypic female who has a 46,XY karyotype in certain types of male pseudohermaphroditism (see Chapter 22). 17-HSD type 3 deficiency (a testosterone biosynthetic defect) and incomplete forms of androgen resistance (resulting from defects in the androgen receptor) may occur in this manner; however, ambiguous genitalia are usually noted early in life in these conditions. True hermaphrodites with ovarian and testicular tissue may undergo both virilization and feminization at puberty.
References


Ross GT. Follicular development: the life cycle of the follicle and puberty. In Grumbach MM, Sizonenko PC, Aubert MI (eds). Control of the Onset of Puberty. Baltimore, Williams & Wilkins, 1990,


1246


References


1248


Ho Q, Ding ZY, Fong DY, Karth berg J. Blood pressure is associated with body mass index in both normal and obese children. Hypertension 2000; 36:165170.


Feinberg I, Carlson VR. Sleep variables as a function of age in man. Arch Gen Psychiatry 1968; 18:1823918250.


References


453. Halpern CT, Joyner K, Udry JR, Suchindran C. Smart teens don’t have sex (or kiss much either). J Adolescent Health 2000; 26:213225.


476. Clark PA, Inmananish A, Veldhuis JD, Regul AD. Comparison of pulsatile luteinizing hormone secretion between prepubertal children and young adults: evidence for a mass/amplitude-dependent difference without gender or day/night contrasts. J Clin Endocrinol Metab 1997; 82:29502965.


482. Roth JC, Kelch RP, Kaplan SL, Grumbach MM. FSH and LH response to luteinizing hormone releasing factor in prepubertal and pubertal children, adult males and patients with hypogonadotropic


1252


Andersson A, Skakkebaek NE. Serum inhibin B levels during male childhood and puberty. Mol Cell Endocrinol 2001; 180:130107.


Baker ML, Hutson JM. Serum levels of mullerian inhibiting substance in boys throughout puberty and in the first two years of life. J Clin Endocrinol Metab 1993; 76:245247.


References

1993; 100:1125-1131; discussion 1131-1132.


691. Duan Y, Tabensky A, DeLuca V, Seeman E. The benefit of hormone replacement therapy on bone mass is greater at the vertebral body than posterior processes or proximal femur. Bone 1997; 21:447451.
700. Thind KK, Goldsmith PC. Infundibular gonadotropin-releasing hormone neurons are inhibited by direct opioid and autoregulatory synapses in juvenile monkeys. Neuroendocrinology 1988; 47:203216.


Wray S, Grant P, Gainer H. Evidence that cells expressing luteinizing hormone-releasing hormone mRNA in the mouse are derived from progenitor cells in the olfactory placode. Proc Natl Acad Sci USA 1995; 86:1128136.

References


Zamorano PL, Mahesh VB, De Sevilla L, Brann DW. Excitatory amino acid receptors and puberty. Steroids 1998; 63:268270.


843. Plant TM, Durrant AR. Circulating leptin does not appear to provide a signal for triggering the initiation of puberty in the male rhesus monkey (Macaca mulatta). Endocrinology 1997; 138:4054-4058.


References


907. Kasuya E, Nyberg CL, Mogi K, Terasawa E. A role of gamma-amino butyric acid (GABA) and glutamate in control of puberty in female rhesus monkeys: effect of an antisense oligodeoxynucleotide for GAD67 messenger ribonucleic acid and MK801 on luteinizing hormone releasing hormone release. Endocrinology 1999; 140:705712.


909. Terasawa E, Luchansky LL, Kasuya E, Nyberg CL. An increase in glutamate release follows a decrease in gamma aminobutyric acid and the pubertal increase in luteinizing hormone releasing hormone release in the female rhesus monkeys. J Neuroendocrinol 1999; 11:275282.


932. Terasawa E, Luchansky LL, Kasuya E, Nyberg CL. An increase in glutamate release follows a decrease in gamma aminobutyric acid and the pubertal increase in luteinizing hormone releasing hormone release in the female rhesus monkeys. J Neuroendocrinol 1999; 11:275282.


934. Terasawa E. Hypothalamic control of the onset of puberty. Curr Opin Endocrinol Diabet 1999; 6:44.


References


120. McCabe ER. Vulnerability within a robust complex system DAX1 mutations and steroidogenic axis development (editorial). J Clin Endocrinol Metab 2002; 87:4143.


131. Phelan JC, McCabe ER. Mutations in NR0B1 (DAX1) and NR5A1 (SF1) responsible for adrenal hypoplasia congenita. Hum Mutat 2001; 18:472487.


144. Tannen JM, Whitehouse RH. A note on the bone age at which patients with true isolated growth hormone deficiency enter puberty. J Clin Endocrinol Metab 1975; 41:78790.


References


1217. Lalande M. In and around SNRPN. Nat Genet 1994; 8:56.


1219. Laurence JZ, Moon RC. Four cases of "retinitis pigmentosa," occurring in the same family, and accompanied by general imperfections of development. Ophthalmic Rev 1866; 2:3241.


References

gonadotropin and LH. J Clin Endocrinol Metab 2000; 85:2281-2286.


1401. Nistal M, Riesola ML, Pianiuaga R. Correlation between testicular biopsies (prepubertal and postpubertal) and spermogram in cryptorchid hum. Hum Pathol 2000; 31(9):10221030.


1411. Skakkebaek NE, Rappert-De Meyts E, Main HM. Testicular dysgenesis syndrome: an increasingly common development disorder with environmental aspects. Hum Reprod 2001; 16:972978.


References


References


1651. Palmert MR, Mansfield MJ, Crowley WF, et al. Is obesity an outcome of gonadotropin-releasing hormone agonist administration? Analysis of growth and body composition in 110 patients with


1277

Med 1981; 305:15461550.

1768.

1766.

1986; 7:4466.

1764.

1762.

1761.

84:21042110.

1759.

1757.

1756.

1755.

1754.

1753.

1752.

1751.

1750.

1749.

1748.

1747.

1746.

1745.

1744.

1743.

1742.

1741.

1740.

1739.

1738.

1737.

1736.

1735.

1734.

1733.

1732.

1731.

1730.

1729.


1278.


References


References


1283


1994. Marshall GR, Plant TM. Puberty occurring either spontaneously or induced precociously in rhesus monkey (Macaca mulatta) is associated with a marked proliferation of Sertoli cells. Biol Reprod 1996; 54:11521159.


2055. The aromatase excess syndrome is associated with feminization of both sexes and autosomal dominant transmission of aberrant P450 aromatase gene transcription. J Clin Endocrinol Metab 1998; 83:13481357.


Oberfield SE. Metabolic lessons from the study of young adolescents with polycystic ovary syndrome: is insulin, indeed, the culprit? J Clin Endocrinol Metab 2000; 85:35203525.


1285
1286


Chapter 25 - Endocrinology and Aging

Steven W. J. Lamberts

The average length of human life is currently 75 to 78 years and may increase to 85 years during the coming ten years, but it is not clear whether these additional years will be satisfactory. Most data indicate a modest gain in the number of healthy years lived but a far greater increase in years of compromised physical, mental, and social function. The number of days with restricted activity and admissions to hospitals and nursing homes increases sharply after 70 years of age. The U.S. National Health Interview Survey indicated that more than 25 million aging people suffer from physical impairment and the number of persons requiring assistance with the activities of daily living increases from 14% at ages 65 to 75 years to 45% in people older than age 85 years.
AGING AND PHYSICAL FRAILTY

Throughout adult life, all physiologic functions start to decline gradually. There is a diminished capacity for cellular protein synthesis, a decline in immune function, an increase in fat mass, a loss of muscle mass and strength, and a decrease in bone mineral density. Most older adults die of atherosclerosis, cancer, or dementia, but in an increasing number of the “healthy” oldest old, loss of muscle strength is the limiting factor that determines their chances of an independent life until death.

Age-related disability is characterized by generalized weakness, impaired mobility and balance, and poor endurance. In the oldest old, this state is termed physical frailty, defined as “a state of reduced physiological reserves associated with increased susceptibility to disability.” Clinical correlates of physical frailty include falls, fractures, impairment in activities of daily living, and loss of independence. Falls contribute to 40% of admissions to nursing homes.

Loss of muscle strength is an important factor in the development of frailty. Muscle weakness can be caused by aging of muscle fibers and their innervation, osteoarthritis, and chronic debilitating diseases. A sedentary lifestyle, decreased physical activity, and disuse, however, are also important determinants of the decline in muscle strength.

In a study of 100 frail nursing home residents (average age, 87 years), lower extremity muscle mass and strength were closely related. Supervised resistance exercise training (45 minutes three times a week for 10 weeks) doubled muscle strength and significantly increased gait velocity and stair-climbing power. This finding demonstrates that frailty in the elderly population is not an irreversible effect of aging and disease but can be influenced and perhaps even prevented. Further, in nondisabled elderly persons living in the community, objective measures of lower extremity function are highly predictive of subsequent disability. Prevention of frailty can be achieved only by working (training). However, exercise is difficult to implement in the daily routine of the aging population, and the number of dropouts from exercise programs is very high.

Part of the aging process involving body composition (i.e., loss of muscle [strength] and bone, increase in fat mass) might also be related to changes in the endocrine system. Current knowledge has shed light on the effects of long-term hormonal replacement therapy on body composition as well as on atherosclerosis, cancer formation, and cognitive function.
THE ENDOCRINOLOGY OF AGING

The two most important clinical changes in endocrine activity during aging involve the pancreas and the thyroid gland.

Approximately 40% of individuals aged 65 to 74 years and 50% of those older than 80 years have impaired glucose tolerance or diabetes mellitus, and in nearly 50% of elderly adults with diabetes the disease is undiagnosed. These adults are at risk for development of secondary, mainly macrovascular, complications at an accelerated rate. Pancreatic, insulin receptor, and postreceptor changes associated with aging are critical components of the endocrinology of aging. Apart from decreased (relative) insulin secretion by the beta cells, peripheral insulin resistance related to poor diet, physical inactivity, increased abdominal fat mass, and decreased lean body mass contributes to the deterioration of glucose metabolism. Dietary management, exercise, oral hypoglycemic agents, and insulin are the four components of treatment for these patients, whose medical care is costly and intensive.

Age-related thyroid dysfunction is also common. Lowered plasma thyroxine (T₄) and increased thyrotropin concentrations occur in 5% to 10% of elderly women. These abnormalities are mainly caused by autonomy and are thus an expression of age-associated disease rather than a consequence of the aging process. Normal aging is accompanied by a slight decrease in pituitary thyrotropin release but especially by decreased peripheral degradation of T₄, which results in a gradual age-dependent decline in serum triiodothyronine (T₃) concentrations without changes in T₄ levels. This slight decrease in plasma T₃ concentrations occurs largely within the broad normal range of the healthy elderly population and has not been convincingly related to functional changes during the aging process. At present, the question of whether healthy aging subjects might benefit from T₃ replacement therapy remains unresolved.

Changes in the hormone levels of normal women (Fig. 25-1) and men (Fig. 25-2) during aging. Note the difference in the distribution of ages in the different panels.

The second hormonal system demonstrating age-related changes is adrenopause, a term that describes the gradual decline in circulating levels of dehydroepiandrosterone (DHEA) and its sulfate (DHEAS). Adrenal secretion of DHEA gradually decreases over time, while corticocortropin secretion, which is physiologically linked to plasma cortisol levels, remains largely unchanged. The decline in DHEA and DHEAS levels in both sexes, therefore, contrasts with the maintenance of plasma cortisol levels and seems to be caused by a selective decrease in the number of functional zona reticularis cells in the adrenal cortex instead of being regulated by a central (hypothalamic) pacemaker of aging.

THE SECONDARY SYSTEMS OF AGING

1. The course of serum insulin-like growth factor (IGF-I) concentrations occurs largely 6 months after the initiation of menopause, serum LH and FSH levels increase sharply. Right, The adrenocortical cells responsible for the production of dehydroepiandrosterone (DHEA) decrease in activity (adrenopause) without clinically evident changes in corticosterone (adrenocorticoctropic hormone, ACTH) and cortisol secretion. There is no central pacemaker in the hypothalamus or higher brain areas (or both) that is hypothalamic or higher brain areas (or both) is hypothesized; which together with changes in the peripheral organs (the ovaries, testicles, and adrenal cortex) regulates the aging process of these endocrine axes.

2. Changes in the activity of the hypothalamic-pituitary-gonadal axis in men are slower and more subtle. During aging, a gradual decline in serum total and free testosterone levels occurs. Changes in the activity of the hypothalamic-pituitary-gonadal axis in men are slower and more subtle. During aging, a gradual decline in serum total and free testosterone levels occurs. Changes in the activity of the hypothalamic-pituitary-gonadal axis in men are slower and more subtle. During aging, a gradual decline in serum total and free testosterone levels occurs. The most dramatic and rapidly occurring change in women around age 50 years is menopause, characterized by a decrease in testicular Leydig cell numbers and their secretory capacity as well as by a decrease in estradiol. Estradiol production during the reproductive years is cycling estradiol production during the reproductive years is replaced by very low, constant estradiol levels. For many years, the prevailing view was that menopause resulted from exhaustion of ovarian follicles. An alternative perspective is that age-related changes in the central nervous system and the hypothalamic-pituitary-ovarian unit initiate the menopausal transition. The evidence that both the ovary and the brain are key pacemakers in menopause is compelling.

3. Changes in insulin sensitivity and thyroid function that occur in the aging population are frequently of clinical importance and recognized and treated as diseases. Changes in the hormone levels of normal women (Fig. 25-1) and men (Fig. 25-2) during aging. Note the difference in the distribution of ages in the different panels.

4. The two most important clinical changes in endocrine activity during aging involve the pancreas and the thyroid gland.

5. Approximately 40% of individuals aged 65 to 74 years and 50% of those older than 80 years have impaired glucose tolerance or diabetes mellitus, and in nearly 50% of elderly adults with diabetes the disease is undiagnosed. These adults are at risk for development of secondary, mainly macrovascular, complications at an accelerated rate. Pancreatic, insulin receptor, and postreceptor changes associated with aging are critical components of the endocrinology of aging. Apart from decreased (relative) insulin secretion by the beta cells, peripheral insulin resistance related to poor diet, physical inactivity, increased abdominal fat mass, and decreased lean body mass contributes to the deterioration of glucose metabolism. Dietary management, exercise, oral hypoglycemic agents, and insulin are the four components of treatment for these patients, whose medical care is costly and intensive.

6. Age-related thyroid dysfunction is also common. Lowered plasma thyroxine (T₄) and increased thyrotropin concentrations occur in 5% to 10% of elderly women. These abnormalities are mainly caused by autonomy and are thus an expression of age-associated disease rather than a consequence of the aging process. Normal aging is accompanied by a slight decrease in pituitary thyrotropin release but especially by decreased peripheral degradation of T₄, which results in a gradual age-dependent decline in serum triiodothyronine (T₃) concentrations without changes in T₄ levels. This slight decrease in plasma T₃ concentrations occurs largely within the broad normal range of the healthy elderly population and has not been convincingly related to functional changes during the aging process. At present, the question of whether healthy aging subjects might benefit from T₃ replacement therapy remains unresolved.

7. Changes in the hormone levels of normal women (Fig. 25-1) and men (Fig. 25-2) during aging. Note the difference in the distribution of ages in the different panels.

8. The most dramatic and rapidly occurring change in women around age 50 years is menopause, characterized by a decrease in testicular Leydig cell numbers and their secretory capacity as well as by a decrease in estradiol. Estradiol production during the reproductive years is cycling estradiol production during the reproductive years is replaced by very low, constant estradiol levels. For many years, the prevailing view was that menopause resulted from exhaustion of ovarian follicles. An alternative perspective is that age-related changes in the central nervous system and the hypothalamic-pituitary-ovarian unit initiate the menopausal transition. The evidence that both the ovary and the brain are key pacemakers in menopause is compelling.

9. Changes in insulin sensitivity and thyroid function that occur in the aging population are frequently of clinical importance and recognized and treated as diseases. Changes in the hormone levels of normal women (Fig. 25-1) and men (Fig. 25-2) during aging. Note the difference in the distribution of ages in the different panels.

10. The two most important clinical changes in endocrine activity during aging involve the pancreas and the thyroid gland.

11. Approximately 40% of individuals aged 65 to 74 years and 50% of those older than 80 years have impaired glucose tolerance or diabetes mellitus, and in nearly 50% of elderly adults with diabetes the disease is undiagnosed. These adults are at risk for development of secondary, mainly macrovascular, complications at an accelerated rate. Pancreatic, insulin receptor, and postreceptor changes associated with aging are critical components of the endocrinology of aging. Apart from decreased (relative) insulin secretion by the beta cells, peripheral insulin resistance related to poor diet, physical inactivity, increased abdominal fat mass, and decreased lean body mass contributes to the deterioration of glucose metabolism. Dietary management, exercise, oral hypoglycemic agents, and insulin are the four components of treatment for these patients, whose medical care is costly and intensive.

12. Age-related thyroid dysfunction is also common. Lowered plasma thyroxine (T₄) and increased thyrotropin concentrations occur in 5% to 10% of elderly women. These abnormalities are mainly caused by autonomy and are thus an expression of age-associated disease rather than a consequence of the aging process. Normal aging is accompanied by a slight decrease in pituitary thyrotropin release but especially by decreased peripheral degradation of T₄, which results in a gradual age-dependent decline in serum triiodothyronine (T₃) concentrations without changes in T₄ levels. This slight decrease in plasma T₃ concentrations occurs largely within the broad normal range of the healthy elderly population and has not been convincingly related to functional changes during the aging process. At present, the question of whether healthy aging subjects might benefit from T₃ replacement therapy remains unresolved.

13. Changes in insulin sensitivity and thyroid function that occur in the aging population are frequently of clinical importance and recognized and treated as diseases. Changes in the hormone levels of normal women (Fig. 25-1) and men (Fig. 25-2) during aging. Note the difference in the distribution of ages in the different panels.

14. The most dramatic and rapidly occurring change in women around age 50 years is menopause, characterized by a decrease in testicular Leydig cell numbers and their secretory capacity as well as by a decrease in estradiol. Estradiol production during the reproductive years is cycling estradiol production during the reproductive years is replaced by very low, constant estradiol levels. For many years, the prevailing view was that menopause resulted from exhaustion of ovarian follicles. An alternative perspective is that age-related changes in the central nervous system and the hypothalamic-pituitary-ovarian unit initiate the menopausal transition. The evidence that both the ovary and the brain are key pacemakers in menopause is compelling.

15. Changes in insulin sensitivity and thyroid function that occur in the aging population are frequently of clinical importance and recognized and treated as diseases. Changes in the hormone levels of normal women (Fig. 25-1) and men (Fig. 25-2) during aging. Note the difference in the distribution of ages in the different panels.

16. The second hormonal system demonstrating age-related changes is adrenopause, a term that describes the gradual decline in circulating levels of dehydroepiandrosterone (DHEA) and its sulfate (DHEAS). Adrenal secretion of DHEA gradually decreases over time, while corticocortropin secretion, which is physiologically linked to plasma cortisol levels, remains largely unchanged. The decline in DHEA and DHEAS levels in both sexes, therefore, contrasts with the maintenance of plasma cortisol levels and seems to be caused by a selective decrease in the number of functional zona reticularis cells in the adrenal cortex instead of being regulated by a central (hypothalamic) pacemaker of aging.

17. The third endocrine system that gradually declines in activity during aging is the growth hormone (GH)insulin-like growth factor (IGF-I) system.
<table>
<thead>
<tr>
<th>Disease</th>
<th>No Treatment</th>
<th>$E_2 + P$</th>
<th>Relative Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coronary heart disease</td>
<td>46.1</td>
<td>30.4</td>
<td>0.66</td>
</tr>
<tr>
<td>Stroke</td>
<td>19.8</td>
<td>19.3</td>
<td>0.96</td>
</tr>
<tr>
<td>Fractures</td>
<td>30.40</td>
<td>152.8</td>
<td>0.500.70</td>
</tr>
<tr>
<td>Dementia</td>
<td>16.3</td>
<td>11.5</td>
<td>0.71</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>10.2</td>
<td>13.51</td>
<td>1.351.80</td>
</tr>
<tr>
<td>Endometrial cancer</td>
<td>2.6</td>
<td>2.6</td>
<td>1.00</td>
</tr>
<tr>
<td>Life expectancy (years)</td>
<td>82.8</td>
<td>83.8</td>
<td></td>
</tr>
</tbody>
</table>

A number of limitations and assumptions must be considered when interpreting this table. The duration of the use and dose regimens of $E_2 + P$ varied considerably between studies included in the meta-analysis (duration, 210 years). It was assumed that the addition of progestagen to the estrogen regimen would increase the risk for breast cancer from 1.35 to 1.80. 

$E_2 + P$, estrogen plus progestagen.


*The estimated lifetime probabilities of developing the conditions mentioned have been derived from mortality and incidence data from the 1987 Vital Statistics of the United States. The relative risks are the best estimates of the relative risk for developing each condition in long-term hormone users compared with nonusers. These estimates were derived from a model of the risks and benefits of hormone therapy developed by Grady et al. 

MENOPAUSE

Menopause is the permanent cessation of menstruation resulting from the loss of ovarian follicular function and is diagnosed retrospectively after 12 months of amenorrhea.

In most women, vasomotor reactions, depressed mood, and urogenital complaints accompany this period of estrogen decline. In the subsequent years, the loss of estrogens is followed by a high incidence of cardiovascular disease, loss of bone mass, and cognitive impairment. The average age of menopause (51.4 years) has not changed over time and seems to be largely determined by genetic factors.

The use of hormone replacement therapy (HRT), consisting of estrogen or estrogen plus progestagen, can alleviate the symptoms of menopause (perimenopausal use), but long-term use of HRT (5 to 10 years or more) may also be advantageous in preventing cardiovascular disease, bone loss, and cognitive impairment.

Perimenopausal Use of Hormone Replacement Therapy

Typical symptoms that result from the sudden decrease in estrogen production around menopause are menstrual cycle disorders, vasomotor changes (hot flushes, night sweats), and urogenital complications (atrophic vaginal irritation and dryness, dyspareunia, atrophic urethral epithelium leading to micturition disorders). Additional symptoms are irritability, mood swings, joint pain, and sleep disturbances. Frequency, severity, onset, and duration of symptoms vary widely between individuals and between ethnic groups. About 75% of women in Western societies experience so few troublesome symptoms during the menopausal transition that HRT is not needed or requested.

HRT rapidly alleviates the symptoms of menopause. Hot flushes and vasomotor instability as well as symptoms of urogenital atrophy rapidly disappear upon the start of HRT.
Long-Term Hormone Replacement Therapy

Because life expectancy is increasing, the time a woman spends after menopause constitutes more than one third of her life. Long-term use of HRT (5 to 10 years) seems to have advantages with regard to the prevention of the three chronic disorders most common in the elderly: (1) cardiovascular diseases, (2) osteoporosis, and (3) dementia. There are, however, also important adverse effects of long-term estrogen-progestagen replacement therapy after menopause. The most compelling problem is the increased incidence of breast cancer. \(^\text{26} 29 30 33\) The excess risk increased by 8% for each year of combined hormone use and by 1% for each year of estrogen-only use. Thus, risk of breast cancer would be predicted to increase by approximately 80% after 10 years of estrogen-progestagen use.

Lifetime probabilities of disease occurrence for a 50-year-old white woman entering menopause without or with subsequent HRT are presented in Table 25-1. \(^\text{27} 28\)

Coronary Heart Disease and Stroke

Nearly every observational study has demonstrated a decreased risk of heart disease in women who ever used estrogen. Meta-analyses of 25 published studies of women who used estrogen and 7 studies that separately assessed estrogen plus progestagen treatment found summary relative risks of 0.70 and 0.66, respectively, for coronary heart disease among women. \(^\text{34} 35\) The apparent benefit is largely limited to current or recent estrogen use. HRT does not play a role in secondary prevention: progression of coronary atherosclerosis in women with established disease was not influenced. \(^\text{36} 37\) HRT was not consistently associated with a reduced risk of stroke. \(^\text{38}\)

The mechanism of cardiosprotection remains uncertain but probably involves multiple actions. Estrogen is an antioxidant and calcium blocker and induces beneficial effects on concentrations of serum low-density lipoprotein (LDL) cholesterol (lowering) and high-density lipoprotein (HDL) cholesterol (increasing). Added progestagen attenuates these estradiol-mediated effects experimentally, but epidemiologically there is no convincing evidence for a decreased effect of combined therapy in comparison with estrogen alone in the primary prevention of coronary disease. \(^\text{39} 40\)

Bone Loss

Osteoporosis is characterized by low bone mass and microarchitectural deterioration of bone tissue, leading to increased bone fragility and, therefore, to fracture susceptibility. The lifetime risk of fractures in 50-year-old white women is 30% to 40% (see Table 25-1). \(^\text{41}\)

The efficacy of HRT on the main sites of osteoporotic fractures has been documented in case-control and cohort studies but only in a few prospective controlled trials. \(^\text{42} 43\) Current use of HRT, especially long-term use, is associated with a reduction in the risk of hip fractures by about 30% and of spine fractures by about 50%. Most osteoporotic fractures occur after age 65. Long-term use is necessary to decrease the incidence of fractures substantially. After HRT is stopped, bone loss resumes within a year and bone turnover increases to the levels observed in untreated women within 3 to 6 months, which probably accounts for the lack of fracture protection in past users. \(^\text{44}\)

Estrogen reduces bone turnover and increases bone density in postmenopausal women in part because it improves calcium homeostasis. The addition of calcium potentiates the effect of estrogen on bone mass. The further addition of androgens (low-dose testosterone) or an antiresorptive drug (bisphosphonates) may further increase bone formation in women most at risk for fractures.

Dementia

A number of experimental studies indicate that estrogens may directly influence the brain by a number of mechanisms, including activation of the cholinergic system, inhibition of oxidative stress and neuronal apoptosis, and an increase in synaptic plasticity. Estrogen may also, through its effects on the cardiovascular system, reduce the risk of vascular dementia. \(^\text{45}\)

Several studies have suggested that HRT improves cognition, prevents development of dementia, and decreases the severity of dementia, but other studies have not shown this benefit of estrogen use. \(^\text{46} 47\)

It is now well recognized that cognition improves in perimenopausal women using HRT. However, most studies suggest that this improvement occurs because menopausal symptoms are alleviated and that there is no clear beneficial effect of HRT on cognition in asymptomatic women. \(^\text{48}\)

Ten observational studies have measured the effect of postmenopausal estrogen use on the risk of development of dementia. A meta-analysis of these studies suggested a 29% decreased risk of dementia among estrogen users. \(^\text{49} 50\) However, results of eight small uncontrolled trials of estrogen use in women with dementia or Alzheimer's disease did not demonstrate a clear benefit for cognition. \(^\text{51}\)

Given the limited data, no definite conclusions can be reached about the effect of HRT in reducing cognitive decline and dementia in older women. Although the data available are promising, HRT is not currently recommended for the prevention or treatment of dementia. \(^\text{52} 53\)

Other Benefits

HRT is associated with slightly longer overall survival. Apart from the clear, rapidly occurring effect on menopausal symptoms, no improvement in quality of life was observed in asymptomatic older women receiving HRT.

Breast and Endometrial Cancer

Late menopause has long been known to be associated with an increased risk, and early menopause with a reduced risk, of breast cancer. This observation is consistent with the idea that prolonged exposure to endogenous estrogen is an adverse risk factor. For every 1-year increase in age at menopause, there is about a 3% increase in the risk of breast cancer. \(^\text{54} 55\)

Most studies have found no increased risk of breast cancer in women who had ever used estrogen, usually for less than 2 years in the perimenopausal period. The relative risk increase for breast cancer in HRT users seems largely confined to current or recent use.

A meta-analysis of more than 50 studies clearly demonstrated that the risk of breast cancer increases with long-term estrogen use. \(^\text{56} 57\) Among women who used estrogen for 5 years or longer (median use, 11 years), the summary relative risk for breast cancer was 1.35. Among 1000 women who used HRT continuously for 10 years starting at age 50 years, it was estimated that there would be an additional six breast cancers, raising the incidence from a background of 45 cases to 51 cases. However, these data mainly refer to the use of estrogens only. In the Breast Cancer Detection Demonstration Project, in which 46,000 women participated, the estrogen-progestagen regimens were associated with greater increases in breast cancer risk compared with estrogen alone. \(^\text{58}\) The excess risk increased by 8% for each year of combined hormone use and by 1% for each year of estrogen-only use. Thus, risk of breast cancer would be predicted to increase by approximately 80% after 10 years of estrogen-progestagen use. \(^\text{59} 60\)
An association between endometrial cancer and estrogen use was observed many years ago. Ten years of unopposed estrogen use increases the risk for endometrial cancer 10-fold.\textsuperscript{26} For this reason, the HRT regimens were supplemented with progestagens, which almost completely prevented this excess risk for endometrial cancer.

Other Risks

HRT doubles a woman's risk of needing gallbladder surgery. It also doubles the risk of deep vein thrombosis and pulmonary embolism; however, the absolute risk is low, about 3 cases per 10,000 treated women per year.
Hormone Replacement Regimens

As described by Barrett-Connor,[27] presently advised doses of estrogen were designed to prevent bone loss, and progestagen regimens were proposed to prevent endometrial cancer. The advice, however, has not been based on studies of a wide range of doses. Several estrogen and progestagen preparations are available for HRT (Table 25-2). Components of available preparations vary in their effects on different target tissues. Commercial preparations differ in their clinical effects by design, and individual women differ in their responses. HRT can be administered orally, transdermally, topically, intranasally, or as subcutaneous implants.

Estrogen has distinct route-dependent effects on somatotropic action. Oral estrogens probably lower serum IGF-I concentrations through impairment of hepatic IGF-I production. This effect does not occur after transdermal estrogen administration. Increasing evidence suggests that transdermally administered estrogen thus has more beneficial effects on protein metabolism and body composition.

Prevention of endometrial hyperplasia and cancer induction by estrogen depends on both dose and duration of progestagen use. Uterine protection requires 12 days of cyclic progestagens or combined continuous regimens. The former causes scheduled bleeding and the latter causes unpredictable spotting or bleeding, which usually resolves within 9 months.[28][42]

Initially after the start of HRT, side effects, including mastalgia, bloating, bleeding, premenstrual tension, and depression, can occur. To prevent these side effects but also to increase compliance, it is generally recommended that the patient start with half the estrogen dose.

Data indicate a close relationship between endogenous circulating estrogen levels and bone loss, bone mineral density, and fractures.[44] Women with detectable serum estradiol concentrations (8 to 92 pmol/L; 5 to 25 pg/mL) had higher bone mineral density, significantly less bone loss, and a lower risk for subsequent hip fractures than women with undetectable estradiol levels. These findings suggest that much lower estrogen doses might be sufficient to maintain bone than those indicated in Table 25-2.
Indications for Hormone Replacement Therapy

Perimenopausal or menopausal HRT is strongly indicated for women having premature menopause, women with clinically important symptoms of the menopausal transition, and women entering menopause with osteoporosis.

The concept that long-term HRT after menopause is an effective risk reduction strategy for coronary heart disease, fractures, and cognitive decline has to be balanced against the increased risk of breast cancer. The decision to start long-term HRT should be based on an individual's risk factors, attitude toward hormonal treatment, and knowledge of its risks and benefits. Individualization of the treatment decision is mandatory. Both knowledge and education influence the decision to start HRT; in a Swedish study, only 24% of women 54 years of age but 72% of female general practitioners and 88% of female gynecologists were receiving estrogen-progestagen replacement therapy. \[45\]
Selective Estrogen Receptor Modulators

A new development in the search for optimal hormone replacement therapy during menopause came from observations that tamoxifen has variable antiestrogenic and estrogenic actions in different tissues. Tamoxifen suppresses the growth of estrogen receptor-positive breast cancer cells. Long-term treatment of menopausal patients with breast cancer with tamoxifen also lowered the incidence of new (contralateral) breast cancer by 40%. In addition, the number of cardiovascular incidents decreased by 70%, and the age-related decrease in bone mineral density was partially prevented.

These initially puzzling observations were explained by the fact that tamoxifen and other compounds such as raloxifene have selective estrogen receptormodulating effects, exerts antiestrogenic actions on normal and cancerous breast tissue but agonistic actions on bone, lipids, and the blood vessel walls. These effects of tamoxifen and raloxifene may be explained by differential stabilization of the conformation of the estrogen receptor but might also be related to the activation of different estrogen receptor forms, in which the form is the classical estrogen receptor, whereas a form mediates the vascular and bone effects of estrogens.

Raloxifene is the second selective estrogen receptor modulator (SERM) available for clinical use in menopausal women. It demonstrates estrogen agonist activity on bone and lipid metabolism and has estrogen antagonist activity in uterine and breast tissue. The 60-mg dose is currently approved for the prevention and treatment of postmenopausal osteoporosis.

The efficacy and safety of raloxifene for the prevention of osteoporosis in postmenopausal women were proved in a study that demonstrated a 2.5% increase in bone mineral density in the lumbar spine and hip in a group of postmenopausal nonosteoporotic women treated with raloxifene for 2 years. A significant reduction of vertebral fracture risk by raloxifene was subsequently demonstrated. A total of 7705 postmenopausal women with existing osteoporosis were studied. After 36 months, bone mineral density at the hip and spine increased in the women treated with 60 mg of raloxifene by 2.1% and 2.6%, respectively, compared with those receiving placebo. At 36 months, 7.4% of women had at least one new vertebral fracture, including 10.1% of women receiving placebo and 6.6% of those receiving raloxifene at 60 mg/day. Compared with the placebo group, those receiving 60 mg of raloxifene had a relative risk for fracture of 0.7 (P < .001). Forty-six subjects needed raloxifene at 60 mg for 3 years to prevent one vertebral fracture in menopausal women without an existing fracture; for those with an existing fracture, 16 subjects required treatment.

Raloxifene has effects on lipids similar to those of estrogen, except for a relatively small effect on high-density lipoprotein cholesterol and no significant effect on triglycerides. Data on cardiovascular event rates and on cognitive function are not yet available.

Raloxifene, in contrast to tamoxifen and estrogen, does not stimulate endometrial thickness or vaginal bleeding. With regard to side effects, raloxifene causes an increased incidence of leg cramps and hot flashes.

A most promising effect of raloxifene is its chemoprotective action against breast cancer. Cummings and colleagues reported the effects in 7705 postmenopausal women (mean age, 66.5 years) with osteoporosis. Statistical significance of the difference between the groups was P < .001. To prevent one case of breast cancer, 126 women would need to be treated. Raloxifene decreased the risk of estrogen receptorpositive breast cancer by 90% but did not affect the risk of estrogen receptornegative invasive breast cancer. This important study demonstrated that among postmenopausal women with osteoporosis, the risk of invasive breast cancer was decreased by 76% during 3 years of treatment with raloxifene (Fig. 25-3).

Figure 25-3 Effect of raloxifene administration (60 to 120 mg/day) on the cumulative incidence of breast cancer in 7705 postmenopausal women (mean age, 66.5 years) with osteoporosis. Statistical significance of the difference between the groups was P < .001. (From Cummings SR, Eckert S, Krueger KA, et al. The effect of raloxifene on risk of breast cancer in postmenopausal women: results from the MORE randomized trial. Multiple Outcomes of Raloxifene Evaluation. JAMA 1999; 281:21892197.)
Hormone Replacement Therapy, Selective Estrogen Receptor Modulators, or No Treatment?

The issue of hormonal intervention for postmenopausal women is controversial, and many aspects remain unresolved. The idea that HRT is a global risk reduction strategy is being reevaluated. Although the general clinical benefits of HRT in the short term during and after the menopausal transition are clear, the balance of the beneficial effects of long-term HRT after menopause versus negative effects, especially on breast cancer incidence, remains worrisome.

Currently, a vast armamentarium of pharmacologic treatments to reduce cardiovascular and bone risks is available; these include cholesterol-lowering statins, -blockers, SERMs, and bisphosphonates. An optimal choice of these different lifestyle drugs for menopausal women requires individualization of the treatment decision. Coronary artery disease, for example, is a complex disorder, resulting from an interaction of genetic predisposition and environmental factors. Risk factor modification (diet, smoking, physical activity) should be advised. Primary prevention of coronary artery disease with HRT seems effective; more effective for existing atherosclerosis are lipid-lowering drugs, aspirin, nitrates, and -blockers.

For women with existing osteoporosis, HRT is very effective. However, SERMs and bisphosphonates come close in their fracture-reducing effects. Recognition of an increased risk for breast cancer in menopausal women is an important consideration in the choice of SERMs. If the impressive preventive effect of raloxifene on breast cancer is confirmed to last much longer than 3 years, chemoprevention of breast cancer will probably become a major consideration in the pharmacologic choice for risk reduction in the long-term preventive treatment of postmenopausal women.
ANDROPAUSE

Role of Testosterone during Aging

Age-associated hypogonadism does not develop as clearly in men at andropause as in women at menopause. The key difference is the gradual, often subtle change in androgen levels in men versus the precipitate fall of estrogen production in women. It is generally agreed that as men age, there is a decline in serum total testosterone concentration that begins after the age of 40 years. In cross-sectional studies, the annual decline in total and free testosterone is 0.4% and 1.2%, respectively. The higher decline in free testosterone levels is related to the increase in sex hormone-binding globulin (SHBG) levels with age.

It remains unclear whether the well-known biologic changes occurring during aging in men (e.g., reduced sexual activity, muscle mass and strength, and skeletal mineralization) are causally related to these changes in testosterone bioactivity (andropause).

In a group of more than 400 independently living elderly men (mean age, 78 years; range, 73 to 94 years), Beld and colleagues observed a positive association between total and free serum testosterone concentrations and muscle strength as well as an inverse relationship with fat mass. Low bioavailable testosterone was associated with a depressed mood in a population-based study of 856 men aged 50 to 89 years.
Testosterone Replacement Therapy

Many persuasive reports in the literature demonstrate that testosterone replacement in men of all ages (young, adult, and old) with clear clinical and biochemical hypogonadism instantly reverses vasomotor activity (flushes and sweats); improves libido, sexual activity, and mood; increases muscle mass, strength, and bone mineralization; prevents fractures; decreases fat mass; and decreases fatigue and poor concentration. Also, the treatment of normal adult men with supraphysiologic doses of testosterone, especially when combined with resistance exercise training, increased fat-free mass and muscle mass and strength.

Most studies reporting the results of androgen therapy in older men were small, short-term, noncontrolled, and without uniform end points. The results of a large randomized study in healthy elderly men have now been published and seem representative for effects expected of androgen therapy. Ninety-six men (mean age 73 years) wore a testosterone patch on their scrotum (6 mg of testosterone per 24 hours) or a placebo patch for 36 months. Mean serum testosterone concentrations in the men treated with testosterone increased from 12.7 ± 2.9 nmol/L (367 ± 7.9 ng/dL) before treatment to 21.7 ± 8.6 nmol/L (625 ± 249 ng/dL; P < .001) at 6 months of treatment and remained at that level for the duration of the study. The decrease in fat mass (3.0 ± 0.5 kg) in the testosterone-treated men during the 36 months of treatment was significantly different from the decrease (0.7 ± 0.5 kg) in the placebo-treated men (P < .001). The increase in lean mass (1.9 ± 0.3 kg) in the testosterone-treated men was significantly different from that in the placebo-treated men (0.2 ± 0.2 kg; P < .001).

Changes in knee extension and flexion strength, hand grip, walking speed, and other parameters of muscle strength and function were not significantly different in the two groups. Bone mineral density in the lumbar spine increased in both the testosterone-treated (4.2% ± 0.8%) and placebo-treated (2.5% ± 0.8%) groups, but mean changes did not differ between groups (see Fig. 25-4). However, the lower the pretreatment serum testosterone concentration, the greater the effects of testosterone treatment on lumbar spine bone density after 36 months (P = .02). A minimal effect (0.9 ± 1.0%) of testosterone treatment on bone mineral density was observed in men with a pretreatment serum testosterone concentration of 13.9 nmol/L (400 ng/dL), but an increase of 5.9% ± 2.2% was found in men with a pretreatment testosterone concentration of 6.9 nmol/L (200 ng/dL).

The subjective perception of physical function decreased significantly during the 36 months of treatment in the placebo-treated group (P < .001) but not in the testosterone-treated group. Interestingly, the effect of testosterone treatment on the perception of physical functioning varied inversely with the pretreatment serum testosterone concentration (P < .01). There was no significant difference between the two treatment groups with regard to the subjective perception of energy or sexual functions.

With regard to the potential adverse effects of testosterone treatment in healthy elderly men, again the study by Snyder and colleagues seems representative. The mean serum prostate-specific antigen (PSA) concentration did not change during the 36 months of treatment in the placebo-treated group but increased by a relatively small but statistically significant (P < .001) amount by 6 months of treatment in the testosterone-treated group and remained relatively stable for the remainder of the study. The urine flow rate, volume of urine in the bladder after voiding, and number of clinically significant prostate events during the 3 years of the study were similar in the two groups. Hemoglobin and hematocrit did not change in the placebo-treated group during treatment, but both increased significantly (P < .001) in the testosterone-treated group within 6 months and remained relatively stable for the remainder of the study. Three men treated with testosterone developed persistent erythrocytosis (hemoglobin > 17.5 g/dL; hematocrit > 52%) during treatment.

Other androgen replacement studies in older men have demonstrated that lipid profiles are not adversely affected by this therapy, but the incidence of cardiovascular events in healthy elderly men who receive androgens for extended periods has not been studied. Numerous studies of large populations of healthy men have shown a marked rise in the incidence of impotence to over 50% in men 60 to 70 years old. Although this increased rate occurs in the same age group who show a clear decline in serum (free) testosterone levels, no causal relationships have been demonstrated. In most instances, testosterone replacement therapy in elderly men is not effective for the treatment of loss of libido or impotence in individuals with serum testosterone concentrations within the normal range in age-matched subjects. Other factors, such as atherosclerosis, alcohol consumption, smoking, and the quality of personal relationships, seem to be more important.

This result suggests that there is a threshold level of testosterone in the low normal range below which libido and sexual function are impaired and above which there is no further enhancement of response.

Summarizing the available literature, the indiscriminate (preventive) treatment of healthy elderly men with testosterone at a dose that increases serum testosterone concentrations to those observed in 20- to 30-year-olds has limited anabolic effects on body composition (a slight decrease in fat mass and a slight increase in muscle mass). No beneficial effects on muscle strength or physical performance are observed.

Detailed analyses of a number of studies of elderly men selected on the basis of low pretreatment serum testosterone concentrations indicated a beneficial effect of testosterone replacement therapy on muscle strength, bone mineral density, mood, and (subjective) aspects of the quality of life. This introduces the question of how to select elderly men who might benefit from testosterone treatment.

There is great interindividual variation in serum testosterone concentrations among healthy men. In adulthood, the biochemical definition of male hypogonadism is generally accepted if serum total testosterone is below 10.4 to 12.1 nmol/L (300 to 500 ng/dL), depending on the population studied. Of healthy men between the ages of 60 to 80 years, 20% demonstrated a serum testosterone concentration below 10.4 nmol/L (300 ng/dL). In the same study, non-SHBG testosterone levels were below the lower limit for young men (<1.6 nmol/L) in over 60% of these men.

An important question is how to define true hypogonadism in elderly men. Because many signs and symptoms of hypogonadism coincide with those observed during the normal aging process, no good single measure of clinically significant hypogonadism is available. In a discussion of testosterone replacement in older men, it was suggested that the biochemical diagnosis of hypogonadism seems certain if the serum total testosterone concentration is less than 7.0 nmol/L (200 ng/dL).
Serum total testosterone levels in a group of healthy young (n = 58; aged 21 to 35 years) and older (n = 96; aged 60 to 80 years) men. Mean ± standard deviation. Serum total testosterone is 18.2 ± 4.2 nmol/L (525 ± 122 ng/dL) for the young men and 14.5 ± 4.5 nmol/L (420 ± 12.9 ng/dL) for the older men. (From Tenover JS. Androgen administration to aging men. Endocrinol Metab Clin North Am 1994; 23:877-892.)

ng/dL). This cutoff remains arbitrary and does not answer the question of whether healthy elderly men with testosterone levels between 7.0 and 10.4 nmol/L are hypogonadal or whether such men would benefit from replacement therapy with testosterone. Also, it has been demonstrated that intercurrent diseases frequently result in a transient, sharp drop in serum testosterone concentrations, whereas frail, elderly men in general tend to have testosterone levels 10% to 15% lower than those of healthy, age-matched control subjects.

When a serum testosterone concentration is found to be below 7.0 nmol/L (200 ng/dL), an additional evaluation with measurements of serum gonadotropins and prolactin is mandatory in order to exclude pituitary pathology.
Conclusions

Testosterone at supraphysiologic doses, when administered to eugonadal men, increases muscle mass and strength. Replacement therapy directed at restoring serum testosterone in healthy elderly men to levels observed between the ages of 30 and 50 years lowers fat mass and increases lean mass to a limited extent without a beneficial effect on muscle strength and physical performance. It remains uncertain whether testosterone replacement produces clinically meaningful improvements in muscle function without significant adverse effects in frail older men or in elderly men with serum testosterone concentrations between 7.0 and 11.4 nmol/L.

If one decides to start testosterone replacement, the major goal of therapy is to return testosterone levels to values as close to "physiologic" levels in age-matched controls as possible. The dose should thus be titrated according to serum levels. Considerations concerning the choice of testosterone preparation as well as the route of administration (oral, injectable, implantable, or transdermal) are discussed in Chapter 8.

At present, the duration of testosterone administration is uncertain. Control of prostate size, PSA levels, and hematocrit is mandatory. The identification of elderly men who might benefit most from testosterone treatment remains uncertain, and the risks to the prostate and possible effects on the process of atherosclerosis require further study. The development of androgenic compounds with variable biologic action in different organs (selective androgen receptor modulation) is currently being pursued.
ADRENOPAUSE

Role of Dehydroepiandrosterone during Aging

Humans are unique among primates and rodents because the human adrenal cortex secretes large amounts of the steroid precursor DHEA and its sulfate derivative DHEAS. Serum DHEAS concentrations in adult men and women are 100 to 500 times higher than those of testosterone and 1000 to 10,000 times higher than those of estradiol. In normal subjects, serum concentrations of DHEA and its sulfate are highest in the third decade of life, after which the concentrations of both gradually decrease, so that by the age of 70 to 80 years, the values are about 20% of peak values in men and 30% of peak values in women (see Fig. 25-2).

DHEA and DHEAS seem to be inactive precursors that are transformed within human tissues by a complicated network of enzymes into androgens or estrogens, or both (Fig. 25-6). The key enzymes are aromatase, steroid sulfatase, 3-hydroxysteroid dehydrogenases (3-HSD-1, -2), and at least seven organ-specific 17-hydroxysteroid dehydrogenases (17-HSD-1 to -7). Labrie and colleagues introduced the term intracrinology to describe this synthesis of active steroids in peripheral target tissues in which the action is exerted in the same cells in which synthesis takes place, without release into the extracellular space and general circulation.

In postmenopausal women, nearly 100% of sex steroids are synthesized in peripheral tissues from precursors of adrenal origin except for a small contribution from ovarian or adrenal testosterone and androstenedione. Thus, in postmenopausal women, virtually all active sex steroids are made in target tissues by an intracrine mechanism. In elderly men, the intracrine production of androgens is also important; less than 50% of the androgen supply is derived from testicular production.

The high secretion rate of adrenal precursor sex steroids in men and women differs from that in laboratory animal models, in which the secretion of sex steroids occurs exclusively in the gonads. In rats and mice, long-term administration of DHEA prevented obesity, diabetes mellitus, cancer, and heart disease and enhanced immune function.

These experimental animal data have been used to argue that DHEA administration in adult or elderly individuals prolongs life span and might be an “elixir of youth.” Supportive data in humans are few, however, and highly controversial. Epidemiologic studies indeed point to a mild cardioprotective effect of higher DHEAS levels in both men and women. Functional parameters of activities of daily living in men older than 90 years were lowest in those with the lowest serum DHEAS concentrations, and in healthy elderly individuals there was an association between the ratio of cortisol to DHEAS levels and cognitive impairment.
Dehydroepiandrosterone Replacement Therapy

Several randomized placebo-controlled studies demonstrated that oral administration of DHEA was beneficial. Three months of daily treatment with 50 mg of DHEA in 20 healthy adults, most of whom were not elderly, increased serum DHEA and DHEAS concentrations to young adult levels, increased androgen and IGF-I concentrations, and induced a remarkable increase in perceived physical and psychological well-being in both sexes without an effect on libido. In a subsequent study, treatment with 100 mg of DHEA for 6 months in eight adult men and eight adult women increased lean body mass in both sexes but increased muscle strength only in the men. A number of shorter, well-controlled trials with DHEA subsequently did not demonstrate a clinically significant effect on the parameters just mentioned.

After 280 healthy elderly women and men, aged 60 to 79 years, were given DHEA at 50 mg or placebo orally daily for 1 year in a randomized trial, no adverse effects were noted. A significant increase in most parameters related to libido was observed in older women. Improved skin status, including hydration, epidermal thickness, serum production, and pigmentation, was also noted, particularly in women. In women older than 70 years, bone turnover slightly improved as a result of a decrease in osteoclast activity.

A physiologic functional role of DHEA in women has been ascertained in a careful double-blind study. In women with adrenal insufficiency, DHEA administration (50 mg/day) normalized serum concentrations of DHEA, DHEAS, androstenedione, and testosterone. DHEA significantly improved overall well-being as well as scores for depression and anxiety, the frequency of sexual thoughts, sexual interest, and satisfaction with both mental and physical aspects of sexuality.
Conclusions

No prospective long-term randomized studies of DHEA administration have been carried out in frail elderly people or in the elderly individuals with the lowest serum DHEA and DHEAS concentrations. DHEAS is a universal precursor for the peripheral local production and action of estrogens and androgens in elderly people. The data suggest that serum estradiol concentrations (which might be a surrogate marker for tissue concentrations) and serum luteinizing hormone concentrations (which reflect serum androgen and estrogen activity) demonstrate important positive relations with bone mineral density, quality of life, cognitive decline, and degree of atherosclerosis.

The addition of DHEA (50 mg) to the existing large pool of DHEA and DHEAS in unselected elderly individuals has limited clinical effects, especially in elderly women. It is not known whether the increase in sex steroid levels induced by long-term DHEA administration is safe with regard to development of ovarian, prostate, or other types of steroid-dependent cancers. DHEA is currently widely used in the United States as an unapproved treatment against aging. With the scientific verdict still out, without further confirmation of the reported beneficial actions of DHEA in humans, and without a better understanding of its potential risks, it is premature to recommend the routine use of DHEA for delaying or preventing the physiologic consequences of aging.
SOMATOPAUSE

Role of Growth Hormone and Insulin-like Growth Factor I during Aging

Elderly men and women secrete GH less frequently and at lower amplitude than do young people. In fact, GH secretion declines approximately 14% per decade in normal individuals. In parallel, serum levels of IGF-I (see Fig. 25-2) are 20% to 80% lower in healthy elderly individuals than in healthy young adults. The concept that this decline in GH and IGF-I secretion contributes to the decline of functional capacity in elderly people (somatopause) is mainly derived from studies in which GH replacement therapy in GH-deficient adults was shown to increase muscle mass, muscle strength, bone mass, and the quality of life. A beneficial effect on the lipid profile and an important decrease in fat mass were also observed in these patients. As in hypogonadal individuals, adult GH deficiency can thus be considered a model of normal aging because a number of catabolic processes that are central in the biology of aging can be reversed by GH replacement.

Several studies of the relationship between body composition and functional capacity and serum IGF-I concentrations demonstrated contradictory results. In a study of healthy individuals of a broad age range, an association was observed between the maximum aerobic capacity and circulating IGF-I levels. However, no relationship with IGF-I was demonstrated in a group of highly active older people when grip strength, physical performance, and cognitive state were measured.
Growth Hormone Replacement Therapy

Rudman and colleagues, after a ground-breaking randomized controlled trial in healthy men 61 to 81 years old with serum IGF-I concentrations in the lower third for their age, reported in 1990 that GH treatment (30 µg/kg three times weekly for 6 months) restored the men's IGF-I levels to "normal." In the treatment group, lean body mass rose by 8.8% and lumbar vertebral density increased by 1.6%. The magnitudes of these initial changes were equivalent to a reversal of the age-related changes by 10 to 20 years. However, during continuation of this study to 12 months, the significant positive effect on bone mineral density at any site was lost.

In the subsequent years, it became clear that GH administration in healthy elderly individuals frequently caused acute adverse effects such as carpal tunnel syndrome, gynecomastia, fluid retention, and hyperglycemia, which were severe enough for an appreciable number of individuals to drop out of these studies. The most disappointing aspect, however, was that no positive effects of GH administration were observed on muscle strength, maximal oxygen consumption, or functional capacity. (In contrast, when GH was administered in combination with resistance exercise training, a significant positive effect on muscle mass and muscle strength was recorded that did not differ from that seen with placebo treatment, which suggests that GH does not add to the beneficial effects of exercise.)

A representative example of a well-controlled study of GH administration in unselected elderly men is given in Table 25-3.

Earlier studies demonstrated that pharmacologic doses of GH prevent the autocannibalistic effects of acute diseases on muscle mass. This finding prompted us to carry out a randomized, placebo-controlled trial of 6 weeks of GH administration in elderly individuals with an acute hip fracture. Our preliminary results indicate that in patients older than 75 years, GH administration causes an earlier return to independent living after the fracture. Comparable studies are being done in several countries, and confirmation is needed before GH can gain a place in the treatment of acute catabolic states in frail elderly people.

Other components in the regulation of the GH-IGF-I axis

are effective in activating GH and IGF-I secretion. Long-acting derivatives of the hypothalamic peptide growth hormonereleasing hormone (GHRH), given twice daily subcutaneously for 14 days to healthy 70-year-old men, increased GH and IGF-I levels to those encountered in 35-year-olds. These studies suggest that somatopause is driven primarily by the hypothalamus and that pituitary somatotropes retain their capacity to synthesize and secrete high levels of GH.

GH-releasing peptides (GHRPs) are oligopeptides with even more powerful GH-releasing effects. They were originally developed by design. Their effects on GH secretion are mediated through endogenous specific receptors. Nonpeptide analogues (e.g., MK-677 and L-692,429) have powerful GH-releasing effects, restoring IGF-I secretion in older adults to levels typical of young adults. Long-term oral administration of MK-677 to healthy elderly individuals increased lean body mass but not muscle strength. If proven to be GH-specific, these orally active GHRP derivatives might be important alternatives to subcutaneously administered GH for studies of the reversal of somatopause, prevention of frailty, and reversal of acute catabolism.

The long-term safety of activating GH and IGF-I levels in older people has become a concern because of reports of an association between serum IGF-I concentrations and cancer risk. Individuals with high IGF-I levels (or low IGF-binding protein 3 levels) within the broad normal range have an increased risk of prostate, colon, and breast cancer. These epidemiologic studies, together with experimental data, suggest that the IGF-I system is involved in tumor development and progression. However, no causal relationship between IGF-I levels and cancer risk has yet been established, and possible medical intervention directed at increasing IGF-I bioactivity in elderly people will in most instances be given toward the end of life, presumably not allowing enough time to affect tumor development or progression.
Conclusions

During the aging process, GHIGF-I axis activity declines. It is unclear whether changes in body composition and functional capacity are directly related. GH administration in older adults causes an increase in lean body mass and an appreciable loss of fat mass. However, the very limited ability of GH treatment to improve muscle strength and functional capacity in elderly people, despite restoration of circulating IGF-I concentrations to young adult levels, limits its application. Furthermore, most dose regimens of GH cause appreciable adverse effects, and long-term safety with regard to tumor development and progression remains uncertain. GHRP and its orally active analogues are capable of restoring GH and IGF-I levels in the elderly population.

In the near future, clinical trials with such orally active molecules in frail elderly people or in elderly individuals with clearly lowered IGF-I levels, or both, should be able to delineate the precise role of the GHIGF-I axis in the aging process. In such trials, much emphasis must be given to safety aspects. At present, there is insufficient evidence to recommend medical intervention in the GHIGF-I axis to rejuvenate healthy elderly people. Only elderly patients with GH deficiency caused by organic diseases, such as pituitary adenomas, clearly benefit from GH replacement therapy.
THE CONCEPT OF SUCCESSFUL AGING

There is considerable variation in the effects of aging on healthy individuals, with some people exhibiting greater and others evidencing few or no age-related alterations in physiologic functions. It has been suggested that it might be useful to distinguish between usual and successful patterns of aging. Genetic factors, lifestyle, and societal investments in a safe and healthful environment are important aspects of successful aging.

Traditionally, the aging process, including the development of physical frailty toward the end of life, has been considered physiologic and unavoidable. It has recently become evident, however, that it might not be necessary to accept the grim stereotype of aging as an unalterable process of decline and loss. As life expectancy rises further in the coming decades, the overarching goal should be "an increase in years of healthy life with a full range of functional capacity at each stage of life." Such a compression of morbidity can be achieved by adapting lifestyle measures, but a number of aspects of the aging process of the endocrine system invite the development of routine medical intervention programs offering long-term replacement therapy with one or more hormones in order to delay the aging process and to allow humans to live for a longer period in a relatively intact state.

### TABLE 25-3 – Effects of Growth Hormone Administration (30 µg/kg three times a week) for 6 Months in 52 Healthy Men (69 years) with Well-Preserved Functional Ability but Low Levels of Insulin-like Growth Factor 1

<table>
<thead>
<tr>
<th>Mean Change in Variable</th>
<th>GH (n = 26)</th>
<th>Placebo (n = 26)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>IGFI (ng/mL)</td>
<td>119.2</td>
<td>7.6</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td><strong>Body Weight and Composition</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weight, kg</td>
<td>0.5</td>
<td>1.0</td>
<td>&gt; .2</td>
</tr>
<tr>
<td>Lean mass, %</td>
<td>4.3</td>
<td>- 0.1</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Fat mass, %</td>
<td>- 13.1</td>
<td>- 0.3</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Bone mineral content, %</td>
<td>0.9</td>
<td>- 0.1</td>
<td>.05</td>
</tr>
<tr>
<td>Skin thickness, %</td>
<td>13.4</td>
<td>1.1</td>
<td>.09</td>
</tr>
<tr>
<td><strong>Muscle Strength, %</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knee extension</td>
<td>3.8</td>
<td>1.3</td>
<td>&gt; .2</td>
</tr>
<tr>
<td>Knee flexion</td>
<td>10.0</td>
<td>8.2</td>
<td>&gt; .2</td>
</tr>
<tr>
<td>Hand grip</td>
<td>- 1.5</td>
<td>3.8</td>
<td>.11</td>
</tr>
<tr>
<td>Maximum oxygen consumption, %</td>
<td>2.5</td>
<td>- 2.0</td>
<td>&gt; .2</td>
</tr>
</tbody>
</table>

GH, growth hormone; IGFI, insulin-like growth factor I.

References


34. Kanis JA. Estrogens, the menopause, and osteoporosis. Bone 1996; 19(Suppl):185S190S.


41. Willett WC, Colditz G, Stampfer M. Postmenopausal estrogen suppression, unopposed, or none of the above. JAMA 2000; 283:534535.

42. Stevenson JC. Do we need different galenic forms of estrogens and progestogens? Menopause 1996; 3:24-32.


1300


Intracellular magnesium, like phosphate, is necessary for a wide range of cellular functions. It is an essential cofactor in enzymatic reactions, including most of the gradient, is unknown, although some evidence for regulated channels has been obtained.

The concentration of which is approximately 5 mM. The intracellular cytosolic free magnesium concentration is approximately 0.5 mM—that is, 1000-fold higher than that of phosphate or other anions, and 55% is present as the free ion.

Magnesium is the fourth most abundant cation in the body. Roughly half is found in bone and half in muscle and other soft tissues. As much as half of the magnesium that disorders of phosphate homeostasis associated with severe depletion of intracellular phosphate lead to profound and global impairment of organ function.

Roles of the mineral ions
Calcium (Ca) and phosphorus (P) are the principal constituents of bone, and together they constitute 65% of its weight. Bone, in turn, contains nearly all of the calcium and phosphorus and over half of the magnesium (Mg) in the human body. Although quantitatively minor in amount, each of these ions in the extracellular fluid and within cells plays a crucial role in normal physiologic processes.

Ninety-nine percent of total body calcium resides in bone, mainly albumin and globulins. The ionized calcium concentration in serum is approximately 1.2 mM (5 mg/dL), and it is this fraction that is biologically active and that is tightly controlled by hormonal mechanisms. Because intracellular cytosolic free calcium concentrations typically are in the range of only 100 nM, a very large chemical gradient (i.e., 10,000:1), augmented by the large negative electrical potential, favors calcium entry into cells through calcium channels. This gradient is maintained by the limited conductance of resting calcium channels and by the energy-dependent extrusion of calcium into the extracellular fluid via high-affinity Ca\(^{2+}\)-ATPases and low-affinity sodium-calcium (Na\(^{+}\)-Ca\(^{2+}\) ) exchangers.

Phosphate is more widely distributed to nonosseous tissues than is calcium. Eighty-five percent of body phosphate is in the mineral phase of bone, and the remainder is located in inorganic or organic form throughout the extracellular and intracellular compartments. In human serum, inorganic phosphate (P\(_{i}\) ) is present at a concentration of approximately 1 mM and exists almost entirely in ionized form as either H\(_2\)PO\(_4\)\(^-\) or HPO\(_4\)\(^{2-}\). Only 12% of serum phosphate is protein-bound, and an additional small fraction is looselycomplexed with calcium, magnesium, and other cations. Intracellular free phosphate concentrations are generally comparable to those in the extracellular fluid (i.e., 1 to 2 mM), although the inside-negative electrical potential of the cell creates a significant energy requirement for translocation of phosphate into cells. This process generally is accomplished through sodium-phosphate cotransport driven by the transmembrane sodium gradient. A number of sodium-phosphate cotransporters have been cloned; various cells and tissues employ different species of such transporters with distinctive regulatory characteristics.

Organic phosphate is a key component of virtually all classes of structural, informational, and effector molecules that are essential for normal genetic, development, and physiologic processes. Phosphate is an integral constituent of nucleic acids; phospholipids; complex carbohydrates; glycolytic intermediates; structural, signaling, and enzymatic phosphoproteins; and nucleotide cofactors for enzymes and G proteins. Of particular importance are the high-energy phosphate ester bonds present in molecules such as adenosine triphosphate (ATP), diphosphoglycerate, and creatine phosphate that store chemical energy. Phosphate plays a particularly prominent role as the key substrate or recognition site in numerous kinase and phosphatase regulatory cascades. Cytosolic phosphate also directly regulates a number of crucial intracellular reactions, including those involved in glucose transport, lactate production, and synthesis of ATP. In light of these diverse roles, it is not surprising that disorders of phosphate homeostasis associated with severe depletion of intracellular phosphate lead to profound and global impairment of organ function.

Magnesium is the fourth most abundant cation in the body. Roughly half is found in bone and half in muscle and other soft tissues. As much as half of the magnesium in bone is not sequestered in the mineral phase but is freely exchangeable with ions in the extracellular fluid and therefore may serve as a buffer against changes in extracellular magnesium concentration. Less than 1% of all magnesium in the body is present in the extracellular fluid, where the magnesium concentration is approximately 0.5 mM. The concentration of magnesium in serum normally is 0.7 to 1.0 mM, of which roughly a third is protein bound, and an additional small fraction is loosely complexed with calcium, magnesium, and other cations. Intracellular free phosphate concentrations are generally comparable to those in the extracellular fluid (i.e., 1 to 2 mM), although the inside-negative electrical potential of the cell creates a significant energy requirement for translocation of phosphate into cells. This process generally is accomplished through sodium-phosphate cotransport driven by the transmembrane sodium gradient. A number of sodium-phosphate cotransporters have been cloned; various cells and tissues employ different species of such transporters with distinctive regulatory characteristics.

Organic phosphate is a key component of virtually all classes of structural, informational, and effector molecules that are essential for normal genetic, developmental, and physiologic processes. Phosphate is an integral constituent of nucleic acids; phospholipids; complex carbohydrates; glycolytic intermediates; structural, signaling, and enzymatic phosphoproteins; and nucleotide cofactors for enzymes and G proteins. Of particular importance are the high-energy phosphate ester bonds present in molecules such as adenosine triphosphate (ATP), diphosphoglycerate, and creatine phosphate that store chemical energy. Phosphate plays a particularly prominent role as the key substrate or recognition site in numerous kinase and phosphatase regulatory cascades. Cytosolic phosphate also directly regulates a number of crucial intracellular reactions, including those involved in glucose transport, lactate production, and synthesis of ATP. In light of these diverse roles, it is not surprising that disorders of phosphate homeostasis associated with severe depletion of intracellular phosphate lead to profound and global impairment of organ function.

Magnesium is the fourth most abundant cation in the body. Roughly half is found in bone and half in muscle and other soft tissues. As much as half of the magnesium in bone is not sequestered in the mineral phase but is freely exchangeable with ions in the extracellular fluid and therefore may serve as a buffer against changes in extracellular magnesium concentration. Less than 1% of all magnesium in the body is present in the extracellular fluid, where the magnesium concentration is approximately 0.5 mM. The concentration of magnesium in serum normally is 0.7 to 1.0 mM, of which roughly a third is protein bound, and 15% is loosely complexed with phosphate or other anions, and 55% is present as the free ion. Over 95% of intracellular magnesium is bound to other molecules, most notably ATP, the concentration of which is 5 mM. The intracellular cytosolic free magnesium concentration is approximately 0.5 mM, which is 1000-fold higher than that of calcium and is maintained by an active sodium-magnesium antiporter. The mechanism whereby magnesium enters cells, presumably down a favorable electrochemical gradient, is unknown, although some evidence for regulated channels has been obtained.

Intracellular magnesium, like phosphate, is necessary for a wide range of cellular functions. It is an essential cofactor in enzymatic reactions, including most of the same glycolytic, kinase, and phosphatase pathways that also involve phosphate. Magnesium serves to directly stabilize the structures of a variety of macromolecules
and complexes, including deoxyribonucleic acid (DNA), ribonucleic acid (RNA), and ribosomes; is a key activator of the many ATPase-coupled ion transporters; and plays a direct role in mitochondrial oxidative metabolism. As a result, magnesium is critical for energy metabolism and the maintenance of a normal intracellular environment. Extracellular magnesium is crucial for normal neuromuscular excitability and nerve conduction, and many of the clinical consequences of magnesium deficiency or excess reflect abnormalities in this sphere.

The importance of the mineral ions for normal cellular physiology as well as skeletal integrity is reflected in the powerful endocrine control mechanisms that have evolved to maintain their extracellular concentrations within relatively narrow limits. The following topics describe the structures, secretory controls, actions, and interactions of parathyroid hormone, calcitonin, and 1,25-dihydroxyvitamin D (1,25(OH)\(_2\) D\(_3\) or calcitriol) the major hormones involved in mineral ion homeostasis. Subsequent topics cover the wide variety of clinical disorders that accompany abnormalities in this hormonal network.
PARATHYROID HORMONE

Parathyroid hormone (PTH) is the peptide hormone that controls the minute-to-minute level of ionized calcium in the blood and extracellular fluids. PTH binds to cell surface receptors in bone and kidney, thereby triggering responses that increase blood calcium. PTH also increases renal synthesis of 1,25(OH)₂D₃, the hormonally active form of vitamin D, which then acts on the intestine to augment absorption of dietary calcium, in addition to promoting calcium fluxes into blood from bone and kidney. The resulting increase in blood calcium (and in 1,25(OH)₂D₃) feeds back on the parathyroid glands to decrease the secretion of PTH. The parathyroid glands, bones, kidney, and gut are thus the crucial organs that participate in PTH-mediated calcium homeostasis.

Parathyroid Gland Biology

Parathyroid chief cells have three properties vital to their homeostatic function:

1. They rapidly secrete stored hormone in response to changes in blood calcium.
2. They can synthesize, process, and store large amounts of PTH in a regulated manner.
3. They replicate when chronically stimulated.

These functional attributes allow for short-term, intermediate-term, and long-term adaptation, respectively, to changes in calcium availability.

Parathyroid Hormone Biosynthesis

PTH, a protein of 84 amino acids, is synthesized as a larger precursor, pre-proparathyroid hormone (pre-pro-PTH). Figure 26-3 illustrates the reported pre-pro-PTH sequences. These pre-pro-PTH sequences share a 25-residue "pre" or signal sequence and a 6-residue "pro" sequence. The signal sequence, along with the short pro sequence, functions to direct the protein into the secretory pathway (Fig. 26-4). During transit across the membrane of the endoplasmic reticulum, the signal sequence is cleaved off and rapidly degraded. The importance of the signal sequence for normal secretion of PTH is illustrated by the hypoparathyroidism inherited in families carrying mutations in the signal sequence of pre-pro-PTH.

The role of the short pro sequence is not completely understood; it may help the signal sequence work efficiently and ensure accurate cleavage of the precursor. After cleavage of the pro sequence, the mature PTH(184) is concentrated in secretory vesicles and granules. One morphologically distinct subtype of granule contains both PTH and the proteases cathepsin B and cathepsin H. This co-localization of proteases and PTH in secretory granules probably explains the observation that a portion of the PTH secreted from parathyroid glands consists of carboxy-terminal PTH fragments. No amino-terminal fragments of PTH are secreted. Although the possible functions of carboxy-terminal fragments of PTH are still poorly characterized, these fragments do not activate the PTH/PTHrP receptor (see later). Thus, the intracellular fragmentation of PTH probably represents an inactivating pathway. The intracellular degradation of newly synthesized PTH provides an important regulatory mechanism. Under conditions of hypercalcemia, the secretion of PTH is substantially decreased, and most of what is secreted consists of carboxy-terminal fragments.

Parathyroid Hormone Secretion

Although catecholamines, magnesium, and other stimuli can affect PTH secretion, the major regulator of PTH secretion is the concentration of ionized calcium in blood. Increased serum ionized calcium leads to a decrease in PTH secretion. The shape of the dose-response curve is sigmoid. Properties of the parathyroid cell determine the conformation of the sigmoid curve but do not alone determine the point on the curve that represents a physiologic steady state for an individual. This point, usually between the mid-point and the bottom of the curve, is determined by how vigorously target organs respond to PTH. Figure 26-5C (solid line) shows how an individual's calcium level rises in response to increases in PTH; the parathyroid gland's sigmoid curve is the dotted line. In the steady state, an individual's blood levels of PTH and calcium represent the intersection of the two lines.

The sigmoid curve reveals several important physiologic properties of the parathyroid gland. The minimal secretory...
For decades it has been known that phosphate elevation stimulates PTH secretion, although the mechanism by which it elevates PTH is uncertain; differing experimental paradigms suggest regulation at the levels of gene transcription, post-translational processing of peptide, and mRNA stability, and mRNA translation. A recent study has shown that acute hypocalcemia in rats leads, within an hour, to an increase in PTH messenger RNA (mRNA).

Calcium also regulates the biosynthesis of PTH. In vivo 1,25(OH) \_2 \text{D}_3 suppression of transcription does not occur when 1,25(OH) \_2 \text{D}_3 level of intracellular calcium despite fluctuations in extracellular calcium. The parathyroid cell is exceptional, in that modest changes in extracellular calcium lead to the calcium-sensing receptor gene also have the expected defects in parathyroid calcium sensing. Whereas activating mutations cause familial hypoparathyroidism with hypercalcemia. Inactivating mutations cause familial hypocalciuric hypercalcemia (FHH), a disease of defective calcium sensing (see later on), whereas activating mutations cause familial hyperparathyroidism with hypercalcemia. Furthermore, mice genetically engineered to have only one functioning copy of the calcium-sensing receptor gene also have the expected defects in parathyroid calcium sensing. Of importance, calcimimetic compounds that activate the cloned calcium-sensing receptor have been shown to inhibit PTH secretion in humans and may prove useful in the treatment of primary and secondary hyperparathyroidism in the future.

The signaling properties of the calcium-sensing receptor explain one of the most unusual features of the parathyroid cell. Most cells maintain a constant and very low level of intracellular calcium despite fluctuations in extracellular calcium. The parathyroid cell is exceptional, in that modest changes in extracellular calcium lead to corresponding changes in intracellular calcium. When calcium activates the cell surface calcium-sensing receptor, intracellular calcium rises because of release of calcium from intracellular stores and opening of plasma membrane calcium channels. This increase in intracellular calcium then leads to a decrease in PTH secretion by mechanisms that remain to be clarified.

The calcium-sensing receptor is expressed widely. Expression in the renal tubules and calcitonin-producing cells of the thyroid contributes to calcium homeostasis, whereas expression in organs such as the brain points to multiple roles for calcium signaling. The observation that the calcium-sensing receptor also responds to certain amino acids suggests that the expression of the calcium-sensing receptor in the gut and other sites may facilitate the assimilation of multiple nutrients.

The minute-to-minute regulation of PTH blood levels can be explained by the two mechanisms already discussedregulation of PTH secretion by the calcium-sensing receptor and amplification of this regulation by intracellular degradation of stored hormone. Over a longer time frame, the parathyroid cell regulates the expression of the PTH gene as well.

Although 1,25(OH) \_2 \text{D}_3, the active form of vitamin D, has no direct effect on PTH secretion, it dramatically suppresses PTH gene transcription. This suppression of transcription does not occur when 1,25(OH) \_2 \text{D}_3 is administered to chronically hypocalcemic animals, however, perhaps because hypocalcemia leads to a fall in parathyroid vitamin D receptors or because hypocalcemia increases the expression of calreticulin in the parathyroid. The capability of hypocalcemia to override the effects of high levels of 1,25(OH) \_2 \text{D}_3 represents an important defense, because it provides a way to make large amounts of PTH and 1,25(OH) \_2 \text{D}_3 at the same time, when both are needed.

Calcium also regulates the biosynthesis of PTH. In vivo studies show that acute hypocalcemia in rats leads, within an hour, to an increase in PTH messenger RNA (mRNA). In contrast, hypercalcemia leads to little or no change in PTH mRNA. Thus, under normal conditions, the inhibition by calcium of PTH biosynthesis already is nearly maximal, just as it is for PTH secretion. When the parathyroid gland is poised to respond to a fall in calcium much more readily than to a rise. The mechanism for the increase in PTH mRNA in response to hypocalcemia is uncertain; differing experimental paradigms suggest regulation at the levels of gene transcription, mRNA stability, and mRNA translation.

For decades it has been known that phosphate elevation stimulates PTH secretion, largely by lowering blood calcium and 1,25(OH) \_2 \text{D}_3 levels. More recently, a series of studies in vitro and in vivo have demonstrated that phosphate can increase PTH secretion directly, independent of effects on blood calcium and 1,25(OH) \_2 \text{D}_3. Phosphate increases PTH secretion acutely only after a delay and probably works largely through regulation of PTH mRNA levels. The direct effects of phosphate may be important at very high and very low levels, although these effects may be important in the setting of renal failure.
The regulation of the PTH gene has particular clinical relevance in patients with renal failure. Hypocalcemia, low levels of 1,25(OH)₂D₃, hyperphosphatemia, and, possibly, uremic toxins disrupt normal calcium homeostasis in this setting. Therapy with 1,25(OH)₂D₃ and calcium increases calcium absorption and also inhibits PTH synthesis by direct effects on the parathyroid gland. Prevention of hyperphosphatemia avoids the direct and indirect actions of phosphate to stimulate PTH secretion.

Regulation of Parathyroid Cell Number

Parathyroid cells divide during the growth of young animals but replicate little in adulthood. Parathyroid cell number can dramatically increase, however, in the setting of hypocalcemia, low levels of 1,25(OH)₂D₃, hyperphosphatemia, or uremia and during neoplastic growth.

Calcium, acting through the parathyroid calcium-sensing receptor, restrains parathyroid proliferation. This effect has been demonstrated clinically in patients who lack both copies of the calcium-sensing receptor gene. These neonates exhibit severe primary hyperparathyroidism with large, diffusely hyperplastic glands that presumably have developed because of insufficient activation of the parathyroid calcium-sensing receptor by extracellular calcium. Furthermore, administration of the calcimimetic compound NPS R-568, which activates the calcium-sensing receptor directly, prevents parathyroid cell proliferation in experimental uremia.

The role of 1,25(OH)₂D₃, independent of blood calcium, in regulating parathyroid cell proliferation is less well established than that of calcium. That 1,25(OH)₂D₃ can dramatically affect parathyroid cell number has been shown in vivo in many settings, but such studies cannot rigorously eliminate effects of transient changes in blood calcium. The suppression of proliferation of cultured parathyroid cells by 1,25(OH)₂D₃ certainly suggests that 1,25(OH)₂D₃ can directly inhibit parathyroid cell replication. Nevertheless, modest changes in 1,25(OH)₂D₃ have no effect on parathyroid cell proliferation in vivo, and calcium alone can prevent parathyroid cell hyperplasia in mice engineered to lack vitamin D receptors. Thus, the importance of regulation of parathyroid cell number by 1,25(OH)₂D₃ in vivo has not been established.

Although the ability to increase parathyroid cell number in response to physiologic challenge represents an important defense against hypocalcemia, it is a slow response that is not easily reversible. When the need for an increased number of parathyroid cells disappears (e.g., after renal transplantation for uremia), persistent hyperparathyroidism can cause vexing clinical problems for months and years thereafter. The mechanisms for decreasing parathyroid cell number, if they exist, are poorly understood. Apoptosis of normal parathyroid cells in response to experimental manipulation has not been demonstrated.
Parathyroid Gland Development

Genes involved in making parathyroid cells during development may also regulate PTH synthesis and parathyroid cell number throughout life; thus, an understanding of parathyroid cell development may have broad clinical implications. Although the genetic mechanisms used to generate parathyroid chief cells during development are largely unknown, the importance of four specific genes has become clear. Studies of gene knockout mice have shown that the hoxa3 and pax9 transcription factors are needed to form parathyroid glands as well as many other pharyngeal pouch derivatives, such as the thymus. People with mutations in the gene encoding the transcription factor GATA3 exhibit a syndrome of hypoparathyroidism, sensorineural deafness, and renal anomalies when only one copy of the gene is mutated. Furthermore, mice or humans missing the gcm2 and GCMB genes, respectively, have no parathyroid glands. In both species, the deletion of gcm2 or GCMB (the human equivalent) is very specific for controlling parathyroid development because no abnormalities in other tissues have been noted. Mice, which have only two parathyroid glands normally, still make PTH in a small number of cells in the thymus after gcm2 gene ablation and secrete this PTH into the circulation. A human patient without GCMB had no detectable circulating PTH at birth and low levels of PTH several years later.
Peripheral Metabolism of Parathyroid Hormone

The earliest radioimmunoassays for PTH demonstrated that the molecular forms of PTH in the circulation differ from those in the parathyroid gland. Characterization of the metabolism of PTH and its fragments has clarified the origins and significance of immunoreactive PTH molecules in the blood stream. As noted previously, both PTH(184) and carboxy-terminal fragments of PTH are secreted from the parathyroid gland; the ratio of inactive PTH to active PTH secretion increases with increasing blood calcium. Secreted intact PTH(184) is extensively metabolized by liver (70%) and kidney (20%) and disappears from the circulation with a half-life of 2 minutes. This rapid peripheral metabolism of PTH is unaffected by widely varying levels of blood calcium or 1,25(OH)\(_2\)D\(_3\). Less than 1% of the secreted hormone finds its way to PTH receptors on physiologic target organs. These features of PTH metabolism ensure that the blood level of PTH is determined principally by the activity of the parathyroid glands and that the PTH level can respond rapidly to small changes in the rate of secretion of the hormone.

In the liver, a small amount of PTH binds to physiologically relevant PTH receptors but most of the intact PTH is cleaved, initially after residues 33 and 36, probably by cathepsins. In the kidney, a small amount of intact PTH binds to physiologic PTH receptors, but most of the intact PTH is filtered at the glomerulus and subsequently bound by a large, membrane-bound luminal protein, megalin; this binding leads to internalization and degradation of PTH by the tubules. Carboxy-terminal fragments are also cleared efficiently by glomerular filtration. In fact, the kidney is the only known site of clearance of carboxy-terminal PTH fragments; these fragments thus accumulate dramatically when the glomerular filtration rate (GFR) falls. Even in the presence of normal renal function, the half-life of carboxy-terminal fragments of PTH exceeds that of PTH(184) by several-fold. Consequently, the concentration of carboxy-terminal fragments in the circulation exceeds that of intact PTH, even though intact PTH usually is the major form of PTH secreted from the parathyroid gland.

Careful analysis of PTH fragments using high-performance liquid chromatography (HPLC) and immunologic methods have revealed almost full-length PTH fragments missing the first several amino acids of the hormone, but containing most or all of the remaining hormone sequence. These still incompletely characterized fragments are both secreted from the parathyroid gland and generated by peripheral metabolism of the hormone. Because they are missing the amino-terminal portion of PTH, they cannot stimulate cyclic adenosine monophosphate (cAMP) production by the PTH/parathyroid hormone-related protein (PTHrP) receptor, and except in renal failure, they circulate in small amounts. Nevertheless, the possible biologic activity of these and other PTH fragments, possibly through novel receptors, remains an unsettled issue. Experiments with PTH(784) suggest that such extended carboxyl fragments may exert potent effects in vivo, antagonistic to those of intact PTH.

---

Actions of Parathyroid Hormone

Actions on the Kidney
Stimulation of Calcium Reabsorption

Almost all of the calcium in the initial glomerular filtrate is reabsorbed by the renal tubules. Sixty-five percent is reabsorbed by the proximal convoluted and straight tubules via a passive, paracellular route. Changes in the transepithelial voltage gradient, determined largely by the rate of sodium reabsorption, control the rate of calcium transport in the proximal tubule; PTH does little to affect calcium flux in this region. The remaining calcium is largely reabsorbed more distally20% of the initial filtrate in the cortical thick ascending limb (cTAL) of Henle's loop and 10% in the distal convoluted and connecting tubules. In the cTAL, calcium reabsorption also is mainly passive and paracellular, although some transcellular, active calcium transport may occur as well.

Efficient paracellular and magnesium movement requires expression of a unique tight junction protein, paracellin-1; mutant paracellin-1 genes underlie a rare renal calcium- and magnesium-wasting disorder. Because paracellular calcium transport in the cTAL is driven by the lumen-positive transepithelial voltage gradient that is established by active Na-K-CaCl\textsubscript{2} reabsorption, calcium reabsorption there is strongly inhibited by loop diuretics such as furosemide. The calcium-sensing receptor, initially characterized in the parathyroid, also is expressed in the cTAL. When activated by high blood calcium or magnesium, this receptor inhibits Na-K-CaCl\textsubscript{2} reabsorption in the cTAL. Consequently, renal concentrating capacity and calcium reabsorption are inhibited as well. This inhibition provides a parathyroid-independent mechanism for controlling renal calcium handling in direct response to changes in blood calcium concentration.

Although PTH modestly stimulates paracellular calcium reabsorption in the cTAL, the primary site for hormonal regulation of renal calcium reabsorption is the distal nephron, which normally reabsorbs nearly all of the remaining 10% of filtered calcium by a unique transcellular active transport mechanism. As depicted in the figure, this process is driven by a transepithelial voltage gradient that is generated by the 

\[
\text{Na}^+\text{Cl}^-\text{K}^+\text{ATPase}\text{,}\quad \text{Na}^+\text{K}^+\text{Cl}^-\text{cotransporter}\text{,}\quad \text{Na}^+\text{HCO}_3^-\text{cotransporter}\text{,}\quad \text{Na}^+\text{HCO}_3^-\text{K}^+\text{cotransporter}\text{,}\quad \text{Na}^+\text{K}^+\text{Cl}^-\text{cotransporter}\text{.}
\]

Inhibition of Phosphate Transport

Phosphate reabsorption occurs mainly in the proximal renal tubules, which reclaim roughly 80% of the filtered load. Some additional phosphate (8% to 10%) is reabsorbed in the distal tubule (but not in Henle's loop), leaving about 10% to 12% for excretion in the urine. The normal overall fractional tubular reabsorption of phosphate (TRP), therefore, is about 88%, although a more reliable measure of renal phosphate handling is the phosphate threshold (TmP/GFR), which can be derived from the TRP through the use of a nomogram (Fig. 26-8) based on studies of experimental phosphate infusions in healthy persons and in patients with a variety of diseases that affect phosphate excretion.

\[
\text{Figure 26-8} \quad \text{Nomogram for determining renal threshold phosphate concentration (TmP/GFR) from the plasma phosphate concentration and the fractional reabsorption of filtered phosphate (TRP) or fractional excretion of filtered phosphate (1TRP, or C_{\text{urea}}/C_{\text{creatinine}}}). \quad \text{Because the blood level of phosphate influences the renal handling of phosphate, the renal threshold phosphate concentration best separates normal from abnormal renal phosphate handling. C_{\text{ur}}: clearance; creat, creatinine; GFR, glomerular filtration rate; TRP, tubular resorption of phosphate. (From Walton RJ, Bijvoet OLM: Nomogram of derivation of renal threshold phosphate concentration. Lancet 1979; 2:309710.}
\]

Dietary intake of phosphate also reciprocally regulates the expression and activity of NaP\textsubscript{2} cotransporters and thus the proximal tubular absorption of phosphate by a mechanism that is independent of PTH. Dietary deprivation of phosphate, for example, leads to a stimulation of phosphate reabsorption that can override the effects of PTH on the proximal tubule. Although not yet demonstrated experimentally, a likely candidate to mediate this effect of dietary phosphate is the putative hormone "phosphatonin," excessive action of which is implicated in the humoral hypophosphatemic syndromes X-linked hypophosphatemic rickets and oncogenic

\[
\text{Figure 26-7} \quad \text{Effects of parathyroid hormone (PTH) on distal tubular calcium transport. PTH acts to increase chloride flux through channels in the basolateral membrane. As indicated, this increase leads to calcium influx through apical calcium channels. PTH also increases basolateral Ca\textsuperscript{2+}-Na\textsuperscript{+} exchange. The apical Na\textsuperscript{+}/Ca\textsuperscript{2+} cotransporter allows chloride to enter the cell and is the target of thiazide diuretics. Cl\textsuperscript{-}, intracellular chloride. (Adapted from Friedman PA, Giesek FA: Calcium transport in renal epithelial cells. Am J Physiol 1993; 264:F181F198.)}
\]

increases this conductance (2) by increasing the transmembrane voltage gradient (i.e., by hyperpolarizing the cell), probably by increasing chloride exit through basolateral channel (1).

A similar hyperpolarizing effect is exerted by thiazide-type diuretics, which reduce intracellular chloride (by inhibiting apical sodium chloride [NaCl] transporters) and thereby also open apical calcium channels, increasing calcium entry. To protect the low physiologic level of cytosolic free calcium from the relatively large amounts of incoming calcium, these cells express the calcium-binding protein calbindin-D28K, which avidly binds calcium at the apical membrane and transports it to the basolateral membrane, where it is then ejected via active processes involving sodium-calcium exchange and an ATP-driven calcium pump. Expression of calbindin-D28K is increased by PTH directly and also indirectly, via increased synthesis of 1,25(OH)\textsubscript{2}D\textsubscript{3}.

\[
\text{Figure 26-7} \quad \text{Nomogram for determining renal threshold phosphate concentration (TmP/GFR) from the plasma phosphate concentration and the fractional reabsorption of filtered phosphate (TRP) or fractional excretion of filtered phosphate (1TRP, or C_{\text{urea}}/C_{\text{creatinine}}}). \quad \text{Because the blood level of phosphate influences the renal handling of phosphate, the renal threshold phosphate concentration best separates normal from abnormal renal phosphate handling. C_{\text{ur}}: clearance; creat, creatinine; GFR, glomerular filtration rate; TRP, tubular resorption of phosphate. (From Walton RJ, Bijvoet OLM: Nomogram of derivation of renal threshold phosphate concentration. Lancet 1979; 2:309710.)}
\]

Because paracellular cation transport in the cTAL is driven by the lumen-positive transepithelial voltage gradient determined largely by the rate of sodium reabsorption, the rate of distal tubular calcium reabsorption is increased by PTH, the total amount of calcium in the final urine is likely to be high, because of the high initial filtered load.
Osteomalacia (see "Disorders of Phosphate Metabolism," later).

Other Renal Effects of Parathyroid Hormone

PTH stimulates the synthesis of 1,25(OH)2D3 in the proximal tubule by rapidly inducing transcription of the 25-hydroxyvitamin D 25(OH)D3, 1-hydroxylase gene, an effect that can be overridden by hypercalcemia or by the action of 1,25(OH)2D3. 12 PTH inhibits proximal tubular transcription of the 25(OH)D3-24-hydroxylase gene and antagonizes the up-regulation of 24-hydroxylase activity by 1,25(OH)2D3. 12 (see "Vitamin D Metabolism"). PTH inhibits proximal tubular sodium, water, and bicarbonate reabsorption, 12 mainly via inhibition of the apical amiloride-sensitive Na+-H+ exchanger (NHE3) and the basolateral Na+-K+-ATPase. 12 PTH also stimulates proximal tubular gluconeogenesis 12 and acts directly on glomerular podocytes to decrease both single-nephron and whole-kidney GFR. 12

**Actions of Parathyroid Hormone on Bone**

The actions of PTH on bone are complicated because PTH acts on a number of cell types both directly and indirectly. PTH increases both bone formation and bone resorption. With regard to calcium homeostasis, the effect of PTH on bone resorption is dominant; continuous administration of PTH leads to a net release of calcium from bone. But the straightforward net effect of PTH on calcium homeostasis belies the highly variable effects of the hormone on bone, which depend on the type of bone (trabecular or cortical), the particular target cell type, and the pattern of PTH administration.

**Figure 26-9** illustrates the cells of the osteoblast lineage (see also Chapter 27). Osteoblasts are derived from multipotent mesenchymal stem cells that can differentiate into chondrocytes, adipocytes, osteoblasts, and probably other cell types. 12 Within the osteoblast lineage, committed osteoprogenitor cells divide, become preosteoblastic stromal cells (which can divide further), and then become osteoblasts. Osteoblasts no longer divide and are cuboidal cells found on the bone surface actively laying down new bone. When these cells become surrounded by bone, they become stellate osteocytes. If, instead, osteoblasts stop synthesizing matrix and remain on the bone surface, they flatten out as bone lining cells. Not all preosteoblasts and osteoblasts mature; a variable number die by apoptotic, programmed cell death. 12

Receptors for PTH are found on preosteoblasts, osteoblasts, lining cells, and osteocytes. PTH changes the osteoblast lineage cell population by stimulating cell proliferation 12 by decreasing apoptosis of preosteoblasts and osteoblasts, thereby increasing the number of osteocytes; and perhaps by converting inactive lining cells to osteoblasts. 12 When added to cells in culture, PTH stops preosteoblastic cells from becoming mature osteoblasts. 12 Furthermore, PTH changes the activity of mature osteoblasts by a variety of mechanisms. When PTH is added to isolated osteoblasts in vitro, the osteoblasts decrease their synthesis of collagen I and other matrix proteins and, at least in part by steering the essential transcription factor core-binding factor-1 (CBFA1) toward proteosomal destruction. 12 In vivo, however, the most obvious effects of PTH are to increase bone formation, probably by indirect actions such as stimulation of synthesis of insulin-like growth factor I (IGF1) and other growth factors by osteoblast lineage cells. This action of PTH can be mimicked in vivo by intermittent administration of PTH to osteoblasts in organ culture 12 or as dispersed cells. 12

Surprisingly, osteoblasts, the bone-resorbing cells derived from hematopoietic precursors, have no PTH receptors. Instead, preosteoblasts and osteoblasts signal to osteoclast precursors to cause them to fuse and form mature osteoclasts and signal to those osteoclasts to allow them to resorb bone and to avoid apoptosis. 12, 26-10 Two osteoblast surface proteins, macrophage colony-stimulating factor (M-CSF) and RANK ligand (RANKL), are essential for stimulation of osteoclast differentiation and osteoclastogenesis. 12, 26-10 RANKL is essential for the activation of mature osteoclasts. The growth factor M-CSF (or CSF1), is expressed both as a secreted protein and as a cell surface protein; the production of both forms is stimulated by PTH. 12 RANKL also

**Figure 26-9** Osteoblast lineage. All precursors of osteoblasts can proliferate; osteoblasts are transformed to osteocytes and lining cells without further proliferation. Some data suggest that lining cells may revert to osteoblast function after parathyroid hormone stimulation. At each stage in the lineage, apoptotic cell death is probably an alternative fate.

named osteoprotegerin ligand (OPGL), osteoclast-differentiating factor (ODF), and TRANCEs a membrane-bound member of the tumor necrosis factor (TNF) family; its synthesis is also increased by PTH. RANKL binds to its receptor, RANK, a member of the TNF receptor family. RANK is found on both osteoclast precursors and on mature osteoclasts. The binding of RANKL to RANK can be blocked by osteoprotegerin (OPG), another member of the TNF receptor family. OPG (also called OCIF and TR1) circulates and is also secreted by cells of the osteoclastic lineage. PTH decreases the synthesis and secretion of OPG from these cells. 12 Thus, PTH, by increasing RANK and decreasing OPG locally in bone, serves to increase bone resorption.

Because PTH can both increase bone formation and increase resorption, the net effect of PTH on bone mass varies from one part of bone to another and also varies dramatically according to whether PTH is administered continuously or intermittently. Intermittent administration of low doses of PTH causes dramatic net increase in trabecular bone mass

**Figure 26-10** Stromal cell control of osteoclastogenesis and osteoclast activity. Parathyroid hormone (PTH) acts on PTH/PTH-related protein (PTHrP) receptors on precursors of osteoblasts to increase the production of macrophage colony-stimulating factor (M-CSF) and RANK ligand and to decrease the production of osteoprotegerin (OPG). M-CSF and RANK ligand stimulate the production of osteoclasts and increase the activity of mature osteoclasts by binding to the receptor RANK. OPG blocks the interaction of RANK ligand and RANK.

with little effect on cortical bone in humans. Continuous administration of PTH, in contrast, leads to a decrease in cortical bone mass; the net effect of PTH on trabecular bone depends on the dose. In mild hyperparathyroidism, there is little net effect of PTH on trabecular bone and a decrease in cortical bone. In all of these settings, the rate of bone formation is increased; the varying rate of osteoclastic resorption determines the net effect of PTH on bone mass.

**Molecular Basis of Parathyroid Hormone Action**

Ever since the discovery that PTH stimulates the secretion of cAMP into the urine, 12 PTH has been thought to act by triggering a cascade of intracellular second messengers. This guiding hypothesis, in its current form, postulates that all of the actions of PTH result from the binding of the hormone to a receptor on the plasma membrane of target tissues. This receptor is a member of a large family of G protein-linked receptors that span the plasma membrane seven times (Fig. 26-11). The binding of hormone on the outside of the membrane causes conformational changes in the receptor molecule that activate the receptor’s ability to release guanosine diphosphate (GDP) from the subunit of a G protein bound to the receptor. The G protein then binds guanosine triphosphate (GTP) in place of GDP. The GTP-binding subunit of the G protein then separates from the subunits, and the separate subunits of the G protein then modulate the activity of enzymes and channels. The activity of these enzymes and channels then affects proteins further downstream, eventually leading to the physiologic responses of bone and kidney cells.

**Parathyroid Hormone and Parathyroid Hormone-related Receptors**

DNA encoding a PTH/PTHrP receptor has been isolated from rat, opossum, human, pig, Xenopus (toad), and zebrafish cells and tissues. 12 The receptor mediates actions of both PTH and PTHrP. The predicted amino acid sequence of the receptor and direct mapping of inserted epitopes suggest that the receptor spans the plasma membrane seven times, but the sequence does not closely resemble the sequences of most known G protein-linked receptors. Instead, it is a member of a distinct subfamily of closely related receptors. Most of these
receptors bind peptides of 30 to 40 amino acids in length. Known members of the family act as receptors for the secretin family of peptides (secretin, vasoactive intestinal peptide (VIP), glucagon, glucagon-like peptide, growth hormone-releasing hormone, pituitary adenylate cyclase-activating peptide, gastric inhibitory peptide), corticotropin-releasing hormone, calcitonin, and insulin diuretic hormones related to corticotropin-releasing hormone. The PTH/PTHrP receptor most closely resembles receptors of the secretin group. The gene encoding the PTH/PTHrP receptor has a complicated structure, with 13 introns interrupting the coding sequence. 

In the mouse, an upstream promoter is used primarily in kidney and to some extent in liver; a downstream promoter is used in all tissues that express the gene, including bone. The human promoter uses the same start sites, but the kidney-specific promoter is much less active in humans, and a third, distinct start site is used in many tissues.

The cloned PTH/PTHrP receptor binds amino-terminal fragments of PTH and PTHrP with equal affinity. The receptor is expressed at high levels in kidney and in osteoblasts of bone but is also expressed in a wide variety of tissues, such as smooth muscle, brain, and a variety of fetal tissues, which are thought to be target tissues more for PTHrP than for PTH. In response to binding of PTH or PTHrP, the receptor activates several G proteins, including G\textsubscript{i}, G\textsubscript{q}, and its close relatives, G\textsubscript{s}, and perhaps other G proteins.

The PTH/PTHrP receptor probably mediates many of the actions of both PTH and PTHrP. The ligand-binding and signaling properties of the receptor, the pattern of expression of the receptor, and the consequences of mutation of the receptor sequence (see later) are persuasive evidence in this regard. Nevertheless, the scheme of PTH action illustrated in Figure 26-11 should be considered a simplified outline. It is unlikely that all of the actions of PTH can be explained by interactions with the cloned PTH/PTHrP receptor: Fragments of PTH that seem not to bind the receptor may be biologically active; some cells respond to PTH in ways not mimicked by the cloned receptor. Furthermore, the carboxyl-terminal portion of PTH(184) binds a cell surface protein distinct from the PTH/PTHrP receptor.

A second PTH receptor, which can be activated by PTH but not by PTHrP, called the PTH2 receptor (PTH2R), has been cloned. This receptor is expressed in multiple tissues, including brain, vascular endothelium and smooth muscle, endocrine cells of the gastrointestinal tract, and sperm. Expression is not seen in osteoblasts or renal tubules, however. Although PTH activates human PTH2R well, PTH only poorly activates the rat PTH2R. Furthermore, a novel ligand called TIP39 has been characterized and shown to be a potent activator of PTH2R. TIP39 bears only a weak resemblance to PTH or PTHrP and is likely to be a physiologically relevant activator of that receptor. The functional role of the PTH2R is unknown; this receptor may well not mediate actions of PTH in vivo. The two cloned PTHRs, as well as distinct receptors for fragments of PTHrP (see later), probably are part of a complex network of ligands and receptors.

Functional implications of Parathyroid Hormone Structure

Amino-terminal fragments of PTH as short as PTH(134) have potency at least as great as that of the full-length PTH(184). Several discrete portions of the PTH(134) peptide interact with the receptor. The first several residues of PTH are particularly important for triggering the conformational change in the receptor that results in activation of G\textsubscript{i} and adenylate cyclase. Sequence responsible for transmembrane activation of G\textsubscript{i} make up most of the first 13 residues of PTH(134); it is these residues that are highly conserved between PTH and PTHrP. At high concentrations, PTH(114) by itself can activate the PTH/PTHrP receptor. This activation domain interacts with the receptor's transmembrane domains and extracellular loops. When the first nine residues of PTH are covalently linked to the receptor's transmembrane domains and extracellular loops, they can activate the receptor. An analogue of PTH(114) can trigger G\textsubscript{i} activation, thus activating phospholipase C. These data, plus the observation that a PTH analogue modified at position 1 selectively loses its ability to activate phospholipase C, demonstrate that the amino-terminal portion of PTH is essential for activation of both G\textsubscript{i} and G\textsubscript{q}. More distal regions of PTH(134) can activate protein kinase C and can raise intracellular calcium levels by mechanisms that have not been fully clarified.

The more carboxy-terminal portions of PTH(134) contribute importantly to the specificity and tight binding of PTH to the PTH/PTHrP receptor, at least partly through interactions with the receptor's amino-terminal transmembrane domain. A variety of studies of genetically altered receptors and biochemical studies using photoactivated crosslinks between PTH and the receptor have reinforced each other and show that the carboxy-terminal portion of PTH makes multiple contacts with the amino-terminal extension of the receptor and with its extracellular loops.

Studies of the structure of PTH by nuclear magnetic resonance spectroscopy suggest that the activation domain and the carboxy-terminal domain are discrete entities dominated by helices separated by a flexible loop of variable size, depending on the hydrophobicity of the solvent. In the crystal structure of human PTH(134), the flexible loop is entirely replaced with helical structure. Taken together, these studies suggest that the carboxy-terminal portion of PTH makes multiple contacts with the receptor that allow high-affinity binding and position the amino-terminal portion of PTH to activate the receptor through contacts with transmembrane domains and associated loops.
Activation of Second Messengers

Precisely how binding of PTH to the extracellular domains of the PTH/PTHrP receptor leads to activation of G proteins is not understood. The crystal structure of rhodopsin, another member of the seven-transmembrane receptor family, as well as the behavior of certain mutant PTH/PTHrP receptors, suggest that the seven transmembrane domains of the PTH/PTHrP receptors form a ring, with the seventh transmembrane domain adjacent to the first and second domains. Presumably, binding of PTH to several different regions of the receptor changes the relationships of the transmembrane domains such that the receptor’s three intracellular loops and carboxy-terminal tail interact with G proteins in an altered way.

Receptors with certain point mutations in the second, sixth, and seventh transmembrane domains can activate Gs even without stimulation by hormone. These mutant receptors were discovered by analyzing the PTH/PTHrP receptors in patients with Jansen’s metaphyseal chondrodystrophy. Patients with this disorder have signs of parathyroid overactivity (hypercalcemia, hypophosphatemia, high levels of 1,25(OH)₂D₃, and urinary cAMP) but low PTH and PTHrP levels. The mutations must change the conformation of the intracellular face of the receptor in a way that resembles the effect of binding of PTH to the normal receptor. The observation that inappropriate activation of the PTH/PTHrP receptor in Jansen’s chondrodystrophy leads to all of the metabolic abnormalities found in primary hyperparathyroidism is one of the most persuasive pieces of evidence that the cloned PTH/PTHrP receptor does, in fact, mediate the actions of PTH in bone and kidney in vivo.

Second Messengers and Distal Effects of Parathyroid Hormone

The activation of multiple G proteins by PTH raises questions about the individual roles of each second messenger and their possible interactions. The importance of cAMP as a mediator of the physiologic actions of PTH has been demonstrated by studies in vivo and in vitro. Furthermore, patients with pseudohypoparathyroidism type I, who cannot increase urinary cAMP levels in response to PTH, show clear renal resistance to PTH (see later).

Phospholipase C, with concomitant activation of protein kinase C and synthesis of IP₃, may mediate physiologic actions of PTH as well, such as inhibition of sodium-phosphate cotransport and stimulation of the renal 25(OH)D₁-hydroxylase. Some actions of PTH may require activation of both adenylate cyclase and phospholipase C for optimal activity.

Target Cell Responsiveness to Parathyroid Hormone

Physiologic responses to PTH depend not only on the concentration of PTH in blood but also on the responsiveness of target cells to PTH. This responsiveness can be modified by previous exposure to PTH or by exposure to a variety of other hormones and paracrine factors. Responsiveness can be changed by alterations at virtually every step in the cellular response to PTH.

Major regulators of PTH/PTHrP receptor gene expression include, not surprisingly, PTH and 1,25(OH)₂D₃, both of which can decrease PTH/PTHrP receptor mRNA in certain target cells. In some settings, PTH decreases the amount of immunoreactive and functional receptor on the cell surface without changing the levels of PTH/PTHrP mRNA. This decrease reflects ligand-induced internalization and degradation of receptors. Internalization of receptor is stimulated by PTH binding and is modulated by sequences found in the membrane-proximal portion of the receptor's cytoplasmic tail. The precise mechanism of PTH-induced internalization of the PTH/PTHrP receptor is not fully understood and may vary in different cellular contexts. Even without change in receptor number, previous exposure to PTH leads to inefficient triggering of G proteins (desensitization).
PARATHYROID HORMONE-RELATED PROTEIN

Parathyroid hormonelated protein (PTHrP) was discovered because the secretion of PTHrP by a wide variety of tumors contributes to the humoral hypercalcemia of malignancy. For this reason, the initial studies of PTHrP in humans and animals stressed the PTH-like structure and properties of the molecule. Subsequent studies soon showed, however, that PTHrP, unlike PTH, is made by a wide variety of tissues, in which it acts locally in ways that may have little relevance to the control of blood calcium.

Gene and Protein Structure

PTHrP sequences from human, rat, mouse, dog, bovine species, chicken, fugu, and sea bream have been cloned\(^1\)\(^2\)\(^3\)\(^4\) (Fig. 26-14). In humans, alternative RNA splicing yields transcripts that encode three distinct proteins of 139, 141, and 173 residues that differ only after residue 139.

Inspection of these sequences suggests that PTHrP has several functionally distinct domains. Eight or nine of the first 13 residues of PTHrP are identical to those in known mammalian PTH sequences. These sequences encompass the known "activation" domain of PTH (see earlier) and are instrumental in the ability of PTHrP to activate PTH/PTHrP receptors. The conserved histidine at position 5 of all PTHrP molecules, which differs from the hydrophobic residue found at the corresponding position of all PTHs, allows PTHrP to activate the PTH/PTHrP receptor but not the PTH2 receptor.\(^5\)\(^6\)\(^7\)

The sequences in PTHrP(1434) are also highly conserved. Although these sequences little resemble the corresponding region of PTH, they can displace PTH from the PTH/PTHrP receptor. Studies of the secondary and tertiary structures of PTHrP(134) and PTH(137) suggest that they have similar structures dominated by helices connected by a flexible hinge.\(^8\)\(^9\)\(^10\)

The remaining portion of the PTHrP molecule bears no resemblance to corresponding sequences in PTH. Nevertheless, residues 35 to 111 of PTHrP are strikingly well conserved, with only nine residues differing between mammalian and chicken PTHrP sequences. This sequence conservation is considerably greater than that found in the carboxyl-terminal portion of PTH, suggesting that this region of PTHrP has unique and important functions. After residue 111, the PTHrP sequences vary considerably from species to species.

Interspersed within the PTHrP sequences are multiple sites containing one or several basic residues that might serve as post-translational cleavage sites (see Fig. 26-14). Extensive analysis of PTHrP fragments in tumors, cell lines, and transfected cells has shown that several of these sites are, in fact, functional cleavage signals.\(^11\) PTHrP is cleaved after the arginine at residue 37; this cleavage, followed by carboxypeptidase cleavage, generates a PTH-like PTHrP(136) fragment as well as the fragments PTHrP(3894) amide, PTHrP(3895), and PTHrP(38101). More carboxy-terminal fragments of PTHrP have been detected in cells as well.

In the blood of patients with humoral hypercalcemia of malignancy, multiple immunoreactive species of PTHrP have been found that may well correspond to the fragments of PTHrP in cells and tissue culture media, although precise characterization of these various immunoreactive species is incomplete (see later).\(^12\)\(^13\)\(^14\)\(^15\)

\(^{1,2,3,4}\) Full-length PTHrP may well not circulate, since an amino-terminal specific immunoaffinity column was unable to extract carboxy-terminal immunoreactivity from the serum of patients with malignant hypercalcemia.
Functions of Parathyroid HormoneRelated Protein

The first actions of PTHrP to be defined were the PTH-like actions associated with the humoral hypercalcemia of malignancy. In this pathologic entity, PTHrP acts as a hormone; it is secreted from the tumor into the blood stream and then acts on bone and kidney to raise calcium levels (see "Hypercalcemia of Malignancy" later). Whether or not PTHrP circulates at high enough levels in normal adults to contribute to normal calcium homeostasis is an unanswered question. With metastases of breast cancer to bone, locally produced PTHrP can raise serum calcium without necessarily raising blood levels of PTHrP.

Growing evidence suggests that PTHrP acts as a calciotropic hormone during fetal life and in lactation. PTHrP secreted from the fetal parathyroid gland stimulates transport of calcium across the placenta in sheep. PTH, in contrast, has no effect on placental calcium transport. Furthermore, fetal mice missing the PTHrP gene transport 45Ca across the placenta inefficiently. This action of PTHrP requires only the mid-region of PTHrP and probably involves a receptor distinct from the PTH/PTHrP receptor.

The second possible setting for humoral actions of PTHrP is lactation. Secretion of PTHrP from the breast into the blood stream may influence the movement of calcium from maternal bone into breast milk. PTHrP, therefore, probably contributes to the dramatic but largely reversible bone loss during lactation, which is only minimally affected by calcium supplementation. An exaggeration of this lactational role of PTHrP may explain the rare presentation of hypercalcemia and high PTHrP levels in pregnant and lactating women. Large amounts of PTHrP are also secreted into breast milk, although the role of PTHrP in milk is unknown.

Most of the actions of PTHrP are likely to be paracrine or autocrine. PTHrP is synthesized at one time or another during fetal life in virtually every tissue. Its role in the development of fetal bone has been demonstrated through the striking abnormalities found in genetically engineered mice missing the PTHrP gene. These abnormalities suggest that PTHrP normally keeps chondrocytes proliferating in orderly columns, thereby delaying chondrocyte differentiation. The role of PTHrP in many other fetal tissues may analogously involve regulation of proliferation and differentiation. The widespread expression of the PTHrP in fetal life probably underlies the expression of PTHrP in a wide variety of malignancies. As is often the case in malignancy, the expression of PTHrP represents the reinitiation of a fetal pattern of gene expression.

PTHrP is synthesized by many adult tissues. In tissues such as skin, hair, and breast, it is likely that PTHrP regulates cell proliferation and differentiation. PTHrP is also synthesized in response to stretch in the smooth muscle of blood vessels and of the gastrointestinal tract, uterus, and bladder and acts in an autocrine fashion to relax the smooth muscle. PTHrP is also widely expressed in neurons of the central nervous system; its function in the brain is unknown, but it may protect neurons from excitotoxicity by decreasing flux through voltage-gated calcium channels; an analogous mechanism may explain the role of PTHrP in smooth muscle relaxation.

Many of the actions of PTHrP are mediated by the PTH/PTHrP receptor. Others, such as the activation of placental calcium transport, are probably mediated by a distinct receptor, and other actions on bone cells probably involve yet another receptor responsive to more distal portions of PTHrP. Increasing evidence suggests, furthermore, that some actions of PTHrP involve direct nuclear actions of PTHrP. Thus, both PTH and PTHrP are likely to use multiple mechanisms to stimulate cells.
CALCITONIN

Calcitonin has an important role in regulating blood calcium in fish and a demonstrable role in rodents; however, the importance of calcitonin in human calcium homeostasis remains uncertain.

The existence of a second calcium-regulating hormone, in addition to PTH, was first demonstrated during perfusion studies of the thyroid and parathyroid glands of dogs. High-level calcium perfusion resulted in a rapid decrease in plasma calcium, even more rapid than after parathyroidectomy. This effect suggested that calcium had stimulated the secretion of a hormone that lowered blood calcium. It was subsequently demonstrated that this missing hormone, named calcitonin for its role in regulating the "tone" or level of calcium, was elaborated by the thyroid gland, not the parathyroids. Calcitonin is found in the nonfollicular cells of the thyroid, called C cells, which originate from the neural crest. In fish, the location of the C cells in discrete organs led to the rapid isolation of calcitonin from these ultimobranchial bodies in dogfish, salmon, and several other species. The identification of the glandular origin of calcitonin enabled the isolation of sufficient quantities of calcitonin for sequence analysis and studies of its structure and biologic function.

Synthesis and Secretion

Calcitonin consists of a 32-amino-acid polypeptide with an intrachain disulfide bond provided by the cysteines at positions 1 and 7. These two cysteine residues, along with the carboxy-terminal proline-amide and six additional residues are the only amino acids conserved among the calcitonins isolated from various species. The disulfide linkage and prolineamide residues are important for the function of the molecule, although biologically active analogues lacking disulfide bonds have been developed.

Of interest, fish calcitonin is more potent in mammals than is the mammalian hormone. The mature peptide is derived from the middle of a 136-amino-acid precursor. The human calcitonin gene, located on the short arm of chromosome 11, contains six exons that are alternatively spliced in a tissue-specific manner to yield the mRNAs encoding calcitonin or calcitonin gene-related peptide (CGRP). The mRNA encoding calcitonin is derived by splicing together the first four exons and represents over 95% of mature transcripts in the thyroid C cells. The splicing of the first three exons to exons 5 and 6 results in an mRNA that encodes the 37-amino-acid -CGRP. The mRNA encoding -CGRP is expressed in multiple tissues and is the only mature transcript of the calcitonin gene detected in neural tissue. A second CGRP gene encodes the closely related -CGRP. In humans, the predicted sequence of the mature peptide differs from that of -CGRP by only three amino acids.

The synthesis and secretion of calcitonin are tightly regulated. Studies in a porcine model reveal a linear relationship between the secretion of calcitonin and ambient calcium levels. Cell culture studies using calcium ionophores and calcium channel blockers demonstrate that the calcium ion concentration within the C cell determines this secretion rate.

The calcium-sensing receptor cloned from parathyroid cells is also expressed in C cells and contributes to the regulation of calcitonin secretion. Other calcitonin secretagogues include glucocorticoids, CGRP, glucagon, enteroglucagon, gastrin, pentagastrin, pancreozymin, and -adrenergic agents. The physiologic role of the gastrointestinal hormones in regulating calcitonin remains unclear; however, they have been postulated to play a role in the regulation of postprandial hypercalcemia. The secretion of calcitonin is inhibited by somatostatin, which is also secreted by the thyroidal C cells. In vivo studies have demonstrated that 1,25(OH)_2 D_3 decreases calcitonin mRNA levels by a transcriptional mechanism.

Calcitonin, when administered acutely, decreases tubular resorption of calcium and impairs osteoclast-mediated bone resorption by a direct action on osteoclasts. In rodents, calcitonin is known to play a role in the regulation of postprandial hypercalcemia. Studies in calcitonin knockout mice reveal a doubling of bone formation rate in the absence of hormone, accompanied by resistance to ovariectomy-induced bone loss. The physiologic role of calcitonin in humans, however, remains elusive. The effect of calcitonin on bone density was examined in patients with long-term hypercalcitoninemia secondary to medullary carcinoma of the thyroid gland (MCT) and in patients with subtotal thyroidectomy resulting in lack of calcitonin secretory reserve. Bone density at the lumbar spine and distal radius were not influenced by the abnormal calcitonin levels. Furthermore, no physiologic abnormalities have been reported with long-term, high-dose administration of exogenous calcitonin.

Many of the effects of calcitonin are mediated by a G protein-coupled cell surface receptor in the PTH/secretin receptor family. The mRNA encoding this receptor has been found in multiple tissues, including kidney, brain, and osteoclasts. The coupling of this receptor to different G proteins results in the activation of either adenylate cyclase or phospholipase C; in some settings, this effect is cell cycledependent. Glycosylation of the receptor is important for both binding and signal transduction. Several isoforms of the calcitonin receptor have been described, but the functional significance of these various isoforms is not known.
Calcitonin Family: Calcitonin Gene-Related Peptide, Amylin, and Adrenomedullin

CGRP, amylin, and adrenomedullin all have been shown to have high-affinity binding sites, and displacement studies suggest that several receptor subtypes are present. However, cloning of specific receptors for these ligands proved difficult because the functional receptors consist of heterodimers between G protein-coupled receptors and single transmembrane proteins of the RAMP (receptor activity-modifying protein) family. Interaction of the calcitonin receptor-like receptor, a relative of the calcitonin receptor, with RAMP1 results in a CGRP receptor, whereas RAMP2 and RAMP3 interactions with the same calcitonin receptor-like receptor generate adrenomedullin receptors. Interaction of RAMP1 with the calcitonin receptor creates an amylin receptor.

CGRP is thought to act as a neurotransmitter and vasodilator rather than as a hormone. In support of this hypothesis, mice lacking -CGRP have an increase in mean arterial pressure. Immunohistochemical studies of CGRP in the brain and peripheral nervous system suggest that this neuropeptide also plays an important role in sensory and integrative motor functions.

Amylin is highly homologous to CGRP and calcitonin. The striking presence of amylin in the pancreas of patients with type 2 diabetes mellitus suggests an etiologic role for this peptide in this disorder.

Adrenomedullin has vasodilatory effects similar to those of CGRP. In addition to activating CGRP receptors, adrenomedullin binds to specific receptors in the vascular system. The physiologic roles of amylin and adrenomedullin and the functional correlates of their receptor interactions have not yet been clarified.
Calcitonin in Human Disease

Calcitonin is secreted by several endocrine malignancies and can therefore serve as a tumor marker. Basal and pentagastrin-stimulated calcitonin levels have been used to identify and follow persons at risk for, or affected by, medullary carcinoma of the thyroid gland (see Chapter 13), although abnormal basal and stimulated levels may be observed in patients on chronic hemodialysis. Calcitonin may also be ectopically secreted by other tumors, including insulinomas, VIPomas, and lung cancers. Severely ill patients, including those with burn inhalation injury, toxic shock syndrome, and pancreatitis, also may have elevated calcitonin levels.
Therapeutic Uses

The observation that calcitonin inhibits osteoclastic bone resorption has led to its therapeutic use for the treatment of several disorders associated with excess bone resorption, including osteoporosis and Paget's disease (see Chapter 27). Calcitonin has also been shown to have an analgesic effect in the treatment of vertebral crush fractures, osteolytic metastases, and phantom limb.
VITAMIN D

Metabolism of Vitamin D

Vitamin D is not a true vitamin, since nutritional supplementation is not required in humans who have adequate sun exposure. When the skin is exposed to ultraviolet radiation,

the cutaneous precursor of vitamin D, 7-dehydrocholesterol, undergoes photochemical cleavage of the carbon bond between carbons 9 and 10 of the steroid ring (Fig. 26-17). The resultant product, previtamin D, is thermally labile and over a period of 48 hours undergoes a temperature-dependent molecular rearrangement that results in the production of vitamin D. Alternatively, this thermally labile product can isomerize to two biologically inert products, lumisterol and tachysterol. This alternative photoisomerization prevents production of excessive amounts of vitamin D with prolonged sun exposure. The degree of skin pigmentation, which increases in response to solar exposure, also regulates the conversion of 7-dehydrocholesterol to vitamin D by blocking the penetration of ultraviolet rays.

The alternative source of vitamin D is dietary. The elderly, the institutionalized, and those living in northern climates probably obtain most of their vitamin D from dietary sources. However, with increasing avoidance of sun exposure by the general population, ensuring adequate dietary intake of vitamin D has become important for the population at large. Vitamin D deficiency is prevalent and has been shown to contribute significantly to osteopenia and fracture risk. The major dietary sources of vitamin D are fortified dairy products, although the lack of monitoring of this supplementation results in marked variation in the amount of vitamin D provided. Other dietary sources include egg yolks, fish oils, and fortified cereal products. Vitamin D provided by plant sources is in the form of vitamin D₃, whereas that provided by animal sources is in the form of vitamin D₂ (see Fig. 26-17). These two forms have equivalent biologic potencies and are activated equally efficiently by the hydroxylases in humans.

Vitamin D is absorbed into the lymphatics and enters the circulation bound primarily to vitamin D-binding protein, although a fraction of vitamin D circulates bound to albumin. The human vitamin D-binding protein is an α-globulin, with a molecular mass of approximately 52 kDa, that is synthesized in the liver. The protein has a high affinity for 25(OH)D but also binds vitamin D and 1,25(OH)₂D₃. Approximately 88% of 25(OH)D circulates bound to the vitamin D-binding protein, 0.03% is free, and the rest circulates bound to albumin. In contrast, 85% of the circulating 1,25(OH)₂D₃ binds to the vitamin D-binding protein, 0.4% is free, and the rest binds to albumin. Mice lacking vitamin D-binding protein have increased susceptibility to 1,25(OH)₂D₃ toxicity as well as to dietary vitamin D deficiency. Thus, the role of vitamin D-binding protein is to maintain a serum reservoir and to modulate the activity of vitamin D metabolites. Recent studies in megalin-null mice suggest that vitamin D-binding protein is filtered by the glomerulus and reabsorbed by a megalin-dependent pathway in the proximal renal tubule. Further investigations are required to determine the importance of this pathway in vitamin D metabolism and homeostasis.

In the liver, vitamin D undergoes 25-hydroxylation by a cytochrome P450 enzyme present in the mitochondria and microsomes. The half-life of 25(OH)D is approximately 2 to 3 weeks. The 25-hydroxylation of vitamin D is not tightly regulated; therefore, the blood levels of 25(OH)D reflect the amount of vitamin D entering the circulation. When levels of vitamin D-binding protein are low, such as in nephrotic syndrome, circulating levels of 25(OH)D are also reduced. The half-life of 25(OH)D is shortened by increases in levels of its active metabolite, 1,25(OH)₂D₃.

The final step in the production of the active hormone is the renal 1-hydroxylation of 25(OH)D to 1,25(OH)₂D₃. The half-life of this hormone is approximately 6 to 8 hours. Like the 25-hydroxylase, the 1-hydroxylase in the renal proximal convoluted tubule is a cytochrome P450 mixed-function oxidase. Unlike the 25-hydroxylase, however, 1-hydroxylation is tightly regulated. PTH and hypophosphatemia are the major inducers of this microsomal enzyme, whereas calcium and the enzyme's product, 1,25(OH)₂D₃, repress it in animal models and in vitro studies. Other hormones such as estrogen, calcitonin, growth hormone, and prolactin increase 1-hydroxylase activity; however, the clinical importance of these observations has not been established. Ketoconazole has been shown to decrease levels of 1,25(OH)₂D₃ in a dose-dependent manner, presumably by interfering with 1-hydroxylase activity.

Figure 26-18 Relative potency of analogues of 1,25(OH)₂D₃. (1,25-Dihydroxyvitamin D₃) in competitive binding to vitamin D receptors of chick intestinal mucosa. Slopes are plotted for (left to right): 1,25(OH)₂D₃, 1,25-dihydroxyvitamin D₃, 1,24R,25(OH)₂D₃, 1-deoxy-1,25(OH)₂D₃, 3-deoxy-1,25(OH)₂D₃, 3-deoxo-1,24,25(OH)₃D₃, 25-hydroxy-5,6-trans-D₃, 25-hydroxy-5,6-cis-D₃, 25-hydroxyvitamin D₃, 25-OH-D₃, 25(OH)-24R,25(OH)₂D₃, 24,25(OH)₂D₃, 24,25(OH)₂D₃, 24,25(OH)₂D₃, 3-deoxo-1,25(OH)₂D₃, 3-deoxy-1,24,25(OH)₃D₃, 3-deoxy-1,25(OH)₂D₃, 3-deoxy-1,24,25(OH)₃D₃, 3-deoxy-1,24,25(OH)₃D₃. Vitamin D₃, D₃; DHT, dihydrotestosterone. (From Prosall D-A, Okamura WH, Norman AW. Structural requirements for the interaction of 1,25-(OH)₃vitamin D₃ with its chick intestinal receptor. J Biol Chem 1975, 250:836388.)

1-Hydroxylase is also found in keratinocytes, the trophoblastic layer of the placenta, some lymphomas and granulomata, including sarcoid granulomata. In malignant and granulomatous tissue, the 1-hydroxylase gene that is expressed is identical to that expressed in the kidney, but the gene is not regulated by PTH, phosphate, calcium, or vitamin D metabolites in these cells. Activation of macrophages with interferon, however, increases the expression of the 1-hydroxylase, whereas treatment of sarcoïdosis-associated hypercalcemia with glucocorticoids, ketoconazole, or chloroquine lowers serum 1,25(OH)₂D₃ levels.

25-Hydroxyvitamin D and 1,25(OH)₂D₃ can also be hydroxylated by the vitamin D 24-hydroxylase that is present in most tissues, including kidney, cartilage, and intestine. 1,25-Dihydroxyvitamin D₃ increases the activity of the 24-hydroxylase, thereby inducing its own metabolism. The 24-hydroxylated vitamin D metabolites, 24,25(OH)₂D₃ and 1,24,25(OH)₃D₃, are not thought to play major biologic roles other than inactivation of 1,25(OH)₂D₃. Mice null for the 24-hydroxylase gene demonstrate hypercalcemia, hypercalciuria, and nephrocalcinosis owing to lethal vitamin D toxicity. Although 24,25(OH)₂D₃ has been shown to have unique actions in a number of biologic systems, no unique receptor for this metabolite has been identified, and the physiologic role of 24,25(OH)₂D₃ is unclear.

Figure 26-17 Vitamin D precursors and alternative reaction products. The numbering system for vitamin D carbons and the distinct structures of vitamin D₃ (ergocalciferol) and D₂ (cholecalciferol) are noted, as is the structure of dihydrocholsterol, a synthetic product not produced in vivo. Note that the 3-hydroxy group of dihydrocholesterol is in a pseudo-1-hydroxy configuration. This may explain the relatively high potency of dihydrocholesterol in conditions associated with low 1-hydroxylase activity.
1,25-Dihydroxyvitamin D₃ is also metabolized to several inactive products by 23- or 26-hydroxylation and side chain oxidation and cleavage. This side chain cleavage, resulting in the formation of calcitroic acid, occurs in the liver and intestine, whereas inactivation of 1,25(OH)₂D₃ in a wide variety of target tissues occurs by 24-hydroxylation. In addition, polar metabolites of 1,25(OH)₂D₃ are excreted in the bile. Some of these metabolites are deconjugated in the intestine and reabsorbed into the enterohepatic circulation.
1,25-Dihydroxyvitamin D₂ exerts its biologic functions by binding to a nuclear receptor, which then regulates transcription of DNA into RNA. Among the other nuclear receptors, the vitamin D receptor most closely resembles the retinoid acid, triiodothyronine, and retinoid-X receptors. The affinity of the receptor for 1,25(OH)₂D₃ is approximately three orders of magnitude higher than that for other vitamin D metabolites. Although 25(OH)D₃ is less potent on a molar basis, its concentration in the serum is approximately three times higher than that of 1,25(OH)₂D₃ [Fig. 26-18]; however, its free concentration is only twice that of 1,25(OH)₂D₃. Therefore, under normal circumstances it is unlikely that 25(OH)D₃ contributes importantly to calcium homeostasis.

Because the affinity of the vitamin D-binding protein for 25(OH)D is greater than for 1,25(OH)₂D₃, in states of vitamin D intoxication, the free levels of 1,25(OH)₂D₃ increase, because 25(OH)D displaces it from the vitamin D binding protein. 25-Hydroxyvitamin D, therefore, may play a role in the clinical syndrome of vitamin D intoxication both by its direct biologic effects, when present at toxic levels, and by increasing free levels of 1,25(OH)₂D₃.

The vitamin D receptor acts by forming a heterodimer with the retinoid-X receptor, binding to DNA elements, and recruiting coactivators in a ligand-dependent fashion. These coactivators link the receptor complex to the basal transcription apparatus, thereby regulating transcription of target genes. In most cases, the up-regulatory response elements for vitamin D contain hexameric repeats separated by three bases. However, vitamin D also promotes the DNA-protein interactions of other transcription factors, such as SP1 and NF-κB, in genes lacking classical response elements, by uncertain mechanisms. The mechanism of transcriptional repression by vitamin D are varied. For example, VDR-RXR heterodimers repress the 1-hydroxylase gene by binding to and blocking the function of another transcription factor, whereas interaction of the vitamin D receptor with the Kᵣ antigen, acting as a transcription factor, is required for transcriptional repression of the human PTHRP gene.

Glucocorticoids have been shown to decrease the expression of the vitamin D receptor gene in osteosarcoma cell lines, whereas 1,25(OH)₂D₃ increases its expression in many cells. In the renal proximal convoluted tubule, however, 1,25(OH)₂D₃ decreases the levels of vitamin D receptors. This decrease has been postulated to lead to decreased activation of the renal 24-hydroxylase by 1,25(OH)₂D₃, thereby protecting the newly synthesized 1,25(OH)₂D₃ from local inactivation.

1,25-Dihydroxyvitamin D₂ also has some biologic effects that occur too rapidly for transcriptional mechanisms to be implicated. These nongenomic actions, including the rapid increase in intracellular calcium, activation of phospholipase C, and opening of calcium channels, are observed in several cell types within minutes of exposure to 1,25(OH)₂D₃. Additional data supporting the hypothesis that nongenomic actions are not dependent on the classical receptor include the identification of specific binding sites for 1,25(OH)₂D₃ on the antiluminal surface of intestinal cells and a disparity between the affinity of the various vitamin D analogues for the nuclear receptor and their potency in these nongenomic actions. Evidently, at least one of the nongenomic actions of vitamin D is available for the nucleus in association with the vitamin D receptor depends on the presence of an intact intracellular receptor, because this effect is not observed in cells derived from patients with vitamin D receptor mutations. The physiologic importance of the nongenomic actions of vitamin D metabolites has not yet been established.

The vitamin D receptor is expressed in most tissues and regulates cellular differentiation and function in many cell types. Nevertheless, the most dramatic physiologic effect of vitamin D acting through the vitamin D receptor, involves regulation of intestinal calcium transport. This effect is most clearly demonstrated by the phenotype of patients and mice with mutant vitamin D receptors (vitamin D-dependent rickets type 2). Dramatic abnormalities in bone mineralization can be reversed by bypassing the defect in intestinal calcium absorption.

Intestinal Calcium Absorption

Under normal dietary conditions, calcium intake is in the range of 700 to 900 mg daily. Approximately 30% to 35% of this calcium is absorbed; however, losses from intestinal secretion of calcium lead to a net daily uptake of only approximately 200 mg. Although vitamin D is the major hormonal determinant of intestinal calcium absorption, the bioavailability of mineral ions in the intestinal lumen may be affected by a number of local factors and dietary constituents. Absorption of calcium and magnesium is impaired by bile salt deficiency, unabsorbed free fatty acids in steatorrheic states, and high dietary content of fiber or phytate. Gastric acid is needed to promote dissociation of calcium from anionic components of food or therapeutic preparations of calcium salts. Administration of calcium salts with meals, especially in patients with achlorhydria, and use of divided doses or more soluble salts, such as calcium citrate, are commonly employed strategies to increase calcium bioavailability.

Calcium is thought to be absorbed by three pathways: (1) the transcellular route, (2) vesicular calcium transport, and (3) paracellular transport. The first two pathways are dependent on 1,25(OH)₂D₃. Although the necessity of vitamin D for paracellular calcium absorption remains controversial, substantial evidence exists that the hormone enhances this pathway as well. The vesicular pathway is thought to involve the nongenomic actions of 1,25(OH)₂D₃. However, this rapid transport (transcalcehia) is observed only in vitamin D-deplete animals. The relative contribution of vesicular calcium transport to intestinal calcium absorption and the specific mediators of this transport have not yet been clarified.

The most extensively studied mechanism of intestinal calcium absorption involves the transcellular route. This pathway is thought to involve three steps: (1) entry of calcium into the enterocyte, (2) transport across the cell, and (3) extrusion across the basolateral membrane.

A number of brush border proteins, including the intestinal membrane calcium-binding protein, brush border alkaline phosphatase, and low-affinity Ca²⁺-Mg²⁺-ATPase, are induced by 1,25(OH)₂D₃. The activity of these proteins correlates with active calcium transport; however, a causal relationship remains to be established. Two calcium channels, CaT₁ and EcaC with six membrane-spanning domains and significant homology to the capsacin receptor, are expressed in the duodenum and jejunum and have been implicated as important regulators of calcium entry into the enteroocyte. Upon entering the enteroocyte, calcium binds to components of the brush border complex subjacent to the plasma membrane. Calmodulin is redistributed to the brush border in response to 1,25(OH)₂D₃.
and may play a role in this process, as may the 1,25(OH)₂D₃-inducible calcium binding protein calbindin 9K.

**Transcellular Transport**

The best-studied effect of vitamin D on the enterocyte is the induction of synthesis of the intestinal calcium-binding protein calbindin 9K. This protein has an EF hand structure that permits the binding of two calcium ions per molecule. The affinity of calbindin for calcium is approximately four times that of the brush border calcium-binding components; thus, calcium is preferentially transferred to calbindin. Calbindin serves to buffer the intracellular free calcium concentration during calcium absorption. It associates with microtubules and may play a role in the transport of calcium across the enterocyte. Organelles such as the mitochondria, Golgi, and endoplasmic reticulum also serve as reservoirs for intracellular calcium.

**Exit from the Enterocyte**

The transport of calcium across the antiluminal surface of the enterocyte, the final process involved in intestinal calcium absorption, is dependent on 1,25(OH)₂D₃. The main mechanism of calcium extrusion is the 1,25(OH)₂D₃-inducible ATP-dependent Ca²⁺ pump. The affinity of the pump for calcium is approximately 2.5 times that of calbindin. With high calcium intake, a 1,25(OH)₂D₃-independent Na⁺–Ca²⁺ exchanger may play a role in the transfer of calcium across the basolateral membrane as well.

**Actions on the Parathyroid Gland**

1,25-Dihydroxyvitamin D₃ has been shown to regulate gene transcription and cell proliferation in the parathyroid glands. The hormone also inhibits the proliferation of dispersed parathyroid cells in culture, although the relative contribution of calcium and 1,25(OH)₂D₃ in the regulation of parathyroid cell proliferation in vivo has not been established. Normocalcemic mice lacking functional vitamin D receptors have normal serum PTH levels and normal-size parathyroid glands, demonstrating that the genomic actions of 1,25(OH)₂D₃ are not essential for parathyroid cellular homeostasis. 1,25-Dihydroxyvitamin D₃, however, decreases the transcription of the PTH gene both in vivo and in vitro. This action has been exploited in the use of 1,25(OH)₂D₃ in the treatment of the secondary hyperparathyroidism associated with chronic renal failure (see "Parathyroid Hormone Biosynthesis" earlier and "Vitamin D Deficiency" later).

**Actions on Bone**

The effects of 1,25(OH)₂D₃ on bone are numerous. 1,25-Dihydroxyvitamin D₃ is a major transcriptional regulator of the two most abundant bone matrix proteins: it represses the synthesis of type I collagen and induces the synthesis of osteocalcin. 1,25-Dihydroxyvitamin D₃ promotes the differentiation of osteoclasts from monocyte-macrophage stem cell precursors in vitro, and it also increases osteoclastic bone resorption in high doses in vivo, by stimulating production of osteoclast-differentiating factor (ODF, i.e., RANK ligand) by osteoblasts. Despite the multiple effects of 1,25(OH)₂D₃ on the biology of bone in vitro, results of in vivo studies in 1,25(OH)₂D₃-deficient rats and in mice lacking functional vitamin D receptors suggest that the major osseous consequences of hormone and receptor deficiency can be reversed when mineral ion homeostasis is normalized. In addition, parietal calcium infusions have been shown to heal the osteomalacic lesions in children with mutant vitamin D receptors. These observations suggest that the major role of 1,25(OH)₂D₃ in bone is to provide the proper microenvironment for bone mineralization through stimulation of the intestinal absorption of calcium and phosphate.

**Other Actions of Vitamin D**

The effects of 1,25(OH)₂D₃ on phosphate transport are less well studied than those on calcium transport; however, vitamin D promotes the already efficient intestinal phosphate absorption.

One of the striking clinical features of profound vitamin D deficiency that remains unexplained is the severe proximal myopathy. Muscle cells express vitamin D receptors, and 1,25(OH)₂D₃ has nongenomic effects on muscle. Furthermore, 1,25(OH)₂D₃ increases amino acid uptake and alters phospholipid metabolism in vitro in muscle cells. Vitamin D administration has been shown to increase the concentration of troponin C, a calcium-binding protein that plays a role in excitation coupling and increases the rate of uptake of calcium by the sarcoplasmic reticulum. However, little is known regarding the direct role of vitamin D in normal muscle physiology. The myopathy that accompanies vitamin D deficiency is characterized by normal creatine kinase levels, a myopathic electromyogram, and biopsy findings of loss of myofibrils, fatty infiltration, and interstitial fibrosis. The myopathy resolves within days to weeks of vitamin D replacement and is not related to normalization of mineral ion homeostasis.

**Vitamin D Analogues**

The recognition that 1,25(OH)₂D₃ promotes cellular differentiation and inhibits cellular proliferation has led to efforts directed at producing new analogues that retain these effects but do not cause hypercalcemia. Several analogues have been shown to have antiproliferative effects on normal cells as well as on malignant cells in vitro and in xenografts in immunosuppressed mice. In addition, analogues of vitamin D have been shown to synergize with cyclosporine in preventing rejection of transplanted pancreatic islet cells in a murine model. One analogue, 22-oxacalcitriol, suppresses PTH synthesis and secretion in rats at doses that stimulate intestinal calcium absorption less than that caused by 1,25(OH)₂D₃. This finding suggests that such analogues may be useful in preventing and treating hyperparathyroidism. The antiproliferative effects of vitamin D have been exploited clinically in the treatment of psoriasis. Although analogues with reduced calcemic activity are predominantly used, hypercalcemic crisis after excessive topical use of such compounds can occur.

The physiologic processes underlying the differential biologic effects of these analogues are not completely understood. Altered affinity for the vitamin D binding protein, metabolism by target tissues, and effects on recruitment of coactivators by the vitamin D receptor may contribute to the unique properties of vitamin D analogues.
CALCIUM AND PHOSPHATE HOMEOSTASIS

The cytosolic concentrations of intracellular calcium, phosphorus, and magnesium differ markedly, as reviewed previously, and their physiologic roles within cells are diverse and largely unrelated (see Fig. 26-1). In contrast, the concentrations of these mineral ions in extracellular fluid are quite comparable (i.e., 1 to 2 mM), and it is here that they exert important interactions, both with cells and with one another, that are critical for bone mineralization, neuromuscular function, and normal mineral ion homeostasis. Extracellular calcium and phosphate, in particular, exist so close to the limits of their mutual solubility that stringent regulation of their concentrations is required to avoid diffuse precipitation of calcium phosphate crystals in tissues.

Serum concentrations and total body balances of the mineral ions are maintained within narrow limits by powerful, interactive homeostatic mechanisms. PTH and 1,25(OH)\textsubscript{2}D\textsubscript{3} regulate mineral ion levels, mineral ion levels regulate PTH and 1,25(OH)\textsubscript{2}D\textsubscript{3} secretion, and both hormones regulate the production of each other. Calcium-sensing receptors in the parathyroid glands control PTH secretion by monitoring the blood concentration of ionized calcium, whereas those in the kidney act to adjust tubular calcium reabsorption, independent of PTH or 1,25(OH)\textsubscript{2}D\textsubscript{3}. The operation of these homeostatic mechanisms can easily be appreciated by considering specific examples of how the organism adapts to changes in calcium loads (Fig. 26-20).

Dietary calcium restriction, for example, is followed by an increase in the efficiency of intestinal calcium absorption. This increased efficiency results from a sequence of homeostatic responses in which lowered blood ionized calcium activates secretion of PTH, PTH augments synthesis of 1,25(OH)\textsubscript{2}D\textsubscript{3} by the proximal tubules of the kidney, and 1,25(OH)\textsubscript{2}D\textsubscript{3} then acts directly upon enterocytes to increase active transcellular transport of calcium. Enhanced intestinal calcium absorption is quantitatively the most important response to calcium deprivation, but a series of other homeostatic events also occur that limit the impact of this stress. Renal tubular calcium reabsorption is increased by PTH, an effect that is enhanced by increased 1,25(OH)\textsubscript{2}D\textsubscript{3}-stimulated expression of calbindin-D\textsubscript{28K} in the distal tubules. Calcium reabsorption is also enhanced directly by any tendency to hypercalcemia, which is detected by calcium-sensing receptors in Henle’s loop (and possibly also in the distal nephron) that control transepithelial calcium movements independent of PTH or 1,25(OH)\textsubscript{2}D\textsubscript{3}.

The impact of dietary calcium deprivation is reduced by approximately 15% through release of calcium from bone in response to PTH and 1,25(OH)\textsubscript{2}D\textsubscript{3}. The concomitant increase in net bone resorption causes release of phosphate as well as calcium into the extracellular fluid. Intestinal phosphate absorption also is increased by 1,25(OH)\textsubscript{2}D\textsubscript{3}. These phosphate loads are problematic, in that phosphate directly lowers ionized calcium in extracellular fluid, suppresses renal synthesis of 1,25(OH)\textsubscript{2}D\textsubscript{3}, and directly inhibits bone resorption. These potentially negative effects of phosphate are obviated by the powerful phosphaturic action of PTH.

Finally, the possibility of unrestrained secretion of PTH, leading to excessive bone resorption and severe hypophosphatemia, is prevented by the effects of calcium on PTH secretion and by the direct suppressive effect of 1,25(OH)\textsubscript{2}D\textsubscript{3} on the synthesis of PTH and of PTH receptors. As a result of these homeostatic responses, calcium-deprived people maintain near-normal serum calcium and phosphate concentrations but display increased intestinal calcium absorption, increased bone resorption and progressive osteopenia, increased renal tubular calcium reabsorption, decreased renal tubular phosphate reabsorption, low urinary calcium excretion, elevated urinary phosphate excretion, and high serum concentrations of PTH and 1,25(OH)\textsubscript{2}D\textsubscript{3}.

Calcium loads induce an opposite series of adaptations: parathyroid suppression, inhibition of renal 1,25(OH)\textsubscript{2}D\textsubscript{3} synthesis, decreased intestinal active transport of calcium, increased renal excretion of calcium and decreased renal excretion of phosphate (secondary to functional hypoparathyroidism), and a decrease in bone resorption sufficient to allow positive skeletal calcium balance. The decline in intestinal calcium absorption is the major safeguard against calcium overload, although this mechanism may be overridden with extraordinarily high intakes of calcium because of the persistence of the passive, non-vitamin D-dependent mode of calcium absorption. Moreover, nonenteral sources of calcium, such as intravenous calcium infusion or excessive net bone resorption (as from immobiliation or malignancy), may readily overwhelm the limited homeostatic adaptations that remain once suppressed intestinal calcium absorption is bypassed. In such situations, the kidney rather than the intestine becomes the principal defense against hypercalcemia, and calcium homeostasis becomes critically dependent on adequate renal function. If renal function is impaired in these settings, as frequently occurs clinically, severe hypercalcemia and pathologic calcium deposition in extraskeletal sites may ensue.
LABORATORY ASSESSMENT OF MINERAL METABOLISM

Parathyroid Hormone

The major challenges in the measurement of blood PTH have been the low levels of circulating PTH and the presence of inactive PTH fragments in far greater abundance than for the intact, biologically active PTH molecule. The measurement of inactive fragments would not be a concern if the ratio of inactive to active PTH molecules remained constant. However, this ratio does change in response to changes in GFR and in parathyroid gland secretory activity (see “Parathyroid Hormone Secretion” and “Peripheral Metabolism of Parathyroid Hormone” earlier). Consequently, radioimmunoassays of PTH have suffered from lack of sensitivity and from the inability to measure the biologically active hormone directly.

For these reasons, two-site assays that require the presence of amino-terminal and carboxy-terminal sequences of full-length PTH(1-84) on the same molecule have generally replaced older radioimmunoassays. It is reassuring that for all assays of this type, the normal ranges for blood PTH are very similar. The assays are sensitive enough to detect PTH in all normal persons. The assays have demonstrated modest circadian variation in PTH levels and some pulsatility in PTH secretion, but these variations have not interfered with the diagnostic usefulness of randomly drawn PTH measurements. Some studies have reported modest increases of PTH levels with age, although others have not. Unlike older radioimmunoassays, the two-site assays demonstrate virtually no overlap in PTH levels between patients with primary hyperparathyroidism and those with nonparathyroid hypercalcemia (Fig. 26-21). Because this distinction represents the most important challenge in the clinical setting, the use of the two-site assay has dramatically facilitated the clinician’s task.

This straightforward picture has been complicated by the realization that many two-site assays do detect small amounts of PTH fragments that are large but do not extend to the hormone’s amino-terminus. These fragments accumulate in significant amounts in patients with renal failure. The clinical importance of the detection of large fragments of PTH by most two-site assays is unknown.
Parathyroid HormoneRelated Protein

The measurement of PTHrP in serum presents a series of challenges that have made the development of clinically useful assays for PTHrP more difficult than for PTH assays. The concentration of PTHrP in the blood stream, even in some patients with PTHrP-mediated malignant hypercalcemia, is not high, and the molecular definition of circulating, biologically active fragments is incomplete. Despite these problems, several groups of investigators have developed assays for PTHrP that can be helpful in the evaluation of a subset of hypercalcemic patients. Radioimmunoassays for amino-terminal portions of PTHrP and two-site assays for amino-terminal and mid-region PTHrP separate healthy persons and patients with nonmalignant hypercalcemia from most patients with the humoral hypercalcemia of malignancy. When measured with the most recently developed assays, PTHrP levels are elevated in almost all patients with malignant hypercalcemia without bone metastases and in most patients with hypercalcemia and bone metastases.

In occasional patients, the PTHrP assay has helped distinguish an occult malignancy from other causes of nonPTH-dependent hypercalcemia. Nevertheless, because the diagnosis of malignancy as the cause of hypercalcemia is usually clinically obvious, and the PTH assay can be used to diagnose primary hyperparathyroidism, the role of PTHrP assays in clinical practice is likely to be limited. Further improvement in assay sensitivity and further understanding of the normal functions of circulating PTHrP fragments may lead to changes in this assessment.
Calcitonin

Several assays for measuring serum calcitonin are commercially available. The measurements are based on single-antibody or double-antibody radioimmunoassays or enzyme immunoassays, several of which are sufficiently sensitive to detect calcitonin deficiency. The calcitonin monomer is thought to be the biologically active molecule; therefore, some investigators believe that extraction of the multimeric forms before radioimmunoassay provides a more sensitive and specific measurement of serum calcitonin levels. However, the double-antibody radioimmunoassay is thought by others to provide the same information with less sample manipulation. The only clinical use of the calcitonin assay is as a tumor marker, primarily in medullary carcinoma of the thyroid.
Vitamin D Metabolites

The currently used radioligand assays for determining the levels of vitamin D metabolites require fractionation and extraction of the hormone from serum proteins by HPLC or silica cartridges. These assays are sufficiently sensitive to detect subnormal values. Because the assays measure both protein-bound and unbound vitamin D metabolites, results may not always reflect the levels of biologically relevant ("free") metabolites. This limitation may lead to misleading results in patients with nephrotic syndrome and vitamin D intoxication.

The levels of 25(OH)D correlate better with the clinical signs and symptoms of vitamin D deficiency than do the levels of 1,25(OH)₂D₃. Because the 25-hydroxylation of vitamin D is not tightly regulated, measurements of 25(OH)D more accurately reflect body stores of vitamin D. Measurement of this metabolite should therefore be performed when vitamin D deficiency is suspected.

Measurements of 1,25(OH)₂D₃ should be reserved for cases when excessive or impaired 1-hydroxylation is suspected. High 1,25(OH)₂D₃ levels can be seen in sarcoidosis, lymphomas, Williams’ syndrome, and intoxication with 1-hydroxylated metabolites (see “Hypercalcemic Disorders”). Impaired 1-hydroxylation can contribute to the hypocalcemia occurring in patients with renal dysfunction, oncogenic osteomalacia, and hereditary defects of vitamin D metabolism (see “Hypocalcemic Disorders”).
The clinical spectrum of primary hyperparathyroidism was changed dramatically in the early 1970s by the introduction of routine multichannel serum chemistry.

Modern Primary Hyperparathyroidism

The bone disease osteitis fibrosa cystica was first described by von Recklinghausen in 1891, but the etiologic link between this disease and parathyroid neoplasms was not established until 1925, when Mandl observed clinical responses to removal of a parathyroid adenoma from a young man with severe bone disease. In early clinical descriptions of primary hyperparathyroidism, the disease emerged as a distinctly uncommon disorder with significant morbidity and mortality, in which nearly all affected patients had radiographically significant or symptomatic skeletal or renal involvement, or both.

CLINICAL DISORDERS

HYPERCALCEMIC DISORDERS

Parathyroid-Dependent Hypercalcemia

It is useful to delineate two categories of hypercalcemia: (1) that associated with dysfunction of the parathyroid cell and (2) that in which hypercalcemia occurs despite appropriate parathyroid suppression. This distinction is particularly useful clinically, because it emphasizes the centrality of the PTH assay in the diagnostic approach to the hypercalcemic patient. Abnormal parathyroid glands are associated with hypercalcemia in three settings: (1) primary hyperparathyroidism, (2) familial hypocalciuric hypercalcemia (FHH), and (3) lithium-induced hypercalcemia.

Primary Hyperparathyroidism

In primary hyperparathyroidism, a primary abnormality of parathyroid tissue leads to inappropriate secretion of PTH. (In contrast, increased secretion of PTH that is an appropriate response to hypocalcemia is called secondary hyperparathyroidism.) The inappropriately high serum concentration of PTH in primary hyperparathyroidism, in turn, sustains excessive renal calcium reabsorption, phosphaturia, and 1,25(OH)₂D₃ synthesis as well as increased bone resorption. These actions of PTH produce the characteristic biochemical phenotypic features of hypercalcemia and hypophosphatemia, loss of cortical bone, hypercalciuria, and the various clinical sequelae of chronic hypercalcemia. Primary hyperparathyroidism results most often in 75% to 80% of the cases from the occurrence of one or more adenomas in previously normal parathyroid glands, although in 20% of the cases, diffuse hyperplasia of all parathyroid glands may be present or, rarely, parathyroid carcinoma may be found (less than 1% to 2%).

Rafai et al.

Classical Primary Hyperparathyroidism

The skeletal involvement in "classical" primary hyperparathyroidism reflects a striking and generalized increase in osteoclastic bone resorption, which is accompanied by fibrovascular marrow replacement and increased osteoblastic activity. The radiographic appearance (Fig. 26-23) features generalized demineralization of bone, with coarsening of the trabecular pattern (due to osteoclastic resorption of the smaller trabeculae); characteristic subperiosteal resorption, often most evident in the phalanges of the hands, which gives an irregular, serrated appearance to the outer, subperiosteal cortex and may progress to extensive cortical resorption; bone cysts, usually multiple, which contain a brownish serous or mucoid fluid, tend to occur in the central medullary portions of the shafts of the metacarpals, ribs, or pelvis, and may expand into and disrupt the overlying cortex; osteolastomas, or "brown tumors," composed of numerous multinucleated osteoclasts ("giant cells") admixed with stromal cells and matrix, which are found most often in trabecular portions of the jaw, long bones, and ribs; and pathologic fractures.

The skull may exhibit a finely mottled, "salt-and-pepper" radiographic appearance, with loss of definition of the inner and outer cortices. Dental radiographs typically show erosion or disappearance of the lamina dura due to subperiosteal resorption, often with extension into the adjacent mandibular bone. The erosion and demineralization of cortical bone may lead to radiographic disappearance of some bones, most notably the tufts of the distal phalanges of the hands, the inferolateral cortex of the distal third of the clavicles, the distal ulna, the inferior margin of the femoral neck and pubis, and the medial aspect of the proximal tibia. The clinical correlates of these changes may include achon bone pain and tenderness, "bowing" of the shoulders, kyphosis and loss of height, and collapse of lateral ribs and pelvis with "pigeon breast" and triradiate deformities, respectively.

The renal manifestations of classical severe primary hyperparathyroidism include recurrent calcium nephrolithiasis, nephrocalcinosis, and renal functional abnormalities that range from impaired function to end-stage renal failure. Associated signs and symptoms include recurrent flank pain, polyuria, and polydipsia. No unique features of the stone disease in primary hyperparathyroidism serve to distinguish it from that associated with other, more common causes of kidney stones. The stone disease more often may be recurrent and severe, and in some patients, the stones may be composed entirely of calcium phosphate, instead of the pure oxalate or mixtures of oxalate and phosphate more commonly encountered in other disorders. In patients diagnosed before 1965, the frequency with which nephrolithiasis complicated primary hyperparathyroidism was as high as 60% to 80% (the frequency is currently less than 25%), yet in studies of unselected patients conducted throughout the past 50 years, primary hyperparathyroidism has accounted for fewer than 5% of all calcium kidney stones.

Nephrocalcinosis refers to the presence of bilateral, extensive but minute calcifications that are evident on plain abdominal radiographs, usually in the renal pyramids and medullary regions, and that correlate with the presence of deposits of calcium in the epithelium of the renal tubules. In classical severe primary hyperparathyroidism, nephrocalcinosis was observed with roughly one-third the frequency of symptomatic stone disease, although it may occur in the absence of stones.

Other clinical features that have been reported in association with classical severe primary hyperparathyroidism are conjunctival calcifications, band keratopathy, hypertension (50%), gastrointestinal signs and symptoms (anorexia, nausea, vomiting, constipation, or abdominal pain), peptic ulcer disease, and acute or chronic pancreatitis. The issue of whether primary hyperparathyroidism increases the risk for peptic ulcer disease and pancreatitis remains controversial. Although hyperparathyroidism is associated with a higher risk of hypertension, successful parathyroidectomy has not been shown to correct the hypertension.

Signs and symptoms in primary hyperparathyroidism may result from the involvement of bone (fracture, bone pain) or kidneys (renal colic, renal failure), peptic ulcer disease, and acute or chronic pancreatitis.

The renal manifestations of classical severe primary hyperparathyroidism include recurrent calcium nephrolithiasis, nephrocalcinosis, and renal functional abnormalities that range from impaired function to end-stage renal failure. Associated signs and symptoms include recurrent flank pain, polyuria, and polydipsia. No unique features of the stone disease in primary hyperparathyroidism serve to distinguish it from that associated with other, more common causes of kidney stones. The stone disease more often may be recurrent and severe, and in some patients, the stones may be composed entirely of calcium phosphate, instead of the pure oxalate or mixtures of oxalate and phosphate more commonly encountered in other disorders. In patients diagnosed before 1965, the frequency with which nephrolithiasis complicated primary hyperparathyroidism was as high as 60% to 80% (the frequency is currently less than 25%), yet in studies of unselected patients conducted throughout the past 50 years, primary hyperparathyroidism has accounted for fewer than 5% of all calcium kidney stones.

Nephrocalcinosis refers to the presence of bilateral, extensive but minute calcifications that are evident on plain abdominal radiographs, usually in the renal pyramids and medullary regions, and that correlate with the presence of deposits of calcium in the epithelium of the renal tubules. In classical severe primary hyperparathyroidism, nephrocalcinosis was observed with roughly one-third the frequency of symptomatic stone disease, although it may occur in the absence of stones.

Other clinical features that have been reported in association with classical severe primary hyperparathyroidism are conjunctival calcifications, band keratopathy, hypertension (50%), gastrointestinal signs and symptoms (anorexia, nausea, vomiting, constipation, or abdominal pain), peptic ulcer disease, and acute or chronic pancreatitis. The issue of whether primary hyperparathyroidism increases the risk for peptic ulcer disease and pancreatitis remains controversial. Although hyperparathyroidism is associated with a higher risk of hypertension, successful parathyroidectomy has not been shown to correct the hypertension.

Signs and symptoms in primary hyperparathyroidism may result from the involvement of bone (fracture, bone pain) or kidneys (renal colic, renal failure), peptic ulcer disease, pancreatitis, or hypercalcemia per se (weakness, apathy, depression, polyuria, constipation, coma). The presence and severity of neuropsychiatric symptoms, in particular, correlate poorly with the serum calcium concentration, although few patients with severe hypercalcemia are entirely asymptomatic. Elderly persons are most likely to exhibit such symptoms. A peculiar neuromuscular syndrome, first described in 1949 but rarely encountered now, includes symmetrical proximal weakness and gait disturbance, with muscle atrophy, characteristic electromyographic abnormalities, generalized hyperreflexia, and tongue fasciculations.

Modern Primary Hyperparathyroidism

The clinical spectrum of primary hyperparathyroidism was changed dramatically in the early 1970s by the introduction of routine multichannel serum chemistry.
screening, which unearthed a large population of patients with previously unsuspected asymptomatic disease. In Rochester, Minnesota, for example, the annual incidence of the disease increased abruptly from 0.15 to 1.12 per 1000 persons between the prescreening era (1965/1974) and 1975, the year after routine screening was introduced. The peak incidence occurs in the sixth decade of life, and the disease rarely is encountered in patients younger than 15 years of age. It is two to three times more common in women, who are slightly older than men at diagnosis.

Annual incidence rates, widely reported to be 0.1 to 0.3 per 1000 persons in the wake of this surge of ascertainment in Europe and the United States, appear to have declined substantially in the past decade to levels as low as 0.04 per 1000. This decrease may represent a true decline in disease incidence or may simply be the residual effect of "sweeping" the population of prevalent subclinical disease over the past three decades.

Ascertainment of mild or asymptomatic disease may decline even further in the future because of prevalent economic disincentives to routine serum chemistry screening in the primary care setting. On the other hand, insistence upon overt hypercalcemia or hypernatremia as a diagnostic criterion may underestimate the true incidence of the disease. For example, when serum calcium and immunoreactive PTH (iPTH) were measured in a large population of Swedish women undergoing routine mammography screening, the prevalence of asymptomatic or suspected primary hyperparathyroidism, defined by criteria that included the combination of high-normal serum calcium plus elevated or high-normal iPTH, was 2.1%. Two thirds of these women (72 of 109) were normocalcemic (10.0 to 10.4 mg/dL), yet bone density was reduced in the group as a whole, and the disease was confirmed in 98% of the 61 who underwent surgery.

It is not surprising, given that primary hyperparathyroidism now usually is diagnosed incidentally, that few patients are found to have overt signs or symptoms of the classical disease and thus are considered to be "asymptomatic." For example, only 2% of patients with primary hyperparathyroidism residing in Olmsted County, Minnesota, and only 17% of 121 patients studied at an academic referral center in New York City had classical disease symptoms. In most of these, the relevant symptom was urolithiasis. Many clinicians argue, however, that most patients regarded as having "asymptomatic" primary hyperparathyroidism and only minimally elevated serum calcium actually suffer from various neuropsychiatric or other symptoms that may abate following curative surgery. However, these symptoms, which include fatigability, weakness, forgetfulness, depression, somatization, polydipsia, polyuria, and bone and joint pain, are common in otherwise normal persons.

The difficulties in designing appropriately controlled studies to determine whether these symptoms can be confidently ascribed to the parathyroid disorder are well described. Although as yet unresolved, this is a critical issue, as the advent of less invasive operative approaches and concerns regarding fracture, cancer, and mortality risk have lowered the threshold for consideration of surgery in many patients with the disease (see later on). Throughout this chapter, "asymptomatic primary hyperparathyroidism" refers to patients who lack signs or symptoms of the classical disease, whether or not they experience any of the subtle symptoms mentioned previously.

The natural history of untreated asymptomatic primary hyperparathyroidism, as currently detected, remains incompletely understood. Few patients seem to progress from this state of health, as measured by severe elevations of serum or urinary calcium, appearance of renal dysfunction or nephrocalcinosis, or worsening osteopenia, over many years of observation. On the other hand, an excess risk of mortality, mainly from cardiovascular disease, has been noted during extended follow-up of large cohorts of patients with chronic hypercalcemia (and presumed primary hyperparathyroidism) identified by population health screening in Sweden, and similar observations have been made during extended follow-up of postsurgical patients with hyperparathyroidism. Associations of hypertension, hyperuricemia, and glucose intolerance with primary hyperparathyroidism have been implicated, together with hypercalcemia per se, as contributors to this elevated risk. Abnormal cardiac calcification and left ventricular hypertrophy (reversible after successful parathyroidectomy) have been reported in primary hyperparathyroidism as well. Increased cardiovascular mortality may be a feature only of severe hyperparathyroidism, as it was restricted to patients with serum calcium values in the highest quartile in the Olmsted County study, which otherwise showed an overall decreased risk of death. A 40% excess risk of malignancy also was reported among 4163 Swedish patients who had undergone surgery more than a year earlier for (presumably symptomatic) primary hyperparathyroidism. It has been argued that these increased risks for mortality and malignancy, even if confirmed, may apply only to persons with primary hyperparathyroidism that is more severe than the "asymptomatic" version typically encountered today.

Abnormalities of bone in modern, mild primary hyperparathyroidism are far subtler than those associated with the classical disease. Histologically, the rate at which new bone remodeling cycles are activated is increased. Because the phase of remodelling before calcium deposition is so much longer than the initial resorptive phase, such an increase in remodeling rate inevitably increases the effective volume of the remodeling space and thus the porosity of bone. Depending on the rate and extent of the accompanying increase in osteoblastic activity and the resulting local balance between net bone formation and resorption, mineralized bone volume may decrease further, remain stable, or even increase (despite an increased remodeling space). For reasons not yet understood, the balance achieved between increased resorption and formation of bone in primary hyperparathyroidism depends not only on the severity of the hyperparathyroidism but also on skeletal location. Thus, net resorption of endosteal bone may predominate in cortical bone sites, whereas net apposition of mineral may occur in trabecular bone (Fig. 26-24).

In mild primary hyperparathyroidism, osteopenia generally is not evident radiographically, although bone mineral density may be reduced, particularly at sites of predominantly cortical bone such as the mid-radius, by as much as 10% to 20%. The mass of trabecular or cancellous bone, as represented in the vertebral bodies, is preferentially preserved and often is normal. A curious finding is that the reduced cortical bone density at the forearm is not improved by successful parathyroidectomy, whereas density at trabecular bonerich sites such as the hip and spine may increase by 10% to 15% over several years postoperatively.

The critical issue of whether fracture risk is increased in patients with primary hyperparathyroidism was addressed by a retrospective analysis of fracture incidence within a cohort of 407 residents of Rochester, Minnesota, who were diagnosed with the disease between 1965 and 1992. Compared with the expected age-adjusted and sex-adjusted rates of incident fractures in that community, the relative risk among persons with hyperparathyroidism was significantly elevated for fractures of the vertebrae (3.2-fold), distal forearm (2.2-fold), and ribs (2.7-fold), although not for fractures of the hip (1.4-fold). Overall risk of fracture at any site was significantly increased as well (1.3-fold) and was as high in persons in whom primary hyperparathyroidism was diagnosed incidentally (following the institution of automated chemistry screening in 1974) as in those in whom the disorder was diagnosed before that time. These results are consistent with those of several previous studies involving smaller cohorts of patients and, absent data from an appropriately controlled prospective study, strongly support the conclusion that patients with primary hyperparathyroidism should be considered to be at increased risk for fracture. This presumably is true of those with both symptomatic

Kidney stones now are reported in only 10% to 25% of patients with primary hyperparathyroidism, although some degree of renal dysfunction, either a significant reduction in creatinine clearance or impaired concentrating or acidifying ability, may be found in up to a third of those with asymptomatic disease. As with the reduction in cortical bone mineral density, these renal abnormalities are not progressive in a majority of affected patients. The association of kidney stones with primary hyperparathyroidism generally is viewed as an indication for parathyroidectomy, however, because successful surgery usually prevents further symptomatic stone disease. Yet it is not possible at present to confidently predict, from biochemical measurements in blood or urine, which asymptomatic patients with primary hyperparathyroidism will go on to develop new stone disease. Stone formers are more likely to be hypercalcicuric than not, but less than one third of hypercalcicuric

[Image 26-24: Iliac crest biopsy specimens from a patient with primary hyperparathyroidism (left) and a normal control (right), viewed by scanning electron microscopy. Note the thin cortices and contrasting preservation of trabecular bone in the patient. (From Stein M, Silverberg SJ, Shane E, et al: The histomorphometry of bone in primary hyperparathyroidism: preservation of cancellous bone structure. J Clin Endocrinol Metab 1990; 70:930938.)]
patients with hyperparathyroidism actually develop stones. As noted later on, however, marked hypercalcemia (calcium levels of >400 mg/day) generally is viewed as an indication for surgery in otherwise asymptomatic patients.

Etiology and Pathogenesis

Parathyroid adenomas are caused by mutations in the DNA of parathyroid cells; these mutations confer a proliferative or survival advantage for affected cells over their normal neighbors. As a consequence of this advantage, the descendants of one particular parathyroid cell, a clone of cells, undergo clonal expansion to produce an adenoma.

Multiple chromosomal regions are missing in the parathyroid cells of individual parathyroid adenomas. These genetic deletions probably reflect the deletion of tumor suppressor genes. These chromosomal loci include portions of chromosome 1p-pter (in 40% of adenomas), 6q (in 32% of adenomas), 15q (in 30% of adenomas), and 11q (in 25% to 30% of adenomas). Many of the 11q deletions are associated, in the undeleted chromosome 11, with mutations in the gene encoding the tumor-suppressor factor menin, the gene mutated in multiple endocrine neoplasia type 1 (MEN 1). Thus, this gene is also involved commonly in somatic mutations in patients with sporadic parathyroid adenomas. The widespread presence of somatic mutations in sporadic parathyroid adenomas, which are detectable only because large numbers of cells in any one tumor contain the same deletion, constitutes the strongest evidence that parathyroid adenomas are clonal expansions of mutant cells.

One parathyroid proto-oncogene, the PRAD 1 or cyclin D1 gene, has been identified. This gene was discovered at the breakpoint of an inversion on chromosome 11 in a parathyroid adenoma. This inversion led to the juxtaposition of the PTH gene's regulatory region and the DNA encoding cyclin D1. As a consequence, the cyclin D1 gene was overexpressed. Cyclin D1 is an important regulator of the transition from the G1 phase of the cell cycle (which follows mitosis) to the S phase (associated with DNA synthesis) and is mutated or amplified in a wide variety of malignancies. Cyclin D1 is overexpressed in about 20% of parathyroid adenomas, though cyclin D1 gene rearrangements have been documented in only 5% of adenomas. Portions of chromosomes 16p and 19p are amplified in a subset of parathyroid adenomas; presumably, the amplified regions contain uncharacterized proto-oncogenes.

As expected for a disease caused by mutations in DNA, parathyroid adenomas occur more frequently in patients who underwent neck irradiation decades earlier, with greater radiation exposure leading to higher risk. Most patients have no definite history of exposure to specific mutagens, however. An intriguing clue that abnormalities of vitamin D physiology may predispose to primary hyperparathyroidism comes from the observation that patients with parathyroid adenomas are more likely than others to inherit a particular allele of the vitamin D receptor gene. These patients have tumors with particularly low levels of mRNA encoding the vitamin D receptor.

The cause of sporadic primary parathyroid hyperplasia is unknown. The known stimulus for parathyroid cell proliferation-low levels of blood calcium or 1,25(OH)2D3—is not present in this disease. Presumably, some other stimulus outside the parathyroid glands or a genetic abnormality present in all four parathyroid glands leads to inappropriating the transient factor menin, the gene mutated in multiple endocrine neoplasia type 1 (MEN 1). As noted later on, however, marked hypercalcemia (calcium levels of >400 mg/day) generally is viewed as an indication for surgery in otherwise asymptomatic patients.

An increase in cell number is not the only abnormality in primary hyperparathyroidism. The ability of the normal parathyroid cell to suppress PTH secretion in response to hypercalcemia might be expected to protect the individual from sustained hypercalcemia, even if the number of parathyroid cells increased moderately. Unfortunately, parathyroid cells from adenomas respond to changes in extracellular calcium with smaller than normal increases in intracellular calcium, and the amount of calcium-sensing receptor protein on the cell surface is reduced.

Inherited Primary Hyperparathyroidism

Although uncommon, inherited forms of primary hyperparathyroidism are clinically important for several reasons. The management of the parathyroid tumors found in familial parathyroid syndromes often differs from that of sporadic primary hyperparathyroidism. Furthermore, extra-parathyroidal manifestations of inherited syndromes may need treatment, and awareness of familial clustering should prompt systematic family screening.

MEN 1 is caused by inactivating mutations in the tumor suppressor gene encoding menin. Menin is a ubiquitously expressed transcription factor. Although MEN 1 includes tumors of the parathyroid, anterior pituitary, and pancreatic islets, the parathyroid tumors are far more prevalent than the others; 95% of affected patients eventually develop hyperparathyroidism. Most of the parathyroid tumors harbor mutations in both copies of the menin gene; one mutation is inherited and the second occurs in the parathyroid cell whose progeny form the tumor.

The onset of hypercalcemia occurs in the second and third decades of life, though occasional patients present in the first decade. Hypercalcemia never presents at birth or in infancy. The disease involves all four parathyroid glands, although the involvement can be asymmetrical and apparently asynchronous. Apart from the
common complicating feature is that hypercalcemia can dramatically increase the gastrin levels and symptomatology of patients who also have gastrinomas.

Treatment of the parathyroid disease in this setting can greatly simplify the management of the gastric hyperacidity. After parathyroid surgery, hypoparathyroidism and recurrent hyperparathyroidism are more common than in other forms of hyperparathyroidism. The timing and type of surgery are therefore more complicated issues in inherited primary hyperparathyroidism. Most authorities agree that parathyroid disease recurs eventually, if fewer than three glands are removed. Some surgeons prefer subtotal parathyroidectomy, whereas others prefer total parathyroidectomy with forearm implantation of a small amount of parathyroid tissue.

Parathyroid disease is usually late and infrequent (5% to 20%) occurrence in MEN 2a, a disease defined by the clustering of medullary carcinoma of the thyroid, pheochromocytoma, and hyperparathyroidism. In some families, hyperparathyroidism is more common; however, these families have the same mutations in the RET gene that are found in families without frequent hyperparathyroidism. Both parathyroid hyperplasia and adenoma have been noted at surgery. Because asymptomatic parathyroid hyperplasia has been noted at the time of thyroid surgery, a progression from hyperplasia to adenoma in MEN 2a has been suggested. The approach to diagnosis and treatment of hyperparathyroidism is similar to that in sporadic primary hyperparathyroidism, but hyperplasia is more frequently the underlying disorder. The pathogenesis of the hyperparathyroidism is uncertain, but the RET gene, mutated in virtually all cases of MEN 2a, is expressed in parathyroid cells, so abnormal net expression in parathyroid cells may directly cause parathyroid tumorigenesis. Hyperparathyroidism does not occur in MEN 2b, the variant associated with mucosal neuromas.

Several other distinct autosomal dominant syndromes of primary hyperparathyroidism have been characterized. Patients with hereditary isolated primary hyperparathyroidism present with parathyroid tumors that can be multiple and occasionally are malignant. No other endocrine glands are abnormal in affected families, and the disease maps to neither the MEN 1 nor MEN 2a locus. Patients with hereditary hyperparathyroidism jaw tumor syndrome present with parathyroid adenomas that are usually cystic and with fibrous jaw tumors that are unrelated to the hyperparathyroidism. Parathyroid cancer, Wilms’ tumor, and polycystic renal disease also have occurred in affected families. This syndrome maps to chromosome 1q21-q31. Occasional families with apparently isolated familial hyperparathyroidism have inherited mutations in the MEN 1 gene, but most such families do not.

Management of Primary Hyperparathyroidism

The strategy for management of primary hyperparathyroidism has evolved in parallel with the changing presentation of the disease. The only opportunity for permanent cure is surgical removal of the abnormal gland(s), an approach that clearly was appropriate for virtually all patients in whom the classical, severe form of the disease was diagnosed four to five decades ago and which still is the treatment of choice for those patients who do present with recurrent kidney stones, nephrocalcinosis, clinically overt bone disease, or severe hypercalcemia. In contrast, the choice of surgical versus medical management for patients with asymptomatic primary hyperparathyroidism remains an open and hotly debated question. Investigators who favor surgery point to the expected improvement in bone mineral density (at the hip and spine) and reversibility of left ventricular hypertrophy following successful surgical intervention; evidence of increased risk for fracture, cardiovascular mortality, malignancy, and neuropsychiatric symptoms associated with primary hyperparathyroidism; and the recent successful development of effective minimally invasive surgical procedures (see later on).

Investigators who favor an observational approach emphasize the evidence for lack of disease progression in most asymptomatic patients; the small but finite risk of surgical failure and postoperative complications; the probability that excess mortality and cancer risks documented in patients with relatively severe disease may not apply to those with mild, asymptomatic primary hyperparathyroidism; the difficulty in assigning vague neuropsychiatric symptoms to the parathyroid disorder; the lack of evidence (or negative evidence) that patients with hypertension or possibly increased risk of cancer, fracture, or cardiovascular mortality, or who already have such disease, are benefited by successful parathyroidectomy; and the availability of sensitive techniques for monitoring disease status in nonoperated patients.

Unfortunately, no prospective studies have compared outcomes in patients with asymptomatic primary hyperparathyroidism randomly assigned to surgery with those in patients assigned to medical management. Furthermore, no large trials of differing strategies in medical management of the disease have been conducted. Surgical series generally reflect outcomes in patients prescreened to receive interventional treatment; thus, results may not be readily extrapolated to those with mild, asymptomatic disease. Evidence that the true incidence of the disease may be declining, and that economic forces may increasingly limit detection to more symptomatic cases of the disease, may indicate that yet another change in the clinical character of the disease is in the offing. For all of these reasons, opinion in this area is rapidly evolving, and all recommendations thus should be considered provisional.

One set of such provisional recommendations was issued by a National Institutes of Health (NIH)-sponsored Consensus Development Conference held in 1990. The major conclusion of the Conference group was that although surgery always should be considered an appropriate option, many patients with asymptomatic primary hyperparathyroidism can be safely monitored without surgery. These patients were defined as those who lacked “significant bone, renal, gastrointestinal, or neuropsychiatric symptoms typical of primary hyperparathyroidism” and who also did not meet the other criteria listed in Table 26-1. Such patients account for at least 50% of persons who currently present with primary hyperparathyroidism.

In other Consensus Conference recommendations, surgery may be preferable if the patient desires surgery even when asymptomatic, if the probability of consistent monitoring seems low, if concomitant illness seems likely to complicate management or obscure significant disease progression, or if the patient is relatively young (under 50 years of age). These recommendations reflect the absence of reliable information about the natural history of the disease over many decades of follow-up, as well as the cumulative cost of medical monitoring, which begins to exceed that of surgery by 5 to 10 years. Conversely, age alone is not a contraindication to parathyroidectomy, as the procedure has been accomplished with excellent results, with a perioperative mortality of 1% to 3%, in large numbers of appropriately selected patients older than 75 years of age. Because hypertension is not thought to be a feature of mild primary hyperparathyroidism, and because hypertension generally is not corrected by parathyroidectomy, hypertension is not viewed as an indication for surgery.

<table>
<thead>
<tr>
<th>TABLE 26-1 -- Indications for Surgery in Primary Hyperparathyroidism</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Overt clinical manifestations of primary hyperparathyroidism</td>
</tr>
<tr>
<td>- Radiographic nephrolithiasis or otherwise documented kidney stone(s)</td>
</tr>
<tr>
<td>- Reduced creatinine clearance (not otherwise explained)</td>
</tr>
<tr>
<td>- Radiographically evident hyperparathyroid bone disease</td>
</tr>
<tr>
<td>- Classical hyperparathyroid neurovascular disease</td>
</tr>
<tr>
<td>- Symptoms attributable to hypercalcemia per se</td>
</tr>
<tr>
<td>- Previous episode of life-threatening hypercalcemia</td>
</tr>
<tr>
<td>b. Severe calcium concentration greater than 12 mg/dL (2.99 mM)</td>
</tr>
<tr>
<td>c. Urinary calcium excretion greater than 400 mg/day (9.98 mmol/day)</td>
</tr>
</tbody>
</table>
Although the Consensus Conference recommendations provide a useful framework for decision making, supporting data from large clinical trials are lacking. In a series of 52 asymptomatic patients selected for nonoperative management mainly on the basis of the Consensus Conference criteria and whose course was followed for 10 years, approximately 25% developed one or more new indications for surgery (Table 26-1). Patients who do not meet the Consensus Conference criteria for surgery may nevertheless experience the same postsurgical increase in bone density as occurs in those who do. Some investigators have emphasized that evidence of baseline vertebral osteopenia, an unusual finding in primary hyperparathyroidism, should be added to the original Consensus Conference criteria for surgery (see Table 26-1) and that surgery also should be considered for menopausal women who exhibit vertebral bone loss in the setting of primary hyperparathyroidism. A common dilemma is the inability to ascertain whether vague but troublesome symptoms such as fatigue, lethargy, weakness (without objective muscle weakness), or depression are due to hyperparathyroidism and thus qualify as "significant" in the context of the decision regarding surgery. Most clinicians do not routinely recommend parathyroidectomy on the basis of such symptoms alone, although dramatic responses to surgery are occasionally seen. With the availability of improved, minimally invasive surgical approaches, the threshold for considering surgery for patients who are significantly disabled by such symptoms clearly is lower now than in the past. Some authorities have advocated, in selected cases, a limited trial of medical therapy to reduce serum calcium (i.e., strict dietary calcium restriction, estrogen, oral phosphate, or bisphosphonate; see later), thereby attempting to predict the symptomatic response to surgical cure. The severity of such symptoms generally cannot be shown to correlate with serum calcium concentration, however, and may be related to other factors, such as blood PTH concentration.

Medical Monitoring of Primary Hyperparathyroidism

The NIH Consensus Conference recommended that patients be followed carefully, at least semiannually initially and at longer intervals if stable, for appearance of symptoms; appearance of adverse effects on blood pressure and serum or urinary calcium or creatinine; review of annual abdominal radiographs; and serial determination of bone mineral density at 1- or 2-year intervals. The most appropriate bone densitometric site was considered to be one that reflects changes in cortical bone (i.e., distal forearm or total body), although the importance of monitoring vertebral bone density, as well, has been emphasized more recently.

Patients undergoing nonoperative medical management must be cautioned to maintain adequate hydration, to avoid diuretics and prolonged immobilization, and to seek prompt medical attention in the event of illnesses accompanied by significant vomiting or diarrhea. Dietary calcium probably should not exceed the recommended daily allowance (RDA) of 800 mg/day, even though short-term studies in highly selected patients with elevated serum 1,25(OH)₂D₃ levels have demonstrated that a high-calcium diet may reduce serum 1,25(OH)₂D₃ and PTH, albeit at the expense of a mild increase in serum and urinary calcium.

Estrogen therapy may be a consideration in postmenopausal women with mild primary hyperparathyroidism or in those who are symptomatic but yet refuse or cannot safely undergo surgery. Estrogens may reduce serum calcium and phosphorus, and urinary calcium and hydroxyproline, and histologic evidence of bone resorption in women with primary hyperparathyroidism, although serum PTH remains elevated and serum 1,25(OH)₂D₃ increases, owing mainly to increased concentrations of the vitamin D binding protein. Progestins may exert similar effects on serum calcium and may slow bone loss in elderly women with primary hyperparathyroidism, but concern over their adverse effects on blood lipids have limited use of these agents in this population. Selective estrogen receptor modulators such as tamoxifen or raloxifene may prove useful in postmenopausal women with osteopenia due to primary hyperparathyroidism, but a beneficial effect has not yet been shown.

Oral phosphate therapy has been advocated for use in occasional patients with primary hyperparathyroidism who have failed or refused surgery and who may have recurrent calcium kidney stones or other serious symptoms. Phosphate in this setting may act by inhibiting both renal synthesis of 1,25(OH)₂D₃ and osteoclastic bone resorption, by promoting precipitation of calcium into bone and soft tissues, and, in the gut, by binding intraluminal calcium and impairing its absorption. Oral phosphate does reduce serum and urinary calcium as well as serum 1,25(OH)₂D₃, but it increases serum PTH, and the long-term effect on bone is unknown.

Bisphosphonates have been employed successfully in the urgent therapy of hypercalcemia due to primary hyperparathyroidism, but their use in chronic management of mild primary hyperparathyroidism has been limited by several adverse effects, including stimulation of PTH secretion and a lowering of the renal phosphate threshold, which aggravates hypophosphatemia and may augment renal 1,25(OH)₂D₃ synthesis. Although not yet available for routine clinical use, calcimimetic drugs capable of binding to and activating the parathyroid calcium-sensing receptor are in development and may offer a new avenue for control of hyperparathyroidism.

Surgical Treatment of Primary Hyperparathyroidism

Parathyroidectomy is a safe and highly effective approach to definitive treatment of primary hyperparathyroidism. The most serious potential complications of parathyroid surgery—cord palsy and permanent hypoparathyroidism—occur after less than 1% and 4%, respectively, of procedures performed by highly skilled surgeons, although the rates for these complications can be much higher with less experienced operators. Such complications occur most often in patients who require subtotal parathyroid resection for hyperplasia or resection of carcinoma. The surgical cure rate for primary hyperparathyroidism in the best hands is at least 95%. Apart from operator inexperience, the usual cause of initial surgical failure is the presence of either unrecognized (often very asymmetrical) parathyroid hyperplasia or ectopic parathyroid tissue (i.e., intrathyroidal, undescended, retroesophageal, or mediatinal glands). Patients who do not meet the Consensus Conference criteria for surgery may nevertheless experience the same postsurgical increase in bone density as occurs in those who do.
usually arises in unresected hyperplastic glands, but rarely it may be due to parathyroid carcinoma, to a second adenoma, or to a multicentric or military "parathyromatosis" engendered by inadvertent local seeding of parathyroid tissue into the neck during previous parathyroid surgery. Until recently, there was broad agreement that the best approach is a bilateral neck exploration in which all four parathyroids are identified and all enlarged glands removed. With this procedure, preoperative parathyroid localization studies before initial cervical exploration are superfluous, as the positive predictive value of even the best technique (technetium Tc 99m sestamibi scanning) falls well short of the success rate of experienced surgeons unaided by previous imaging. Some surgeons prefer to biopsy all identified glands, even though this approach risks a 20% to 40% incidence of transient postoperative hypocalcemia, whereas others report excellent results without routine biopsy of nonenlarged glands. In the event that one gland is not identified after extensive exploration, ipsilateral thyroid lobectomy commonly is performed, as the frequency of intrathyroidal parathyroid glands is approximately 5%.

With the advent of preoperative technetium Tc 99m sestamibi scanning, which can accurately localize 80% to 90% of the single adenomas that account for 75% to 85% of cases, there has been renewed interest in performance of directed unilateral explorations, which reduce operative and recovery room time, minimize the number of frozen section required, are associated with significantly fewer postoperative complications, and can more readily be performed using minimally invasive techniques (including local anesthesia and intravenous sedation) that enable same-day discharge. Sestamibi scanning also can identify the occasional medistinal adenoma and thereby direct the surgery away from neck exploration. On the other hand, because the sensitivity of sestamibi scanning ranges from 75% to 90% and the technique is least reliable in the presence of multiglandular disease (hyperplasia or double adenomas), the test may falsely localize an adenoma or miss the presence of bilateral disease in 10% to 20% of patients. To reduce this failure rate, which is unacceptably high in comparison with that of bilateral exploration, supplemental preoperative ultrasonic imaging (with or without needle biopsy) has been employed, and rapid intraoperative PTH assays have been developed to verify successful excision.

Because the half-life of intact PTH in blood is very short (<2 minutes), a decline of 50% or more from baseline within 10 minutes or so can signal successful removal of all hyperfunctioning parathyroid tissue. This approach has worked well in patients with single adenomas but can be misleading in those with multiglandular disease unless strict criteria for cure are applied (i.e., >70% decline in plasma calcium at 20 minutes). No surgical series has analyzed a sufficient number of patients with hyperparathyroidism to permit rigorous definition of the role of intraoperative PTH measurements.

Another adjunct to minimally invasive parathyroidectomy has been the use of a hand-held gamma probe intraoperatively, both to map the location of abnormal glands and to verify their removal, following immediate preoperative imaging technetium Tc 99m using sestamibi. The challenging logistics of arranging the sequence of injection, imaging, and surgery necessary for success with this technique have limited its use to a few specialized centers, however.

At present, preoperative imaging enables consideration of a minimally invasive unilateral parathyroidectomy in approximately 70% of those patients thought preoperatively to have sporadic primary hyperparathyroidism due to a solitary adenoma. The ultimate dominant of this approach will depend on the balance between advantages (reduced operative and recovery time, fewer ancillary procedures, and lower hospital charge) and disadvantages (the added costs of the imaging and intraoperative PTH assays and the possibly greater demands for surgical skill and experience). Alternatively, successful application of minimally invasive techniques to bilateral explorations has begun. In one center, for example, 97% of patients with primary hyperparathyroidism currently undergo bilateral exploration under local anesthesia, with completion of the procedure in less than 1 hour and discharge to home within 6 hours.

In patients found or suspected to have multiglandular hyperplasia, several different approaches have been recommended, including removal of all but approximately 30 to 50 mg of parathyroid tissue ("3/g-land parathyroidectomy"), removal of all enlarged glands (but at least two) with biopsy of the remaining normal-sized glands, and total resection of all parathyroid tissue from the neck followed by autotransplantation of gland fragments to the forearm, with cryopreservation of remaining tissue. A concerted effort is made to search for a possible fifth gland, and any unresected (or transplanted) parathyroid tissue is marked with clips or sutures to facilitate identification in the event that reoperation becomes necessary. Great care must be taken by both the surgeon and the pathologist to describe the origin of, and label, all resected putative parathyroid tissue, particularly in parathyroid hyperplasia, as one of the most common handicaps facing the surgeon, should a subsequent operation become necessary, is inadequate knowledge of the original surgical findings and pathology. In contrast to initial cervical exploration, preoperative localization should be attempted routinely in those few patients who require reoperation for failed surgery, and intraoperative PTH determinations may be helpful in guiding the surgeon during these often difficult procedures.

The incidence of parathyroid carcinoma in primary hyperparathyroidism is less than 1%, but this possibility should be strongly considered in patients with unusually severe hyperparathyroidism, a palpable neck mass, hoarseness, evidence of local invasion at surgery, or recurrent hypercalcemia (see later on). Even so, parathyroid carcinoma rarely is suspected preoperatively and often eludes diagnosis at the time of initial surgery. When the disease is recognized, vigorous attempts should be made to remove all of the tumor on bloc. The incidence of local recurrence approaches 50%, however, and distant metastases, particularly to lung, may be heralded by recurrent, severe hyperparathyroidism.

The immediate postoperative management of parathyroidectomy focuses on establishing the success of the surgery and monitoring the patient closely for symptomatic hypercalcemia and for uncommon but potentially serious acute complications such as bleeding, vocal cord paralysis, or laryngospasm. After successful resection of a parathyroid adenoma, serum intact PTH levels decline rapidly, often to undetectable concentrations, with a disappearance half-time of about 2 minutes, whereas serum calcium typically reaches a nadir between 24 and 36 hours. Serum PTH returns to the normal range within 30 hours, although measurements of the parathyroid secretory response to hypercalcemia suggest that it does not fully normalize for at least several weeks.

In the past, patients generally were maintained on a low-calcium diet until normalization of serum calcium was clearly documented, ampules of injectable calcium and other seizure precautions were maintained at the bedside, serum calcium was measured at least every 12 hours until the patient was stable, and symptomatic hypercalcemia was promptly treated with calcium, either intravenously (90-mg bolus, 50 to 100 mg/hour) or oral (1.5 to 3.0 g/day). This approach is no longer appropriate for most patients, who are discharged less than 24 hours after surgery. Instead, oral calcium supplements routinely are provided as soon as oral intake is re-established, and moderate doses of 1.25(OH)2 D3 (0.5 to 1.0 µg daily) are added in those with large adenomas and severe hyperparathyroidism or in whom alkaline phosphatase had been elevated preoperatively.That is, patients in whom an impressive calcium requirement can be anticipated, often for many weeks postoperatively, as they remineralize their skeletons. This Hungry bone syndrome is associated with hypocalcemia, hypophosphatemia, and low urinary calcium excretion.

Serum calcium should be checked at intervals of several days initially to guide adjustment of calcium and vitamin D therapy as needed to achieve a stable result. In patients in whom hypercalcemia persists for more than several days, serum PTH should be measured to exclude the possibility of postoperative hyperparathyroidism. In view of evidence that bone mineral density continues to increase for at least a year after successful parathyroidectomy, it is expected that calcium and vitamin D supplementation for at least that long.

The approach to patients with persistent or recurrent hyperparathyroidism is informed by the recognition that parathyroid hyperplasia or carcinoma, ectopic or supernumerary parathyroid tissue, and postoperative hyperparathyroidism and other complications of further surgery all are more common in this population. The first issue to address is whether surgery is indicated. When a presumed adenoma is not identified initially, the original indications for surgery generally still exist, although some patients may not be suitable candidates for more extensive surgery, such as a median sternotomy, because of concurrent

![Image](https://via.placeholder.com/150)
medical illness. Patients with parathyroid hyperplasia have experienced significant clinical improvement even after incomplete parathyroidectomy, although those with MEN 1 are likely to experience further progression of their disease. As noted previously, preoperative localization studies, although unnecessary for initial neck exploration, are justified for patients in whom a first operation failed to effect a cure. Since the disease has recurred uncommonly and with technetium Tc 99m sestamibi offers the highest sensitivity and accuracy, although other studies (ultrasound, computed tomography (CT), or magnetic resonance imaging (MRI)) may provide additional or confirmatory information. Sestamibi does localize to thyroid nodules, which may accompany parathyroid disease in 20% to 40% of patients, although it tends to wash out of thyroid tissue much more rapidly than from parathyroids. Technetium Tc 99m sestamibi can be combined with 131I scanning to improve distinction of parathyroids from thyroid nodules or with single photon emission CT (SPECT) imaging to achieve accuracy in localization not possible with planar imaging. Sestamibi scanning may fail to reveal small glands or to demonstrate multiple abnormal glands in cases of parathyroid hyperplasia, the most common cause of persistent postoperative hyperparathyroidism.

More invasive techniques have been employed as well, including angiography and selective venous sampling for measurement of PTH, although the sensitivity of these procedures for detection of residual abnormal parathyroid glands is only 50% to 65%. Angiography does offer the opportunity to attempt angioablation of any identified tissue, although this procedure is not routinely successful, and when all other parathyroid tissue has been previously removed, it precludes the opportunity to avoid hyperparathyroidism by autotransplantation of parathyroid fragments at surgery. Ultrasound- or CT-guided fine-needle aspiration of suspected parathyroid tissue may be used to obtain cytologic or immunohistochemical confirmation prior to surgery. Intraoperative ultrasonography has been useful in some cases to locate cervical or intrathyroidal glands. Success with video-assisted thoracoscopic resection of documented mediastinal lesions offers a less invasive alternative to median sternotomy for this relatively common cause of persistent hyperparathyroidism.

The need for these procedures depends on the expertise of the original surgeon and the surgeon's confidence that the neck was adequately explored initially. For example, at reoperation at one center, more than half of the missed hyperplastic parathyroid glands in those cases previously explored by a highly experienced parathyroid surgeon were found in the mediastinum or another ectopic location, whereas more than 90% of those referred by less experienced surgeons were discovered in a normal anatomic location in the neck.

After successful surgery for primary hyperparathyroidism, bone mass generally improves by as much as 5% to 10% in the first year at sites rich in trabecular bone (spine, femoral neck) but does not improve at cortical bone sites (distal radius). Bone density increases at trabecular bone sites may continue for several years, to as much as 12% to 15% after 10 years, although normal bone mineral density may not be achieved. This improvement, which is most apparent in patients with the greatest preoperative reductions in bone mass, may be related in part to rapid remodeling of the previously enlarged portion of bone undergoing remodeling but the continued improvement over years suggests a more sustained increase in bone formation and total bone volume, as well.

Familial Hypocalciuric Hypercalcemia

Familial hypocalciuric hypercalcemia (FHH), also appropriately called familial benign hypercalcemia, is, in most families, a disorder of autosomal dominant inheritance. The mutations, which cause complete or partial loss of function of the calcium-sensing receptor, lead to a shift in the parathyroid cell's set-point for calcium. As a consequence, higher than normal levels of blood calcium are needed to suppress PTH secretion. Furthermore, abnormal function of the calcium-sensing receptor in the renal thick ascending limb leads to increased, PTH-independent calcium reabsorption and consequent hypercalcemia.

The presence of a normal sensing receptor gene with the abnormal one usually leads to a very mild clinical disorder, although the receptor functions as a dimer, and certain mutations can worsen the function of the normal allele. Rare patients who inherit mutant calcium-sensing receptor genes from both parents present at birth with severe, life-threatening, primary hyperparathyroidism and almost always require immediate parathyroid surgery. In another genetic variation, a familial form of calcium-sensing receptor-dependent hypercalcemia has been described in association with other autoimmune disorders such as Hashimoto's thyroiditis and celiac sprue, in which autoantibodies directed against the sensor apparently antagonize calcium recognition by the parathyroids and renal tubules.

FH is manifested at birth by hypercalcemia. Although some controversy exists, most observers note that the condition is asymptomatic and that apparent symptoms represent ascertainment bias. Possible exceptions include the occurrence of chondrocalcinosis and perhaps pancreatitis. The blood calcium level is usually less than 12 mg/dL but can be higher. Phosphate measurements are low, as in primary hyperparathyroidism. Blood magnesium levels are high normal or slightly elevated.

When patients present as adults, the distinction from mild primary hyperparathyroidism can be difficult. The distinction between FH and primary hyperparathyroidism is a crucial one, however. Young patients with primary hyperparathyroidism are usually treated surgically and cured. In contrast, hyperparathyroidism always recurs after surgery for FHH, unless the patient is rendered hypoparathyroid by the removal of all parathyroid tissue. Therefore, surgery is contraindicated as therapy for FHH, except in the very rare patient with severe, symptomatic hypercalcemia. No blood or urine measurements are completely reliable for distinguishing between the two conditions, though the ratio of calcium clearance to creatinine clearance distinguishes most patients with FHH from those with primary hyperparathyroidism. The most helpful diagnostic information is the presence of hypercalcemia in an infant relative; such early hypercalcemia does not occur in MEN 1. Furthermore, a past history of clearly normal blood calcium, considerably lower than current measurements, makes FHH unlikely, if no other reason for a change in blood calcium exists.

Lithium Toxicity

Treatment of bipolar affective disorders with lithium commonly leads to mild, persistent increases in blood calcium occasionally out of the normal range, in affected persons. After several years of therapy, clear elevations of PTH levels and modest increases in parathyroid gland size, detected by ultrasonography, often occur. Usually, when lithium therapy is stopped, the blood calcium and PTH normalize within several months. Uncommonly, substantial hypercalcemia and clear hyperparathyroidism ensue. At surgery, parathyroid hyperplasia and, occasionally, parathyroid adenomas have been found.

The management of patients with mild, lithium-induced hypercalcemia is somewhat complicated. Like patients with mild primary hyperparathyroidism, patients taking lithium usually tolerate mild hypercalcemia without obvious symptoms. These patients can be monitored with protocols similar to those for patients with asymptomatic primary hyperparathyroidism. Close attention must be paid to urine concentrating ability in these patients, however, because the nephrogenic diabetes insipidus associated with lithium therapy can lead to dehydration and sudden worsening of hypercalcemia. Substantial hypercalcemia should lead to withdrawal of lithium therapy, if possible, with substitution of newer psychopharmacologic agents. If hypercalcemia persists after withdrawal of lithium, decisions about surgery follow the
same guidelines as those for patients with primary hyperparathyroidism.

Lithium increases the set-point for PTH secretion when it is added to isolated parathyroid cells in vitro. The set-point for PTH secretion in vivo is shifted to the right in patients who have received lithium for several years as well. A corresponding shift in the concentration of extracellular calcium needed to raise intracellular calcium levels suggests that lithium interferes with the action of the parathyroid calcium-sensing receptor, perhaps by interfering with inositol phosphate metabolism.
Parathyroid-Independent Hypercalcemia

In parathyroid-independent hypercalcemia, PTH secretion is appropriately suppressed. PTH levels, measured using two-site assays, are invariably lower than 25 pg/mL and are usually lower than normal or undetectable. Most affected patients have malignant hypercalcemia, although parathyroid-independent hypercalcemia occurs in a number of other settings as well.

Hypercalcemia of Malignancy

The diagnosis of malignant hypercalcemia is seldom a subtle one. Most malignancies produce hypercalcemia only when they are far advanced; the diagnosis becomes evident after routine studies, guided by the history and physical examination. Patients with malignant hypercalcemia usually die a month or two after hypercalcemia is discovered. Patients present with the classic signs and symptoms of hypercalcemia: confusion, polydypsia, polyuria, constipation, nausea, and vomiting. Perhaps because of the acuteness of the hypercalcemia and the elderly patient population involved, dramatic changes in mental status, culminating in coma, are relatively common. The diagnosis can be missed because the manifestations often overlap those of the underlying malignancy and because low blood albumin may lead to an apparently normal total blood calcium, despite an elevated blood ionized calcium. Even though the overall prognosis is grim, the diagnosis of malignant hypercalcemia is important to make.

Treatment is usually simple and effective in the short term; such treatment can importantly reverse the patient’s symptoms for several weeks, and even provide time for a fundamental attack on the underlying tumor, if it is treatable. Only effective treatment of the underlying neoplasm can significantly influence the long-term prognosis for patients with malignant hypercalcemia.

Although mechanisms in a given patient may be multiple, it is still useful to distinguish hypercalcemia associated with local involvement of bone from that caused by humoral mechanisms. In all cases, resorption of bone plays a pivotal role in the pathogenesis.

Local Osteolytic Hypercalcemia

Hypercalcemia resulting from tumors invading bone occurs most clearly in multiple myeloma and some patients with breast cancer. There is little evidence that the tumor cells themselves resorb bone. Instead, active osteoclasts found near the tumor cells are thought to be the proximate mediators of bone resorption. Myeloma cells and marrow cells associated with myeloma cells secrete numerous cytokines capable of stimulating bone resorption, including lymphotoxin (tumor necrosis factor-) and interleukins 1 and 6 (IL-1 and IL-6). RANKL is also found on the surface of myeloma cells and therefore may stimulate the production and activity of osteoclasts, just as RANKL on the surface of osteoblast-like cells can. In patients with myeloma, treatment with intermittent intravenous pamidronate (a bisphosphonate) inhibits this resorption and reduces the incidence of bone pain, fracture, and hypercalcemia.

The pathogenesis of hypercalcemia in breast cancer is not completely understood. Extensive metastases to bone are detected in most patients with hypercalcemia and breast cancer; this finding suggests that factors produced in bone by the metastatic tumor cells may be important. Breast cancer cells make a host of cytokines capable of resorbing bone. The role of tumor-produced PTHrP may be particularly important. A majority of breast cancer patients with hypercalcemia have elevated levels of PTHrP. Circulating PTHrP, as well as PTHrP produced in bone by metastatic tumor cells, may generate the hypercalcemia. Primary breast tumors that stain for PTHrP are more likely to result in bone metastases than are those that do not stain for PTHrP. This PTHrP may be instrumental in the establishment of lytic metastases. Animal models indicate that transforming growth factor- (TGF-), released from bone matrix by PTHrP-stimulated osteoclastic resorption, may further augment PTHrP secretion by the tumor cells. The latter may be further promoted by estrogen, which may explain the occasional occurrence of hypercalcemia following institution of estrogen or tamoxifen therapy in this disease.

Humoral Hypercalcemia of Malignancy

Albright, in 1941, was the first to propose that a PTH-like humoral factor caused the hypercalcemia in patients with malignancy but few or no bone metastases. Four decades later biochemical analysis demonstrated that such patients have high blood calcium levels, low blood phosphate levels, and high urinary cAMP levels like those found in primary hyperparathyroidism, but no elevation in PTH levels. The stimulation of cAMP production was used as an assay to eventually purify PTHrP from human tumors associated with the humoral hypercalcemia of malignancy. Animal models indicate that transforming growth factor- (TGF-), released from bone matrix by PTHrP-stimulated osteoclastic resorption, may further augment PTHrP secretion by the tumor cells. The latter may be further promoted by estrogen, which may explain the occasional occurrence of hypercalcemia following institution of estrogen or tamoxifen therapy in this disease.

The evidence that PTHrP mediates the humoral hypercalcemia of malignancy in most patients is substantial. As noted previously, PTHrP binds to the PTH/PTHrP receptor and mimics all of the actions of amino-terminal fragments of PTH. Blood levels of PTHrP are elevated in most patients with solid tumors and hypercalcemia.
Vitamin D Intoxication

Because the synthesis of 1,25(OH)₂D₃ is so tightly regulated, extremely large doses of vitamin D, on the order of 100,000 units per day, are required to cause hypercalcemia. Such doses are available in the United States only by prescription; therefore, most cases of vitamin D intoxication are iatrogenic. Occasionally, inadvertent ingestion occurs. Patients present with nausea, vomiting, weakness, and altered level of consciousness. Hypercalcemia can be severe and prolonged, because of the storage of vitamin D in fat. As expected, PTH levels are suppressed, and levels of 25(OH)D, which are poorly regulated and reflect levels of ingested vitamin D, are dramatically elevated. In contrast, the levels of 1,25(OH)₂D₃ are only modestly elevated or can be normal or even low. The modest changes in 1,25(OH)₂D₃ levels result from the down-regulation of the renal 1-hydroxylase by low levels of PTH and high levels of phosphate, calcium, and 1,25(OH)₂D₃ itself. The cause of the hypercalcemia, when it occurs in the face of normal levels of 1,25(OH)₂D₃, is uncertain but may reflect the direct action of 25(OH)D and possibly other vitamin D metabolites, which are capable of binding the 1,25(OH)₂D₃ receptor weakly. Also, the weaker vitamin D metabolites may displace 1,25(OH)₂D₃ from the circulating D-binding protein and increase the concentration of active, free 1,25(OH)₂D₃.

The hypercalcemia of vitamin D intoxication results both from increased intestinal absorption of calcium and from the direct effect of 1,25(OH)₂D₃ to increase resorption of bone. In severe cases, therefore, glucocorticoid therapy, which competes the action of 1,25(OH)₂D₃ on both bone and intestine, should be added to the therapeutic regimen of hydration and omission of dietary calcium.

Sarcoidosis and Other Granulomatous Diseases

Sarcoidosis may be associated with hypercalcemia and, even more commonly, hypercalcuria. Hypercalcemic patients have high levels of 1,25(OH)₂D₃; the high level of 1,25(OH)₂D₃ probably causes the hypercalcemia, although overproduction of bone-resorbing cytokines and PTHrP may contribute in some patients. As expected in 1,25(OH)₂D₃-dependent hypercalcemia, intestinal absorption of calcium is increased and PTH levels are suppressed. Furthermore, the hypercalcemia and high levels of 1,25(OH)₂D₃ fall upon treatment with glucocorticoids. The unregulated synthesis of 1,25(OH)₂D₃ found even in an anephric patient occurs not in the kidney but rather in the sarcoid granulomas. Removal of a large amount of granulomatous tissue can reverse hypercalcemia. Furthermore, isolated sarcoid macrophages can synthesize 1,25(OH)₂D₃ from 25(OH)D, as can normal macrophages stimulated with interferon. Such macrophages express the gene encoding the identical 25(OH)D 1-hydroxylase found in the kidney.

The unregulated synthesis of 1,25(OH)₂D₃ by activated macrophages explains many of the findings in sarcoid patients. These patients have unusual sensitivity to vitamin D and can become hypercalcemic in response to ultraviolet radiation or oral vitamin D intake. Abnormalities in calcium metabolism are usually found only in patients with active disease and large, clinically obvious total-body burdens of granulomas. Nevertheless, hypercalcemia can present in patients without obvious pulmonary disease. Furthermore, subtle abnormalities of vitamin D metabolism can be demonstrated even in patients with mildly active sarcoidosis. For example, patients without elevated levels of angiotensin-converting enzyme in their blood have normal levels of 1,25(OH)₂D₃, but these levels do not fall normally in response to an oral calcium challenge.

Hypercalcemia is also associated with other granulomatous diseases, such as tuberculosis, fungal infections, and berylliosis and has been reported in Wegener's granulomatosis, in acquired immunodeficiency syndrome (AIDS)-related Pneumocystis carinii infection, and even in association with extensive granulomatous foreign body reactions.

Hyperthyroidism

Hypercalcemia can result from thyrotoxicosis. Blood calcium levels seldom exceed 11mg/dL, but mild elevations are found in a quarter of patients. Patients have low PTH levels, low 1,25(OH)₂D₃ levels, and hypercalcuria. Hypercalcemia is caused by a direct action of thyroid hormone to stimulate bone resorption.

Beta-adrenergic blocking agents can reverse the hypercalcemia.

Vitamin A Intoxication

Excess ingestion of vitamin A (retinol) results in a syndrome of dry skin, pruritus, headache from pseudotumor cerebri, bone pain, and, occasionally, hypercalcemia. Hypercalcemia occurs only with the ingestion of 10 times the RDA (5000 IU/day). The identical syndrome can result from ingestion of the vitamin A derivatives all-trans-retinoic acid [Accutane] and tretinoin (all-trans-retinoic acid [Retin-A]), used to treat acne. Bones can show characteristic periosteal calcification on radiographs. The hypercalcemia is probably caused by the action of retinoids to directly stimulate bone resorption. The diagnosis is made by the association of a history of excess ingestion of retinoids with the characteristic syndrome and abnormal results of liver function tests; elevated vitamin A levels confirm the diagnosis. Treatment involves hydration and, if necessary, glucocorticoids.

Adrenal Insufficiency

Hypercalcemia occurs in the setting of adrenal insufficiency. Blood calcium is elevated partly as a result of hemoconcentration and increased albumin levels, and even in association with extensive granulomatous foreign body reactions.

Thiazide Diuretics

Thiazide diuretics do not cause hypercalcemia by themselves, but can exacerbate the hypercalcemia of primary hyperparathyroidism. The mechanism of the hypercalcemia may involve the action of thiazide diuretics to increase distal tubular calcium reabsorption. Thiazides block sodium chloride cotransport into these cells (see Figs. 28-26-7). The fall in intracellular chloride hyperpolarizes the cell, thereby increasing calcium influx through voltage-sensitive channels. Decreased renal clearance of calcium alone would be expected to raise blood calcium in the normal human only transiently because the transient hypercalcemia would be expected to suppress PTH secretion and lead to return of the blood calcium to normal.

As predicted by this model, thiazide administration leads to chronic hypercalcemia only in patients with abnormal parathyroid physiology. In primary hyperparathyroidism, thiazide administration exacerbates the hypercalcemia, and in hypoparathyroidism, thiazide administration facilitates the maintenance of normocalcemia when given in conjunction with 1,25(OH)₂D₃ and calcium.

Milk-Alkali Syndrome

The triad of hypercalcemia, metabolic alkalosis, and renal failure can be the consequence of massive ingestion of calcium and absorbable alkali. This syndrome was first described when milk and sodium bicarbonate were used in large amounts to treat peptic ulcer disease. With the change in ulcer treatment to nonabsorbable antacids and suppression of acid secretion, milk-alkali syndrome became rare. In the last several years, however, the increased use of calcium carbonate to treat dyspepsia and osteoporosis has led to the reappearance of milk-alkali syndrome. In most cases, a history of ingestion of several grams per day of calcium in the form of calcium carbonate can be elicited. The pathogenesis of the syndrome is not understood in detail but may well involve a vicious circle in which alkalosis decreases renal calcium clearance and hypercalcemia helps maintain alkalosis. Nephrocalcinosis, nephrogenic diabetes insipidus, decrease in GFR associated with hypercalcemia, and hypovolemia from vomiting all lead to renal failure, which can be severe. PTH levels, measured with currently available two-site assays, are invariably low in hypercalcemic patients as are levels of 1,25(OH)₂D₃. After clearance of the calcium by hydration or dialysis, if necessary, renal function generally returns to normal, unless the disorder has been severe and long-standing.
Immobilization

Immobilization can lead to bone resorption sufficient to cause hypercalcemia. The immobilization is usually caused by spinal cord injury or extensive casting after fractures. Hypercalcemia of immobilization occurs predominantly in the young or in patients with other reasons for a high rate of bone turnover, such as Paget’s disease or extensive fractures. Hypercalcuria and substantial bone loss are more common than hypercalcemia is. After spinal cord injury, the hypercalcuria is maximal at 4 months and can persist for more than a year. T T PTH and 1,25(OH)₂D₃ levels are suppressed T T T T ; bone biopsies show increased resorption and decreased formation of bone. T T T The combination of calcitonin and bisphosphonates T T T T or simply bisphosphonates alone T T T T have been used to reverse the hypercalcemia and hypercalciuria of spinal cord injury.

Renal Failure

Following rhabdomyolysis, during the oliguric phase of acute renal failure, severe hypocalcemia can result from acute hyperphosphatemia and calcium deposition in muscle. T T PTH levels are high. In the diuretic phase that follows, hypercalcemia can occur. The hypercalcemia results from the high 1,25(OH)₂D₃ levels observed in some patients and from mobilization of the calcium deposits. T T

In chronic renal failure, hypercalcemia can result from tertiary hyperparathyroidism or may appear during therapy of aplastic bone disease associated with low PTH levels and sometimes with aluminum toxicity. T T

Williams’ Syndrome

Williams’ syndrome is a developmental disorder in which supravalvular aortic stenosis is associated with elfin facies and mental retardation. T T T T Hypercalcemia can occur transiently in the first four years of life. Affected hypercalcemic infants have been found to have increased intestinal absorption of calcium and associated elevations of 1,25(OH)₂D₃ that fall to normal as the blood calcium normalizes. T T Levels of 25(OH)D are normal. The hypercalcemia can generally be controlled by dietary manipulation.

Molecular analysis has clarified the origin of the connective tissue component of Williams’ syndrome. T T T T Isolated supravalvular aortic stenosis is associated with deletion or translocation of the distal portion of the elastin gene. Williams’ syndrome, with more protean connective tissue abnormalities and mental retardation, is associated with large deletions that include the elastin gene and a gene encoding the protein kinase LIM-kinase 1. T T It is possible that the subgroup of patients with infantile hypercalcemia have deletion of another gene near the elastin locus. This possible genetic heterogeneity may explain the conflicting literature, which has found no consistent abnormality of calcium metabolism in normocalcemic patients with Williams’ syndrome. T T

Jansen’s Metaphyseal Chondrodysplasia

Jansen’s metaphyseal chondrodysplasia is a rare disease in which affected persons present in childhood with short stature and hypercalcemia. T T (Fig. 26-29) Blood chemistry studies suggest hyperparathyroidism, with high calcium, low normal phosphate, high 1,25(OH)₂D₃, high alkaline phosphatase, and high urinary hydroxyproline, but PTH levels are suppressed. T T A generalized defect in endochondral bone formation results from abnormally organized chondrocytes in growth plates. Metaphyses appear disordered and rachitic on radiographs. The bones may show signs of osteitis fibrosa cystica. Constitutive activation of the PTH/PTHrP receptor, caused by point mutations in the transmembrane domains of the receptor, explains the findings in this disorder. T T T T T T The abnormalities on serum chemistry studies result from PTH-like actions of the receptor in bone and kidney. T T T T T T The growth plate disorder results from PTHrP-like actions of the receptor on the growth plate.
Approach to the Hypercalcemic Patient

The diagnostic approach to the hypercalcemic patient is strongly influenced by the clinical setting and the knowledge that primary hyperparathyroidism is at least twice as common as all other causes combined (Table 26-2). These considerations are particularly significant in the patient who seems otherwise well and in whom the hypercalcemia is detected incidentally or is mild, stable, or known to be of long duration (i.e., years). Among outpatients referred to endocrinologists for evaluation of hypercalcemia, for example, more than 90% are found to have primary hyperparathyroidism. In ill or hospitalized patients, malignant disease is the cause in more than 50% of cases. The differential diagnosis is seldom complicated, however, because malignant hypercalcemia usually presents in the context of advanced, clinically obvious disease.

Because hypercalcemia usually is first detected as an elevation of total serum calcium, it is important to distinguish hemoconcentration or rare instances of calcium-binding paraproteinemia from a true increase in serum ionized calcium (Fig. 26-30). The presence of hypercalcemia should be confirmed by direct measurement of ionized calcium, and total calcium should be repeated, together with albumin, globulin, electrolytes, blood urea nitrogen (BUN), creatinine, and phosphate. Especially when hypercalcemia is mild, it is prudent to repeat the serum total or ionized calcium measurement at least twice, preferably with the patient fasting and without venous occlusion, before proceeding with more costly studies directed at discovering its etiology.

A careful history and physical examination, combined with efforts to assess chronicity by seeking previous results of routine multichannel serum chemistry determinations, most often points to the likely diagnosis. Serum phosphate often is low in hyperparathyroidism; however, because phosphate is often low in PTHrP-secreting malignancies as well, the presence of hypophosphatemia is not helpful in distinguishing these possibilities. When serum phosphate levels are normal or high despite correction of dehydration, the possibility of PTH- or PTHrP-independent hypercalcemia should be considered more strongly, however. Elevations in serum chloride and alkaline phosphatase, often observed in primary hyperparathyroidism, cannot be reliably employed in the differential diagnosis of hypercalcemia.

Important elements of the medical history of hypercalcemic patients include inquiries about kidney stones or fractures, weight loss, back or bone pain, fatigue or weakness, cough or dyspnea, ulcer disease, or pancreatitis; ingestion of vitamins, calcium preparations, lithium, or thiazides; dates of most recent mammograms and chest radiographs; and a family history of hypercalcemia, kidney stones, ulcer disease, endocrinopathy, or tumors of the head or neck. Because malignancy is a common cause of hypercalcemia and may occur concomitantly with primary hyperparathyroidism, clinical findings strongly suggestive of malignancy should be acted on by proceeding directly to a search for an underlying tumor, regardless of serum PTH levels.

The single most important test in the differential diagnosis of hypercalcemia is the measurement of serum PTH, preferably in a two-site assay specific for the intact, biologically active molecule (see Fig. 26-21). A consistently elevated serum PTH

<table>
<thead>
<tr>
<th>TABLE 26-2 – Causes of Hypercalcemia</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Parathyroid-Dependent Hypercalcemia</strong></td>
</tr>
<tr>
<td>Primary hyperparathyroidism</td>
</tr>
<tr>
<td>Tertiary hyperparathyroidism</td>
</tr>
<tr>
<td>Familial hypocalciuric hypercalcemia</td>
</tr>
<tr>
<td>Lithium-associated hypercalcemia</td>
</tr>
<tr>
<td><strong>Parathyroid-Independent Hypercalcemia</strong></td>
</tr>
<tr>
<td>Neoplasms</td>
</tr>
<tr>
<td>Parathyroid hormone-related protein-dependent</td>
</tr>
<tr>
<td>Other humoral syndromes</td>
</tr>
<tr>
<td>Osteolytic metastases and multiple myeloma</td>
</tr>
<tr>
<td>Excess vitamin D/1,25(OH)D</td>
</tr>
<tr>
<td>Vitamin D ingestion</td>
</tr>
<tr>
<td>1,25-Dihydroxyvitamin D intoxication</td>
</tr>
<tr>
<td>Topical vitamin D analogues</td>
</tr>
<tr>
<td>Granulomatous disease</td>
</tr>
<tr>
<td>Williams’ syndrome</td>
</tr>
<tr>
<td>Thyrotoxicosis</td>
</tr>
<tr>
<td>Adrenal insufficiency</td>
</tr>
<tr>
<td>Renal failure</td>
</tr>
<tr>
<td>Acute renal failure</td>
</tr>
<tr>
<td>Chronic renal failure with aplastic bone disease</td>
</tr>
<tr>
<td>Immobilization</td>
</tr>
<tr>
<td>Jansen’s disease</td>
</tr>
<tr>
<td>Drugs</td>
</tr>
<tr>
<td>Vitamin A intoxication</td>
</tr>
<tr>
<td>Milk-alkali syndrome</td>
</tr>
<tr>
<td>Thiazide diuretics</td>
</tr>
<tr>
<td>Theophylline</td>
</tr>
</tbody>
</table>
level in the presence of true hypercalcemia is always abnormal and almost always indicates the presence of primary hyperparathyroidism. The exceptions are FHH, autonomous parathyroid secretion complicating secondary hyperparathyroidism (*tertiary* hyperparathyroidism), ongoing lithium therapy in some patients, and, very rarely, ectopic PTH secretion by a malignant neoplasm.

Diagnosis of primary hyperparathyroidism is complicated, however, by the fact that some patients fail to exhibit both hypercalcemia and elevated iPTH. In up to 10% of patients with hypercalcemia and primary hyperparathyroidism, PTH levels may fall within the (high) normal range with current PTH assays. Such PTH levels are inappropriate in the face of hypercalcemia, however, and support the diagnosis of PTH-dependent hypercalcemia. In fact, many such patients will demonstrate frankly elevated serum PTH if retested, especially if dietary calcium is restricted beforehand. As noted previously, some may present with serum calcium in the high-normal range (>10.0 mg/dL), together with an elevated or high-normal PTH. This abnormality may be discovered incidentally in an otherwise asymptomatic person or in the course of evaluating recurrent urolithiasis or osteopenia. Patients with persistently high-normal serum calcium and high-normal iPTH should be retested at intervals and, meanwhile, given a provisional diagnosis of hyperparathyroidism and evaluated accordingly.

In patients with PTH-dependent hypercalcemia (Fig. 26-31), calcium and creatinine should be measured in a 24-hour urine collection and a simultaneous serum sample to measure total urinary calcium output (mg/day) and the clearance ratio of urinary and serum calcium and creatinine (i.e., UCa/Cr x SCr/Cr). A daily calcium excretion of less than 100 mg per day, or a clearance ratio less than 0.01, should prompt consideration of familial hypocalciuric hypercalcemia, especially in patients younger than 40 years of age, those with high-normal PTH levels, and those with a family history of hypercalcemia. A urinary calcium excretion of greater than 4 mg/kg per day or a clearance ratio greater than 0.02 effectively excludes FHH. If FHH, serum phosphate is normal or slightly low, serum magnesium may be slightly high, and serum

1,25(OH)2D3 is normal or low (unlike in primary hyperparathyroidism).

A definite diagnosis of FHH, as in the MEN syndromes, may be provided by confirming the presence of mutations in the relevant genes, although such studies are not invariably informative (presumably because of mutations in introns and other unchecked regions) and usually are unnecessary. The identification of RET gene mutations is now an essential part of the management of families with MEN 2, because this information most effectively guides the decision for preventive thyroidectomy to prevent medullary cancer of the thyroid. In contrast, the identification of MENIN gene mutations has not yet led to any effective preventive strategies; thus, genetic analysis may be useful only for genetic counseling in families with MEN 1. Even for this purpose, the incomplete ascertainment of mutations limits the effectiveness of such analysis.

In patients with suspected lithium-induced hyperparathyroidism, a trial off lithium, if feasible clinically, may confirm the diagnosis or indicate the presence of persistent primary hyperparathyroidism. Patients with primary hyperparathyroidism should undergo bone densitometry, preferably at both cortical bone and trabecular bonerich sites (i.e., forearm and hip or lumbar spine, respectively) to assist in the decision about surgery. Those younger than 40 years of age or who have a family history of hypercalcemia (or other MEN manifestations) should be evaluated for these syndromes as well. Patients who do not meet criteria for parathyroidectomy should be followed medically, as should those with FHH.

A low or undetectable serum PTH level signifies the presence of nonparathyroid hypercalcemia and should prompt a detailed evaluation for malignancy or other causes of PTH-independent hypercalcemia (see Table 26-2). Breast and lung cancers alone account for over 50% of all malignancy-associated hypercalcemias. Mammography, chest radiography with or without CT, abdominal CT, and serum and urinary immuno-electrophoresis are among the more useful tests for detecting the cause of nonparathyroid hypercalcemia. Although humoral mechanisms, especially secretion of PTHrP, are implicated in the pathogenesis of most cancer-associated hypercalcemias, bone metastases are common, particularly in breast cancer. Technetium

99m bone scanning, therefore, generally is useful for detecting this syndrome and identifying bones vulnerable to fracture. The utility of serum PTHrP measurements probably is limited to the unusual situation in which serum PTH is suppressed but an underlying malignancy cannot readily be demonstrated.

In the absence of evident malignancy, unusual causes of hypercalcemia should be sought. Vitamin D and vitamin A intoxication can be excluded by measurement of serum 25(OH)D and retinoids, respectively. Elevated 1,25(OH)2D3 and hypercalcemia may occur in several settings, including sarcoidosis and other granulomatous diseases, B-cell and T-cell lymphomas (including AIDS-associated lymphomas) and, uncommonly, in epithelial neoplasms such as lung cancer. Very rarely, patients with severe idiopathic hypercalcemia and excessive absorption of dietary calcium may manifest mild, dietary-dependent hypercalcemia. Overtreatment of hyperparathyroidism or other conditions with oral 1,25(OH)2D3 or topical use of analogues of the active metabolite in psoriasis should be obvious from the history. Because hypercalcemia and hypercalciuria are observed in up to 10% and 30%, respectively, of patients with thyrotoxicosis, measurement of serum thyroid-stimulating hormone (TSH) may be helpful, especially in older patients who may be less overtly symptomatic. Adrenal insufficiency and pheochromocytoma usually are accompanied by characteristic clinical features, but a definite diagnosis may be sought with appropriate studies. Granulomatous diseases are among the more common disorders that underlie initially unexplained hypercalcemia.
Therapy of Severe Hypercalcemia

Causes of Severe Hypercalcemia

The need for urgent therapy of acute, severe hypercalcemia, usually defined as a serum calcium concentration greater than 14 mg/dL (3.5 mM), is unusual. This is because most patients with hypercalcemia have primary hyperparathyroidism, in which hypercalcemia is typically chronic and mild. Episodes of acute, severe hypercalcemia may occur occasionally in primary hyperparathyroidism (“parathyroid crisis”), usually in patients with large parathyroid adenomas and very high PTH levels. The severe hypercalcemia appears in the setting of dehydration due to diarrheal illness, protracted vomiting or diuretic therapy, recovery from major surgery, immobilization, ingestion of large amounts of oral calcium salts, or parathyroid carcinoma.

Most often, acute, severe hypercalcemia is encountered in patients with underlying malignancy, in whom accelerated bone resorption dramatically increases the filtered load of calcium. The ensuing profound hypercalcemia impairs renal tubular sodium reabsorption, which induces progressive extracellular volume depletion, reduces GFR, impairs renal calcium clearance, and further aggravates the hypercalcemia. In many such patients, elevated circulating levels of PTH and PTHrP compound the problem by mimicking the action of PTH to enhance distal tubular calcium reabsorption.

Clinical Features of Severe Hypercalcemia

The indications for urgent therapy of hypercalcemia usually relate more to the presence of clinical symptoms of hypercalcemia than to the absolute level of serum calcium, although few clinicians would hesitate to treat patients in whom total serum calcium exceeded 14 mg/dL (3.5 mM). Many patients with previously mild hypercalcemia become symptomatic when serum calcium concentrations exceed 12 mg/dL (3.0 mM). It is important to remember that hypoalbuminemia may mask significant elevations of ionized calcium. The most common symptoms of severe hypercalcemia are referable to disturbances of nervous system and gastrointestinal function—fatigue, weakness, lethargy, confusion, coma (rarely), anorexia, nausea, abdominal pain (rarely due to pancreatitis), and constipation. Polyuria, nocturia, and polydypsia are present also.

Bone pain is often present but is usually due to underlying metastatic disease. Cardiac arrhythmias may occur, particularly bradyarrhythmias or heart block, and digitalis toxicity may be potentiated. Patients who suffer a fatal outcome from acute severe hypercalcemia may manifest coma, hypotension, acute pancreatitis, acute renal failure, widespread soft tissue calcification, heart failure, or venous thrombosis, particularly of the renal veins.

Management of Severe Hypercalcemia

The first decision to be made in the management of acute, severe hypercalcemia is whether or not to treat the problem at all. The question of definitive treatment may become an issue for the patient with an incurable, widely disseminated malignancy, when all other approaches to controlling the neoplasm have been exhausted, and the patient has chosen not to have complications treated. Otherwise, as noted previously, patients with serum calcium levels above 14 mg/dL ordinarily should receive aggressive treatment. Treatment most often entails rehydration and administration of a bisphosphonate intravenously. Calcitonin can be useful as a temporary measure early in therapy. Gallium nitrate, plicamycin, and intravenous phosphate are infrequently used, because of toxicity.

Volume Repletion

When treatment is indicated, the first priority is to correct the extracellular volume depletion that is invariably present, usually by infusing isotonic saline at a rate of 2 to 4 liters/day. The aggressiveness with which the individual patient is rehydrated must be considered in relation to both the patient's volume status and the risk of precipitating or aggravating congestive heart failure or asosites. Diuretics, particularly thiazides, should be discontinued. The conventional use of furosemide or other potent "loop" diuretics to promote calciuresis may exacerbate extracellular volume depletion if used early in the course of treatment. In light of the current availability of highly effective alternatives for the therapy of hypercalcemia, such drugs probably are best avoided, except when vigorous rehydration fails to improve severe hypercalcemia or may precipitate congestive heart failure. In any case, prolonged use of saline-induced calciuresis without the early introduction of an effective anti-resorptive agent is ill advised and ultimately futile.

Bisphosphonates

Several drugs are available for specific therapy of acute hypercalcemia (Table 26-3). Each of these agents inhibits osteoclastic bone resorption, although their relative efficacy and toxicity vary substantially. The antiresorptive drug of first choice in most situations is an intravenous bisphosphonate. Of the two currently available in the United States, pamidronate more reliably normalizes serum calcium and frequently does so following even a single intravenous infusion, although the dose can be repeated in refractory cases. The drug is generally well tolerated, although local pain or swelling at the infusion site, low-grade fever 1 or 2 days after the infusion, transient lymphopenia, and mild hypophosphatemia or hypomagnesemia may occur in a significant fraction of patients. Serum calcium concentration usually declines rapidly following administration of 60 to 90 mg, reaching the normal range within 2 to 3 days in over 80% of cases and occasionally falling below normal at its nadir. The duration of the response to pamidronate varies dramatically ranging from a week or so to several months and cannot be predicted in an individual patient.

Calcitonin

Calcitonin, which directly inhibits osteoclast function, is frequently used with other antiresorptive agents to achieve more rapid control of severe hypercalcemia. Calcitonin rarely produces a decline in serum calcium of more than 2 mg/dL, and its efficacy typically is limited to a few days at most, possibly because of receptor

### TABLE 26-3 – Therapy for Severe Hypercalcemia

<table>
<thead>
<tr>
<th>Therapy</th>
<th>Usual Dose</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rehydration</td>
<td>24 L/day of 0.9% sodium chloride IV</td>
<td>qd x 15 days</td>
</tr>
<tr>
<td>Pamidronate</td>
<td>6090 mg IV over 424 hr</td>
<td>qd x 1 day</td>
</tr>
<tr>
<td>Etidronate</td>
<td>7.5 mg/kg IV over 4 hr</td>
<td>qd x 37 days</td>
</tr>
<tr>
<td>Calcitonin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salmon</td>
<td>4 IU/kg SC</td>
<td>q612 hr</td>
</tr>
<tr>
<td>Human</td>
<td>0.5 mg SC</td>
<td>q1224 hr</td>
</tr>
<tr>
<td>Gallium nitrate</td>
<td>200 mg/m² of body surface area IV</td>
<td>qd x 5 days</td>
</tr>
<tr>
<td>Plicamycin</td>
<td>1525 µg/kg IV over 46 hr</td>
<td>qd or qod x 15 days</td>
</tr>
<tr>
<td>Glucocorticoids</td>
<td>2000300 mg hydrocortisone IV</td>
<td>qd x 35 days</td>
</tr>
<tr>
<td>Dialysis</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
down-regulation in target cells of bone and kidney. Its major advantages are a more rapid onset of action (by several hours) and its potential to augment renal calcium excretion directly. Calcitonin is also unique in that it may produce a significant analgesic effect in many patients with fracture or metastatic bone disease. Calcitonin is generally well tolerated, although transient nausea, vomiting, abdominal cramps, flushing, and local skin reactions may occur.

**Gallium Nitrate**

Gallium nitrate can normalize serum calcium within 5 to 7 days in most patients given the recommended 5-day infusion of 200 mg/m²/day, but it is rarely used because of significant nephrotoxicity and other adverse effects such as hypotension, nausea, vomiting, hypophosphatemia, and anemia and is not currently available in the United States.

**Plicamycin**

Plicamycin (formerly mithramycin), an antiresorptive drug that for many years was considered the first-line agent for treatment of severe hypercalcemia, can cause a broad range of significant hepatic, renal, and hematologic toxic effects and offers no practical or theoretical advantage over bisphosphonates. The drug is no longer used for this purpose.

**Phosphate**

Intravenous phosphate infusion was employed in the past to control hypercalcemia in patients with concomitant hypophosphatemia, but the mechanism of action, which may partly depend on direct inhibition of osteoclasts, almost certainly involves also generalized precipitation of calcium phosphate salts in soft tissues, notably kidneys, pancreas, heart, lungs, stomach, and vessels. For this reason, and also because intravenous phosphate administration may produce severe, life-threatening hypocalcemia or hypotension, this therapy can no longer be considered standard for severe hypercalcemia. Oral phosphate, in doses up to 2000 mg daily, has been advocated for therapy of hypercalcemia in hypophosphatemic patients with normal renal function, but its efficacy is limited, it is not appropriate for urgent therapy of severe hypercalcemia, and its use is accompanied by frequent nausea, abdominal cramps, and diarrhea.

**Other Approaches**

Other strategies for control of severe hypercalcemia may be applicable in specific circumstances. High-dose glucocorticoids, for example, may be very helpful in vitamin D intoxication, in granulomatous diseases such as sarcoidosis, and with hematologic malignancies known or likely to be glucocorticoid-responsive. Because the effects of antiresorptive agents are somewhat delayed (i.e., days) and acute reduction of serum calcium during the first 12 to 24 hours depends mainly on inducing calciuresis through adequate rehydration, peritoneal or, better, hemodialysis may be required early in the management of life-threatening hypercalcemia. In this case, it is desirable to employ low- or zero-calcium dialysate.
HYPOCALCEMIC DISORDERS

Clinical Presentation

The predominant clinical symptoms and signs of hypocalcemia are those of neuromuscular irritability, including perioral paresthesias, tingling of the fingers and toes, and spontaneous or latent tetany. Tetany can be elicited by percussion of the facial nerve below the zygoma, resulting in ipsilateral contractions of the facial muscle (Chvostek’s sign) or by 3 minutes of occlusive pressure with a blood pressure cuff resulting in carpal spasm, which, on occasion, can be very painful (Trousseau’s sign) (Fig. 26-32). The usefulness of these signs in diagnosing hypocalcemia and in following therapeutic responses cannot be overemphasized.

Electrocardiographic abnormalities also result from hypocalcemia, including prolonged QT intervals and marked QRS and ST segment changes that may mimic acute myocardial infarction or conduction abnormalities. Ventricular arrhythmias are a rare complication of hypocalcemia, although congestive heart failure, corrected by normalization of serum calcium, has been reported.

In profound hypocalcemia or during acute falls in serum calcium, grand mal seizures or laryngospasm also may be observed. Chronic hypocalcemia is associated with milder symptoms and signs of neuromuscular irritability and may even be asymptomatic. Long-standing hypocalcemia associated with hyperphosphatemia (observed with PTH deficiency or resistance) may lead to calcification of the basal ganglia and occasional extrapyramidal disorders. In addition, mineral ion deposits in the lens may lead to cataract formation.

Chronic hypocalcemia, particularly when associated with hypophosphatemia, as in vitamin D deficiency, is associated with growth plate abnormalities in children (rickets) and defects in the mineralization of new bone (osteomalacia) (see Chapter 27). Severe symptomatic hypocalcemia constitutes an emergency that requires immediate attention to prevent seizures and death from laryngospasm or cardiac causes.

Total calcium in serum includes both the free (biologically active) and protein-bound components; the major binding protein is albumin (see "Roles of the Mineral Ions"). Therefore, measurements of total calcium cannot be interpreted without concurrent measurement of albumin. Studies of hypoalbuminemic patients with cirrhosis have led to a formula for correction of total calcium based on concurrent albumin levels (a decrease in calcium of 0.8 mg/dL for every 1 g/dL decrease in albumin). No formula has proved accurate, however, for assessment of calcium concentration in acutely ill patients. This difficulty probably relates to the variety of factors that may increase protein binding and decrease the fraction of total calcium present as the free ion, including alkalosis, elevated circulating free fatty acids, and lipid infusions. Consequently, ionized calcium should be measured when the diagnosis of hypocalcemia is considered in the setting of acute illness or severe hypoalbuminemia.

Chronic hypocalcemia is most often due to deficiency of PTH or 1,25(OH)₂D₃ or to resistance to the biologic effects of these calcium-regulating hormones (Table 26-4).
Parathyroid-Related Disorders

Hypocalcemia associated with parathyroid dysfunction can be differentiated from other causes of hypocalcemia by routine laboratory tests. Serum calcium is low owing to lack of PTH-mediated bone resorption and urinary calcium reabsorption. Serum phosphate is increased owing to impaired renal clearance. Serum 1,25(OH)₂D₃ is low because PTH and hypophosphatemia stimulate the renal 25(OH)D₁-hydroxylase. Consequently, 1,25(OH)₂D₃-mediated intestinal calcium absorption is markedly decreased, further exacerbating the hypocalcemia. PTH levels measured using sensitive two-site PTH assays (see Fig. 26-21) are usually low or undetectable but may be inappropriately normal if some degree of PTH production is preserved. Elevated levels of PTH are found in syndromes associated with resistance to the biologic effects of PTH.

Congenital or Inherited Parathyroid Disorders

Several rare syndromes associated with congenital or inherited hypoparathyroidism appear sporadically, or in a variety of inheritance patterns suggesting multiple causes. Mutation of a transcription factor called glial cell missing (GCM)b (6p23), which is expressed in the PTH-secreting cells of the developing parathyroids, is a cause of familial hypoparathyroidism in humans and mice. Though as yet uncloned, the gene responsible

<table>
<thead>
<tr>
<th>TABLE 26-4 -- Causes of Hypocalcemia</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Parathyroid-Related Disorders</strong></td>
</tr>
<tr>
<td>Absence of the parathyroid glands or of PTH</td>
</tr>
<tr>
<td>Congenital</td>
</tr>
<tr>
<td>DiGeorge's syndrome</td>
</tr>
<tr>
<td>X-linked or autosomally inherited hypoparathyroidism</td>
</tr>
<tr>
<td>Autoimmune polyglandular syndrome type I</td>
</tr>
<tr>
<td>PTH gene mutations</td>
</tr>
<tr>
<td>Postsurgical hypoparathyroidism</td>
</tr>
<tr>
<td>Infiltrative disorders</td>
</tr>
<tr>
<td>Hemochromatosis</td>
</tr>
<tr>
<td>Wilson’s disease</td>
</tr>
<tr>
<td>Metastases</td>
</tr>
<tr>
<td>Hypoparathyroidism following radioactive iodine thyroid ablation</td>
</tr>
<tr>
<td>Impaired secretion of PTH</td>
</tr>
<tr>
<td>Hypomagnesemia</td>
</tr>
<tr>
<td>Respiratory alkalosis</td>
</tr>
<tr>
<td>Activating mutations of the calcium sensor</td>
</tr>
<tr>
<td>Target organ resistance</td>
</tr>
<tr>
<td>Hypomagnesemia</td>
</tr>
<tr>
<td>Pseudohypoparathyroid</td>
</tr>
<tr>
<td>Type I</td>
</tr>
<tr>
<td>Type II</td>
</tr>
<tr>
<td><strong>Vitamin D Related Disorders</strong></td>
</tr>
<tr>
<td>Vitamin D deficiency</td>
</tr>
<tr>
<td>Dietary absence</td>
</tr>
<tr>
<td>Malabsorption</td>
</tr>
<tr>
<td>Accelerated loss</td>
</tr>
<tr>
<td>Impaired enterohepatic recirculation</td>
</tr>
<tr>
<td>Anticonvulsant medications</td>
</tr>
<tr>
<td>Impaired 25-hydroxylation</td>
</tr>
<tr>
<td>Liver disease</td>
</tr>
<tr>
<td>Isoniazid</td>
</tr>
<tr>
<td>Impaired 1-hydroxylation</td>
</tr>
<tr>
<td>Renal failure</td>
</tr>
<tr>
<td>Vitamin D dependent rickets type I</td>
</tr>
<tr>
<td>Oncogenic osteomalacia</td>
</tr>
<tr>
<td>Target organ resistance</td>
</tr>
<tr>
<td>Vitamin D dependent rickets type II</td>
</tr>
<tr>
<td>Phenytoin</td>
</tr>
<tr>
<td><strong>Other Causes</strong></td>
</tr>
<tr>
<td>Excessive deposition into the skeleton</td>
</tr>
<tr>
<td>Osteoblastic malignancies</td>
</tr>
<tr>
<td>Hungry bone syndrome</td>
</tr>
<tr>
<td>Chelation</td>
</tr>
<tr>
<td>Foscarnet</td>
</tr>
<tr>
<td>Phosphate infusion</td>
</tr>
<tr>
<td>Infusion of citrated blood products</td>
</tr>
</tbody>
</table>
Neonatal hypocalcemia
Prematurity
Asphyxia
Diabetic mother
Hyperparathyroid mother
HIV infection
Drug therapy
Vitamin D deficiency
Impaired PTH responsiveness
Critical illness
Pancreatitis
Toxic shock syndrome
Intensive care unit patients

EDTA, ethylenediaminetetra-acetic acid; HIV, human immunodeficiency virus; PTH, parathyroid hormone.

for X-linked recessive hypoparathyroidism has been linked to Xq26.7-q27. DiGeorge's syndrome occurs sporadically and is associated with an embryologic defect in the formation of the third, fourth, and fifth branchial pouches, resulting in the absence of parathyroid glands. DiGeorge's syndrome is often associated with other congenital abnormalities in a syndrome referred to by the acronym CATCH 22 (cardiac defect, abnormal facies, thymic hypoplasia, cleft palate, hypocalcemia, and 22q11 deletions). Microdeletion of 22q11.21q11.23 and a t(2;22)(q14;q11) balanced translocation suggest that a gene at chromosome 22q11 may be pathogenic in this syndrome. Hypoparathyroidism also has been reported in two patients with a 22q11 deletion. In a number of cases of DiGeorge's and velocardiofacial syndromes, there is a recognizable abnormality at 22q11; rather, terminal 10p deletions or interstitial 10p13/10p14 deletions have been identified, suggesting that two loci may be critical for development of branchial pouch structures. Terminal deletions of 10p accompanied by hypoparathyroidism can be further subdivided into DiGeorge critical region II (10p13.14) and a more telomeric region (10p14.10pter) wherein mutation of the transcription factor GATA3 causes the syndrome of hypoparathyroidism, sensorineural deafness, and renal anomaly (HDR).

DiGeorge's syndrome may in fact be a neurocraniopathy, because ablation of the premigratory cephalic neural crest in chick embryos produces the same phenotype. The contribution of homeobox genes to parathyroid development and their potential relationship to DiGeorge's syndrome also have been demonstrated by the absence of thymic and parathyroid tissue, accompanied by cardiac and craniofacial abnormalities, in mice lacking the homeobox gene hoxa11. Familial hypoparathyroidism is seen in conjunction with mucocutaneous candidiasis, Addison's disease, and other immune disorders in autosomal recessive autoimmune polyglandular syndrome type I (see Chapter 37). Hypoparathyroidism may also be observed in association with mitochondrial myopathies such as mitochondrial trifunctional protein deficiency and the Kearns-Sayre syndrome. With other inherited forms of hypoparathyroidism, the disorder may be observed as an isolated defect, or affected patients may present with other features such as lymphedema, dysmorphism, hearing impairment, and renal and cardiac abnormalities.

Abnormalities in the Parathyroid Hormone Gene

Specific defects have been found in the PTH gene in a small number of kindreds affected by congenital hypoparathyroidism. These defects include point mutations in the signal peptide and in an intron border, leading to aberrant splicing. No abnormalities in the sequences encoding PTH(1-84) have been discovered in familial hypoparathyroidism.

Destruction of the Parathyroid Glands

The most common cause of chronic hypocalcemia is post-surgical hypoparathyroidism. This problem may occur after removal of all parathyroid tissue during thyroidec- tomy and radical neck dissection for malignancies, or after inadvertent interruption of the blood supply to the parathyroid glands during head and neck surgery. Transient hypoparathyroidism, attributed to chronic suppression of the remaining normal glands, is common after parathyre idec-tomy; permanent hypoparathyroidism may occur after vascular or surgical injury or inadvertent removal of all parathyroid tissue. Rarely, transient hypoparathyroidism may follow spontaneous infarction of autonomou s tissue in primary hyperparathyroidism. Hypoparathyroidism is a rare complication of radioactive iodine ablation of the thyroid gland for Graves' disease. Hypoparathyroidism also can occur as a result of infiltrative diseases of the parathyroids. This association is seen in diseases of iron overload such as hemochromatosis and in patients with thalassemia major who have received numerous transfusions. Copper deposition in Wilson's disease may also cause parathyroid dysfunction. Metastatic disease to the parathyroid glands is a rare cause of hypoparathyroidism, presumably because of the need for four-gland involvement before significant hypoparathyroidism is observed.

Impaired PTH Secretion

Impaired secretion of PTH from the parathyroid glands can lead to functional hypoparathyroidism. This problem is commonly seen in profound hypomagnesemia, in which target organ resistance to PTH can also occur. Both of these abnormalities are reversible on magnesium repletion (see "Disorders of Magnesium Metabolism"). Copper deposition in Wilson's disease may also cause hypoparathyroidism. Chronic respiratory alkalosis leads to hyperphosphatemia and decreased ionized calcium levels accompanied by impaired renal calcium resorption and inappropriately normal PTH levels. This biochemical phenotype suggests both an abnormality of PTH secretion and renal resistance to PTH.

Activating mutations in the calcium-sensing receptor cause autosomal dominant hypocalcemia (ADH) associated with inappropriately normal PTH levels. The clinical syndrome is variable; patients present with hypocalcemia and seizures, whereas their affected relatives may be only subsequently diagnosed with asymptomatic hypocalcemia. Unlike in patients with inactivating mutations of the calcium sensor, homozgyously affected persons do not appear to have a more severe phenotype. The presence of hypercalculia in these patients makes medical management uniquely challenging.

Treatment with vitamin D metabolites often results in a marked increase in renal calcium excretion, associated with renal calcification and resultant renal impairment. On the basis of these observations, it has been suggested that asymptomatic persons be left untreated and that the goal of therapy in patients with symptomatic hypocalcemia be solely to relieve symptoms, not to achieve normocalcemia.

Pseudohypoparathyroidism

The idiopathic and inherited forms of PTH resistance are referred to as pseudohypoparathyroidism (PHP). The first cases of documented PTH resistance were
described by Albright in 1942. The patients were hypocalcemic and hyperphosphatemic and also exhibited a number of clinical features including short stature, round face, foreshortened fourth metacarpals, obesity, and subcutaneous calcifications (Fig. 26-33 and Fig. 26-34). These are the characteristics of the disorder now called Albright’s hereditary osteodystrophy (AHO).

PTH administration to affected patients failed to provoke a phosphate diuresis or an increase in serum calcium. It was subsequently demonstrated that hypocalcemic patients with features of AHO had elevated PTH levels and that PTH infusions failed to stimulate renal production of cAMP. Failure of stimulation of cAMP production suggested a defect in the PTH receptor or in its cAMP-mediated signal transduction. The measurement of cAMP in the urine following an infusion of synthetic PTH(1-34) is now used to establish the diagnosis of PTH resistance.

The variable presence of AHO and renal resistance to PTH in PHP has led to the subclassification of pseudohypoparathyroidism (Table 26-5). Type Ia is characterized by AHO and diminished Gs activity (approximately 50% of normal). The diminished Gs activity has been demonstrated in several tissues and cells, including kidney, fibroblasts, transformed lymphocytes, platelets, and erythrocytes. Decreased amounts of Gs mRNA (50%) are present in fibroblasts of many patients with PHP Ia, and inactivating mutations in the Gs gene have been identified in several kindreds.

Impaired mentation is seen in approximately half of the patients with PHP Ia and appears to be related to the Gs deficiency rather than to chronic hypocalcemia, because patients with other forms of PHP and hypocalcemia have normal mentation. The Gs deficiency in PHP Ia may be associated not only with PTH resistance but also with resistance to other hormones such as TSH, glucagon, and gonadotropins, resulting in thyroidal and gonadal dysfunction. Paradoxically, two unrelated males with both PHP Ia and gonadotropin-independent precocious puberty have been described. The Gs point mutation found in these patients is thought to lead to a protein that is unstable at 37°C and, therefore, to confer renal resistance to PTH. At the lower temperature of the testes, however, the protein is not degraded. In this setting, the stable but mutated protein is constitutively active and stimulates the Leydig cell, in a manner similar to the skeletal effects of the Gs gene.

Patients with PHP type Ib and gonadotropin-independent precocious puberty have been described. The Gs gene from their fathers, they exhibit pseudo-PHP; when they inherit the mutant Gs gene from their mothers, they exhibit PHP. This pattern, in which the phenotype depends on the parent of origin, is termed genetic imprinting; mice with targeted ablation of the Gs gene (GNAS1) also display such imprinting.

The observation of a phenotype in a heterozygous "loss of function" mutation in Gs is in contrast to the findings in mice with targeted deletions of the other G genes (G2, G13), in which a phenotype is observed only in the homozygous state. The fact that the GNAS1 gene is imprinted has partly resolved the dilemma of this dominant phenotype. Notably, mice with targeted ablation of one GNAS1 gene fail to express mRNA in the renal cortex when

<table>
<thead>
<tr>
<th>Disorder</th>
<th>Urinary cAMP Response to PTH</th>
<th>Urinary PO4 Response to PTH</th>
<th>Other Hormonal Resistance</th>
<th>AHO</th>
<th>Pathophysiology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pseudohypoparathyroidism Ia</td>
<td>Decreased</td>
<td>Decreased</td>
<td>Yes</td>
<td>Yes (Gs mutation)</td>
<td></td>
</tr>
<tr>
<td>Pseudohypoparathyroidism Ib</td>
<td>Normal</td>
<td>Normal</td>
<td>No</td>
<td>Yes (Gs mutation)</td>
<td>20q13.3 defect (GNAS1 locus)</td>
</tr>
<tr>
<td>Pseudohypoparathyroidism Ic</td>
<td>Decreased</td>
<td>Decreased</td>
<td>No</td>
<td>No</td>
<td>Vitamin D deficiency or myotonic dystrophy in some cases</td>
</tr>
<tr>
<td>Pseudohypoparathyroidism II</td>
<td>Normal</td>
<td>Decreased</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
</tbody>
</table>

AHO, Albright’s hereditary osteodystrophy; cAMP, cyclic adenosine monophosphate; PO4, phosphate.

the mutant gene is inherited from the mother but have normal expression in the cortex when the mutant gene is inherited from the father. No such imprinting pattern is seen in the inner medulla; this finding correlates with demonstration of PTH but not vasopressin resistance in the mice (and patients).

Patients with PHP type Ib present with hypocalcemia and high PTH levels, and PTH infusions fail to increase urinary cAMP production. However, this disorder is not accompanied by any of the clinical features of AHO, nor is it associated with abnormal Gs levels in fibroblasts. Renal resistance to PTH is the only consistent feature of type Ib; therefore, several investigators have postulated that this syndrome is due to an isolated abnormality of the PTH receptor. However, a search for mutations in the coding exons of the receptor gene failed to reveal a functional receptor abnormality. The target organ manifestations of PHP Ib are variable, with some affected persons demonstrating PTH overactivity in bone and PTH resistance in kidney. Cultured osteoblast-like cells from a patient with this disorder demonstrated normal cAMP responsiveness to PTH, despite the lack of renal responsiveness.

The locus responsible for PHP Ib has been found to reside on chromosome 20q13.3, the same region that contains the GNAS1 gene, encoding Gs. The disease is inherited with the imprinting characteristic of PHP Ia, but mapping studies suggest that the disease mutations are close to but distinct from the Gs-coding region; furthermore, some kindreds have no abnormalities in Gs exonic sequences. In contrast, in one family, a 3-base-pair deletion in the Gs gene was found in three brothers with the syndrome of PHP Ib and in their unaffected mother. In vitro studies demonstrated that this mutation prevented coupling of Gs to the PTH/PTHrP

![Figure 26-33](Image) Daughter (left) and mother (right) with pseudohypoparathyroidism and Albright’s hereditary osteodystrophy.

![Figure 26-34](Image) Radiograph of hand from a patient with pseudohypoparathyroidism and Albright’s hereditary osteodystrophy. Note the shortened fourth metacarpal.
receptor, but not to the TSH or LH receptors. The mutations in a large majority of patients, which do not involve the \( G_s \)-coding region, remain to be defined.

Several patients with AHO and PTH resistance have been found to have normal \( G_s \) activity; the disorder in this subgroup has been designated PHP Ic. Biochemical characterization in one case revealed a significant decrease in the manganese-stimulated adenylate cyclase activity in fibroblast membranes of the affected person, raising the possibility that a second defect in the cAMP pathway may lead to the phenotype of PHP Ic. In a more recently analyzed patient with PHP Ic, a small deletion in the \( G_s \) gene blocked activation by receptors but did not change \( G_s \) activity in the usual in vitro assays.

A second bone disorder distinct from AHO has been associated with \( G_s \) gene mutations that are paternally inherited. These patients have progressive osseous heteroplasia, a disabling disorder associated with ossification of skin, muscle, and other connective tissue.

In PHP II, PTH infusions increase urinary cAMP normally; however, PTH does not elicit a phosphaturic response. In this syndrome, as in PHP Ib, signs of AHO or resistance to other hormones are lacking, but unlike in PHP Ib, the disorder is not familial in origin. The age at onset of this disorder is variable, ranging from infancy to senescence, suggesting that it is an acquired defect or that the biochemical phenotype may be unmasked by intercurrent abnormalities. A subset of patients with myotonic dystrophy display the biochemical features of PHP II, the degree of PTH resistance correlating with the degree of expansion of the pathogenetic CTG repeats in the myotonin protein kinase gene. A similar biochemical phenotype can also be observed in vitamin D deficiency, and some authors have suggested that PHP II is a manifestation of vitamin D deficiency rather than a distinct clinical entity.

Minagawa and colleagues reported on cases of three neonates with no signs of rickets and with normal levels of vitamin D who presented with transient PHP II that resolved at about 6 months of age. These workers postulated that PTH responsiveness is subject to maturation during fetal and neonatal development. PHP II, therefore, seems to reflect a heterogeneous clinical disorder associated with defects in PTH responsiveness distal to cAMP generation or involving a separate signal transduction pathway.

The resistance to PTH in PHP has not been documented in bone cells; rather, several patients with PHP Ib have been reported to have skeletal changes consistent with hyperparathyroidism. Patients with PHP have lower bone density than that in normal persons and hypoparathyroid patients. Basal urinary hydroxyproline excretion in patients with PHP is twice that in hyperparathyroid patients, and they have similar increases in response to parathyroid extract. Because the markers of bone turnover in patients with PHP are not as high as those in hyperparathyroid patients with similar or lower PTH levels, it has been postulated by some authors that the PTH resistance in bone is relative. However, normal cAMP response has been documented in osteoblasts isolated from patients with PHP Ia and PHP Ib. This finding suggests that the hypocalemia in PHP is not secondary to skeletal resistance but is a consequence of the renal resistance to PTH that results in both increased urinary calcium losses and impaired 25(OH)D \( \rightarrow \) hydroxylation. The lack of activation of vitamin D results in diminished intestinal calcium absorption and osteomalacia, both of which further exacerbate the hypocalemia. Deficiency of 1,25(OH)\(_2\)D \( \rightarrow \) and the resultant hypocalemia can, in turn, impair the phosphaturic responses to PTH but not the urinary cAMP responses to PTH; therefore, it is imperative that studies to confirm the diagnosis of PHP II be performed in normocalcemic patients who have normal vitamin D status.
Hypocalcemia secondary to vitamin D deficiency or resistance to the biologic effects of 1,25(OH)\textsubscript{2}D\textsubscript{3} is easily differentiated from the hypocalcemia of hyperparathyroidism by routine clinical and laboratory evaluation. The primary cause of hypocalcemia in vitamin D deficiency is decreased intestinal absorption of calcium. In the setting of normal renal function, the hypocalcemia of vitamin D deficiency, unlike that of hypoparathyroidism, is accompanied by hypophosphatemia and increased renal phosphate clearance. This increase in phosphate clearance is a direct result of compensatory (secondary) hyperparathyroidism. The hyperparathyroidism is a consequence of the hypocalcemic stimulus to PTH secretion and the stimulation of PTH gene expression and parathyroid cell proliferation caused by hypocalcemia (see “Parathyroid Hormone Biosynthesis” earlier). Therefore, measurement of serum phosphate and PTH are very useful in distinguishing these disorders from hyperparathyroidism. The secondary hyperparathyroidism results in increased calcium mobilization from the skeleton, increased renal reabsorption of calcium, and increased renal 1-hydroxylation of 25(OH)D. In severe vitamin D deficiency, the increased levels of PTH no longer lead to increased bone resorption, perhaps because osteoclasts appear not to resorb unmineralized osteoid.

In profound vitamin D deficiency, the level of 1.25(OH)\textsubscript{2}D\textsubscript{3} is usually low; in moderate vitamin D deficiency, the stimulation of the renal 1-hydroxylase by PTH can result in a normal or even elevated 1.25(OH)\textsubscript{2}D\textsubscript{3} level. These high levels of 1.25(OH)\textsubscript{2}D\textsubscript{3} reflect the action of PTH on the renal 1-hydroxylase. The ineffectiveness of the high levels of total 1.25(OH)\textsubscript{2}D\textsubscript{3} to normalize serum calcium may be explained by increased binding of this metabolite to vitamin D-binding protein when the levels of 25(OH)D are very low.

**Vitamin D Deficiency**

Because the two sources of vitamin D are the diet and cutaneous synthesis after ultraviolet radiation, lack of solar radiation and decreased intake or impaired absorption of vitamin D can lead to vitamin D deficiency. As the population has become increasingly educated about the risks of skin cancer from solar radiation, the avoidance of long periods of intense sun exposure, coupled with the use of high-SPF (sun protection factor) sun blocks, has resulted in increased reliance on dietary sources of vitamin D. The RDA for vitamin D is 200 IU; however, in the absence of solar exposure, this recommendation is two to three times lower than that required to prevent vitamin D deficiency.

Vitamin D is present in many food sources, both vegetable and animal. In addition, many prepared foods, especially cereals, are fortified with vitamin D. Although dairy products have been fortified with vitamin D as well, the actual amount of vitamin D provided does not correlate well with the purported content. The vitamin D derived from vegetable sources is vitamin D\textsubscript{3}, and that from animal sources is vitamin D\textsubscript{2}. These two forms of vitamin D are metabolized identically and have equivalent biologic potency in humans. Both forms have been used to fortify foods.

Early vitamin D deficiency can be detected when the serum level of 25(OH)D falls below 15 ng/dL, because this level has been shown to be associated with the development of secondary hyperparathyroidism. Although elderly, homebound individuals are at high risk, several studies have demonstrated that vitamin D deficiency is prevalent in the general population (reviewed in reference 618). The clinical relevance of this vitamin D deficiency has been confirmed by a study demonstrating that vitamin D administration (800 IU/day) to an ambulatory elderly population decreases serum PTH levels as well as the incidence of hip fracture.

Malabsorption also remains an important cause of vitamin D deficiency in all age groups. Because vitamin D is a fat-soluble vitamin, its absorption is dependent on emulsification by bile acids. Any cause of fat malabsorption or short bowel syndrome can result in vitamin D deficiency; therefore, malabsorption should be ruled out in patients with low 25(OH)D levels (<8 ng/dL).

**Accelerated Loss of Vitamin D**

25-Hydroxyvitamin D and 1,25(OH)\textsubscript{2}D\textsubscript{3} are secreted with bile salts and undergo enterohepatic circulation; therefore, intestinal disease may also result in vitamin D deficiency owing to excessive losses. Increased metabolism of vitamin D, leading to low blood levels of 25(OH)D, is seen in individuals given anticonvulsant medications and antituberculous therapy. Phenobarbital, primidone, phenytoin, rifampin, and glutethimide have all been reported to accelerate the hepatic inactivation of vitamin D.

**Impaired 25-Hydroxylation of Vitamin D**

The vitamin D that is absorbed undergoes 25-hydroxylation in the liver; therefore, severe hepatic parenchymal damage can result in 25(OH)D deficiency. Clinically, severe vitamin D deficiency secondary to liver disease is rare, because the degree of hepatic destruction necessary to impair 25-hydroxylation is incompatible with long-term survival. However, isoniazid has been shown to decrease the 25-hydroxylation of vitamin D. Two kindreds have been described in whom the clinical and biochemical presentations and therapeutic responses suggest an inherited 25-hydroxylation defect.

**Impaired 1-Hydroxylation of 25-Hydroxyvitamin D**

The clinical relevance of this vitamin D deficiency has been confirmed by a study demonstrating that vitamin D administration (800 IU/day) to an ambulatory elderly population decreases serum PTH levels as well as the incidence of hip fracture.

The metabolic consequences of chronic renal failure on the parathyroid glands and the skeleton are complex (see Chapter 27). Impaired renal 1-hydroxylation leads to decreased intestinal absorption of calcium, resulting in hypocalcemia. The diminished phosphate clearance associated with renal failure leads to elevated levels of blood phosphate; this change, in turn, lowers levels of calcium and 1,25(OH)\textsubscript{2}D\textsubscript{3}. The resultant secondary hyperparathyroidism increases release of calcium and phosphate from bone; however, because of the renal insufficiency, PTH does not have a phosphaturic effect. As a result, the increased serum phosphate rises further.

Oral phosphate binders are used to lower blood phosphate. Calcium-containing antacids are now used as oral phosphate binders, in preference to the more toxic aluminum-containing antacids (see Chapter 27). Calcium administration also attenuates the hypocalcemic stimulus to parathyroid secretion. 1,25-Dihydroxyvitamin D\textsubscript{3} therapy is crucial for the absorption of this calcium and should be instituted early in the course of renal failure (when the creatinine clearance falls below 30 to 40 mL/min) to avoid the development of secondary hyperparathyroidism, with careful monitoring to avoid hypercalcemia. Once secondary hyperparathyroidism has developed, pharmacologic doses of 1,25(OH)\textsubscript{2}D\textsubscript{3} delivered intravenously or orally may be required to suppress PTH gene transcription and parathyroid cellular proliferation.
of malignancies whose proliferation is inhibited by pharmacologic doses of 1,25(OH)\(_2\)D\(_3\).

Decreased levels of 1,25(OH)\(_2\)D\(_3\) may also be observed in patients taking ketoconazole\(^{61}\) and in those with oncogenic osteomalacia\(^{28}\) (see Chapter 27).

A rare heritable defect of vitamin D activation has been described in several kindreds. Biochemically, vitamin D-dependent rickets type I (VDDR I) is characterized by hypocalcemia and secondary hyperparathyroidism. The only feature that differentiates this disorder from dietary vitamin D deficiency is the presence of normal or elevated levels of vitamin D and 25(OH)D accompanied by low levels of 1,25(OH)\(_2\)D\(_3\).\(^{62,63}\) The disease is inherited in an autosomal recessive fashion; those affected present in infancy with rickets, osteomalacia, and seizures. Cloning of the 1-hydroxylase gene has confirmed that mutation of this gene constitutes the molecular basis for the disorder.\(^{64,65,66}\) and as expected, physiologic replacement doses of 1-hydroxylated metabolites of vitamin D result in clinical remission.\(^{67,68}\)

**Target Organ Resistance to 1,25-Dihydroxyvitamin D**

A second rare inherited disorder characterized by resistance to the biologic actions of 1,25(OH)\(_2\)D\(_3\) has been described in several kindreds. This disorder, referred to as vitamin D-dependent rickets type II (VDDR II), is also characterized by autosomal recessive inheritance. Its biochemical presentation, with hypocalcemia, hypophosphatemia, and secondary hyperparathyroidism, resembles that of vitamin D deficiency, but it is accompanied by elevated levels of 1,25(OH)\(_2\)D\(_3\). The molecular basis for this disease is mutation of the vitamin D receptor gene (VDR), resulting in impaired target organ responsiveness. Most of the mutations that have been described involve the DNA binding domain of the receptor.\(^{69,70,71,72,73}\) These mutations result in a decreased affinity of the receptor for its response elements on target genes,\(^{74,75}\) leading to impaired regulation of these genes. Mutations in the hormone-binding domain\(^{76}\) of the receptor have also been described in kindreds with VDDR II.

The clinical presentation of VDDR II is variable; however, most patients present in infancy with rickets, hypophosphatemia, and seizures, although presentation in late adolescence has also been described.\(^{77}\) Alopecia totalis, developing in the first 2 years of life, is present in some kindreds.\(^{78}\) The finding of alopecia in mice with VDR mutations\(^{79,80}\) confirms the association of alopecia with disruption of the vitamin D receptor gene.

Because of the target organ resistance to the active metabolite of vitamin D, there is no ideal treatment for VDDR II. Pharmacologic doses of vitamin D, 25(OH)D, 24,25(OH)\(_2\)D, and 1,25(OH)\(_2\)D\(_3\) have been administered in an attempt to overcome this target organ resistance.\(^{81,82,83}\) with variable effects. In those patients in whom the hypocalcemia and osteomalacia are resistant to such therapeutic interventions, parenteral calcium infusions have been used to heal osteomalacic lesions.\(^{84}\) Studies in VDR-ablated mice have demonstrated that maintenance of normal mineral ion homeostasis prevents all of the complications of VDR ablation except alopecia.\(^{85,86}\)

On the basis of these observations, patients with VDR mutations should receive early and aggressive treatment to prevent skeletal abnormalities and parathyroid hyperplasia. Lifelong therapy is usually required, although spontaneous remissions off therapy have been described.\(^{87,88,89,90}\) The pathophysiology of the spontaneous remissions is not well understood, because the underlying genetic defect remains. It is likely that these "remissions" reflect compensated calcium homeostasis that obtains once the needs of the growing skeleton are met. In support of this hypothesis is a report of a relapse in a pregnant woman, followed by a remission post partum.\(^{91}\)

Phenytoin causes target organ resistance to the biologic effects of 1,25(OH)\(_2\)D\(_3\), in addition to its acceleration of the hepatic catabolism of vitamin D metabolites. Phenytoin has been shown to impair intestinal calcium absorption in vivo in rats\(^{92}\) and to impair PTH- and 1,25(OH)\(_2\)D\(_3\)-mediated bone resorption in vitro. Combination chemotherapy with 5-fluorouracil and low-dose leucovorin has been reported to cause hypocalcemia in 65% of patients, associated with an acute decrease in plasma 1,25(OH)\(_2\)D\(_3\) levels.\(^{93}\)
Other Causes of Hypocalcemia

Excessive Deposition into the Skeleton

Excessive deposition of calcium into the skeleton can occur in association with osteoblastic metastases, with chondrosarcomas, 93 or in the hungry bone syndrome. This syndrome is characterized by prolonged hypocalcemia, hypocalcinuria, and hypophosphatemia following parathyroidectomy for primary hyperparathyroidism (see "Primary Hyperparathyroidism" earlier). It is a consequence of remineralization of a skeleton that has been subjected to the bone-resorbing effects of PTH over a prolonged period. Hungry bone syndrome can also be observed after treatment of other diseases that are associated with excessive bone resorption. It has been described following radioactive iodine treatment for Graves’ disease. 113

Chelation

Decreases in ionized calcium have been reported with foscarnet, a pyrophosphate analogue that is used as an antiviral agent. 29 This effect may be secondary to complex formation between ionized calcium and the drug.

Hyperphosphatemia, due to phosphate administration or rapid destruction of soft tissue (e.g., with rhabdomyolysis or chemotherapy of hematologic malignancies), may produce profound hypocalcemia by directly complexing and precipitating calcium in bone or soft tissues, by inhibiting bone resorption, and by blocking renal synthesis of 1,25(OH)2 D3 (see "Hyperphosphatemia" later).

Massive infusions of citrated blood products may cause hypocalcemia, presumably because citrate complexes calcium in the recipient's plasma. 29 Large doses of ethylenediaminetetraacetic acid (EDTA)-containing radiographic contrast dyes have also been reported to cause hypocalcemia. Hypocalcemia, due to formation of complexes of calcium and fluoride, has been reported with hydrofluoric acid burns 159 and ingestion. 159

Neonatal Hypocalcemia

Neonatal hypocalcemia is seen in infants of hyperparathyroid mothers and of diabetic mothers, in premature infants, and in infants with birth asphyxia. The cause of hypocalcemia in infants of diabetic mothers is probably multifactorial. Prematurity per se does not account for the higher incidence. 106 The response of premature infants and infants of diabetic mothers to exogenous PTH suggests that functional hypoparathyroidism may account in part for the increased hypocalcemia in these two populations. 113, 116 The hypocalcemia in infants of hyperparathyroid mothers is presumably secondary to the maternal hypercalcemia, which in turn suppresses fetal parathyroid function. 159

Human Immunodeficiency Virus Infection

Hypocalcemia is 6 times more prevalent in human immunodeficiency virus (HIV)-infected patients than in the general population. 110 Although hypocalcemia is often a consequence of antiretroviral and antibiotic or antifungal therapy, vitamin D deficiency and hypomagnesemia are also common in patients with AIDS. Impaired parathyroid responsiveness to hypocalcemia has also been documented.

Critical Illness

Hypocalcemia is commonly seen in critically ill patients and is thought to be a reflection of parathyroid gland suppression, failure to activate vitamin D, calcium chelation or sequestration, and hypomagnesemia. However, one study has demonstrated an increased basal level and secretory response of PTH to lowering serum calcium in both septic and nonseptic intensive care unit (ICU) patients. 29 A correlation between cytokine levels and hypocalcemia was noted in this and other studies, suggesting that these inflammatory agents may play a role in redistribution of calcium to the intracellular or other pools.

Severe acute pancreatitis is often associated with hypocalcemia, and this association is a negative prognostic indicator. The hypocalcemia occurs shortly after the onset of the pancreatitis and is associated with an increase in PTH levels, suggesting that parathyroid function is normal. It has long been thought that this hypocalcemia is secondary to deposition of calcium "soaps" consisting of calcium and fatty acids. Supporting this hypothesis, studies in a patient with a pancreatic fistula demonstrated hypocalcemia (calcium concentration of 4.3 mg/dL).

**TABLE 26-6 -- Therapeutic Mineral Ion Preparations**

<table>
<thead>
<tr>
<th>Available Formulation</th>
<th>Compound</th>
<th>Oral Preparation</th>
<th>Parenteral Preparation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Compound</td>
<td>Ion Content</td>
<td>Compound</td>
</tr>
<tr>
<td>Calcium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calcium carbonate</td>
<td>100</td>
<td>10.0</td>
<td>1250 mg</td>
</tr>
<tr>
<td>Calcium phosphate</td>
<td>310</td>
<td>9.6</td>
<td>1565 mg</td>
</tr>
<tr>
<td>Calcium acetate</td>
<td>158</td>
<td>6.3</td>
<td>668 mg</td>
</tr>
<tr>
<td>Calcium citrate</td>
<td>498</td>
<td>6.0</td>
<td>950 mg</td>
</tr>
<tr>
<td>Calcium lactate</td>
<td>218</td>
<td>4.6</td>
<td>650 mg</td>
</tr>
<tr>
<td>Calcium gluconate</td>
<td>64</td>
<td>1.7</td>
<td>5 mL</td>
</tr>
<tr>
<td>Calcium gluconate</td>
<td>430</td>
<td>2.3</td>
<td>1000 mg</td>
</tr>
<tr>
<td>Calcium gluconate</td>
<td>468</td>
<td>2.0</td>
<td>82 mg</td>
</tr>
<tr>
<td>Calcium chloride</td>
<td>147</td>
<td>6.8</td>
<td>273 mg</td>
</tr>
<tr>
<td>Magnesium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Magnesium oxide</td>
<td>40</td>
<td>24.8</td>
<td>400 mg</td>
</tr>
<tr>
<td>Magnesium gluconate</td>
<td>450</td>
<td>2.2</td>
<td>500 mg</td>
</tr>
<tr>
<td>Magnesium chloride</td>
<td>203</td>
<td>4.9</td>
<td>535 mg</td>
</tr>
<tr>
<td>Magnesium chloride</td>
<td>246</td>
<td>4.1</td>
<td>99 mg</td>
</tr>
<tr>
<td>Magnesium sulfate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phosphorus</td>
<td>Sodium/potassium phosphate (neutral)</td>
<td>capsule</td>
<td>250 mg</td>
</tr>
<tr>
<td></td>
<td>Potassium phosphate (neutral)</td>
<td>capsule</td>
<td>250 mg</td>
</tr>
<tr>
<td>Sodium phosphate (neutral)</td>
<td>soln</td>
<td>94 mg/mL</td>
<td>3.0 mmol/mL</td>
</tr>
<tr>
<td>---------------------------</td>
<td>------</td>
<td>----------</td>
<td>-------------</td>
</tr>
</tbody>
</table>

* Molecular weight (MW) shown is for the usual chemical form, including water molecules (e.g., MgSO₄· 7 H₂O).

Other formulations exist. Those shown are among those approved in the United States. Phosphate preparations contain buffered mixtures of monobasic (H₂PO₄⁻) and dibasic (HPO₄²⁻) ions; the phosphorus content therefore is specified in millimoles. Oral phosphates contain 7 mEq sodium and potassium per capsule (sodium/potassium form) or 14 mEq potassium/capsule (potassium form). Parenteral solutions typically contain 4 mEq of sodium or potassium per mL.


in the setting of high levels of calcium (26 mg/dL) and fatty acids in ascitic fluid. Subsequent studies in a rat model supported this finding and demonstrated that oleate has a high binding capacity for calcium. However, other investigations in a porcine model of experimental pancreatitis have demonstrated that hypocalcemia does not occur if the animals are subjected to thyroidectomy before the induction of pancreatitis. This finding suggests a role for calcitonin in the development of hypocalcemia with acute pancreatitis, although several clinical studies have documented normal calcitonin levels in hypocalcemic persons with pancreatitis. Severe hypocalcemia with hypercalcitoninemia and hypophosphatemia has been reported in patients with the toxic shock syndrome or sepsis and in critically ill patients. As in acute pancreatitis, this hypocalcemia is usually accompanied by increases in serum levels of PTH, and the degree of hypocalcemia is a negative prognostic indicator. The mechanism of hypocalcemia in these patients is likely to be heterogeneous and has not been clearly defined.
hours with close monitoring of calcium levels. In hypocalcemia associated with hypomagnesemia, magnesium replacement also is required. Magnesium should be given intravenously. 100 mEq infused over 24 hours in the acute setting. Because most of the parenteral magnesium is excreted in the urine, oral magnesium oxide should be instituted as soon as possible for repletion of body stores. Special caution and use of reduced doses are necessary in administering magnesium to patients in renal failure (see “Disorders of Magnesium Metabolism” later).

The treatment of hypocalcemia should be directed at the underlying disorder. In all cases, replacement with exogenous calcium (1 to 3 g of elemental calcium daily) should be instituted. Calcium carbonate is the least expensive formulation but requires acidification for efficient absorption. This requirement becomes important in patients with achlorhydria and those in whom gastric acid production is being suppressed with pharmacologic agents. Notable in this respect is the acid-buffering capacity of calcium carbonate. Because of this, it is recommended that patients take their calcium carbonate supplements in divided doses of 1 g or less. In such cases, the calcium should be taken with food or citrus drinks to promote maximal absorption.

In cases of vitamin D deficiency or resistance, the metabolite of vitamin D chosen depends on the underlying disorder. If impaired renal 1-hydroxylation is present, such as in renal failure, hyperparathyroidism (or PTH resistance), or the vitamin D-dependent rickets syndromes, metabolites that do not require this modification should be administered (calciotrol 0.25 to 1 µg per day or dihydrotachysterol 0.2 to 1 mg/day). If decreased intake or increased losses are the problem, vitamin D should be administered and the treatment directed at the underlying disorder. Initial repletion of stores can be undertaken with 50,000 IU of vitamin D daily for 2 to 3 weeks, followed by weekly or bimonthly administration until the underlying problem has been corrected.

In patients with resistance to vitamin D, such as those receiving phenytoin, high doses (50,000 IU one to three times weekly) should be used as maintenance therapy. In other patients, once treatment of the underlying disorder and repletion of body stores have been addressed, two multivitamins (800 IU) should provide sufficient maintenance therapy. In cases of severe malabsorption, vitamin D can be administered parenterally (in intravenous hyperalimentation formulations or as intramuscular injections of as much as 500,000 IU every 6 months).

Patients should be monitored closely both to assess response to therapy and to prevent therapeutic complications. Serum calcium should be monitored frequently (daily with profound hypocalcemia, weekly with moderate hypocalcemia) for the first month of therapy. Concomitant with resolution of hypocalcemia, the serum PTH level should decline as the secondary hyperparathyroidism resolves. Measurement of serum PTH and assessment of 24-hour urinary calcium excretion should be performed within 2 to 4 weeks of institution of therapy. The urinary calcium measurement reflects the effect of therapy on the patient’s ability to absorb calcium. In addition, it provides important information on which to base therapeutic modifications to avoid nephrolithiasis.

Once normalization of serum and urinary calcium and a decrease in PTH levels are observed, a transition from aggressive replacement therapy to maintenance therapy should be undertaken to prevent hypercalcemia and nephro lithiasis. These same parameters should be monitored 1 and 3 months after a dosage change to assess the effect of the therapeutic intervention. Monitoring of alkaline phosphatase levels can also be performed at this time. Alkaline phosphatase levels may actually increase soon after treatment is started because of healing of the osteomalacic lesions; however, by 3 to 4 months after institution of therapy, a clear downward trend should be observed. Alkaline phosphatase and PTH values may remain elevated for 6 to 12 months after institution of therapy; persistence of elevated values should not be a cause for alarm, provided that levels are declining and that the other parameters suggest that therapy is effective.

The treatment of hyperparathyroidism is similar to that of vitamin D deficiency, with the exception that these patients have impaired renal 1-hydroxylation of 25(OH)D and therefore require treatment with 1-hydroxylated metabolites. PTH has been used experimentally for the treatment of hyperparathyroidism but is not approved for therapy. Further work is needed to determine the role of this expensive parenteral therapy. Oral calcium and 1-hydroxylated vitamin D metabolites, therefore, remain the mainstay of therapy. Monitoring of serum and urinary calcium levels should be performed as in the treatment of vitamin D deficiency. Therapy in these patients is lifelong; accordingly, careful monitoring is required to avoid renal or hypercalcemic complications. The aim of therapy should be to maintain serum calcium in the low-normal range without causing frank hypercalcuria. Because PTH plays an important role in renal calcium reabsorption, difficulties are often encountered attaining these therapeutic goals. In such cases, renal calcium losses can be minimized by the addition of a thiazide diuretic to the treatment regimen.

One of the frustrations often encountered in treating patients with hyperparathyroidism is the fluctuating response to a seemingly stable therapeutic regimen. Episodes of hypercalcemia are occasionally observed without any discernible cause. Because of this, serum calcium should be monitored every 3 months to permit temporary withdrawal of 1,25(OH)2D3 should a hypercalcemic trend be observed. Fortunately, the half-life of this metabolite is short, so that discontinuation for a few days to a week with resumption of a lower dose is usually efficacious.

All patients receiving vitamin D metabolites and calcium need to be aware of potential therapeutic complications. The mild symptoms of hypercalcemia must be emphasized to the patient. It is essential that these patients be aware that their calcium level should be monitored more frequently during intercurrent illnesses that may affect calcium absorption or hydration status, to prevent the development of hypocalcemia or severe hypercalcemia.
DISORDERS OF PHOSPHATE METABOLISM

Hyperphosphatemia

The kidney exerts dominant control over the serum concentration of phosphate by adjusting the rate of tubular phosphate reabsorption, mainly in response to changes in the serum concentration of PTH. At near-normal daily loads of filtered phosphate, the capacity of the kidney to excrete phosphate is not easily exceeded. Consequently, when hyperphosphatemia occurs, it is generally in the setting of significantly impaired renal function, hypoparathyroidism, or a greatly increased flux of phosphate into the extracellular fluid, or some combination of these factors (Table 26-7).

The most common cause of hyperphosphatemia is acute or chronic renal failure in which GFR is so reduced that the usual daily load of phosphate cannot be excreted at a normal level of serum phosphate, despite maximal inhibition of phosphate reabsorption in the remaining functional nephrons. In hypoparathyroidism (or pseudohypoparathyroidism), serum phosphate levels may rise to levels as high as 6 to 8 mg/dL, owing to loss of the tonic inhibitory effect of PTH on phosphate reabsorption. The hyperphosphatemia of hypoparathyroidism is only partly due to the absence of PTH per se. Hypocalcemia may further impair phosphate clearance in this setting, and correction of hypocalcemia by treatment with vitamin D metabolites and oral calcium may reduce serum phosphate, for example, even though PTH levels remain low, a response attributed to a direct renal tubular phosphaturic effect of the raised serum calcium level.

Other circumstances in which renal tubular phosphate excretion is decreased, in the absence of renal failure, include (1) acromegaly, or heparin, and (3) tumoral calcinosis. Tumoral calcinosis is an unusual disorder of unknown cause that appears frequently to have a genetic basis consistent with autosomal recessive inheritance, although dominant transmission also has been reported. Affected patients may display focal hyperostosis, large, lobulated perilobular ectopic calcifications, especially around shoulders or hips; hyperphosphatemia due to increased renal tubular reabsorption of phosphate; increased serum 1,25(OH)2D3 levels and intestinal hyperabsorption of calcium; and increased intestinal calcium absorption, consistent with the elevated serum 1,25(OH)2D3 concentration. The disorder may present in childhood or adulthood, is more common in blacks, and is lifelong, with a tendency for the tumoral calcifications to progress at affected sites.

Tumoral calcinosis appears to result from a primary overproduction of 1,25(OH)2D3, although unlike in idiopathic hypercalciuria, urinary calcium usually is not high, in a response attributed to a direct renal tubular phosphaturic effect of the raised serum calcium level.

Other circumstances in which renal tubular phosphate excretion is decreased, in the absence of renal failure, include (1) acromegaly, or heparin, and (3) tumoral calcinosis. Tumoral calcinosis is an unusual disorder of unknown cause that appears frequently to have a genetic basis consistent with autosomal recessive inheritance, although dominant transmission also has been reported. Affected patients may display focal hyperostosis, large, lobulated perilobular ectopic calcifications, especially around shoulders or hips; hyperphosphatemia due to increased renal tubular reabsorption of phosphate; increased serum 1,25(OH)2D3 levels and intestinal hyperabsorption of calcium; and increased intestinal calcium absorption, consistent with the elevated serum 1,25(OH)2D3 concentration. The disorder may present in childhood or adulthood, is more common in blacks, and is lifelong, with a tendency for the tumoral calcifications to progress at affected sites.

Hyperphosphatemia may result from overly rapid administration of therapeutic phosphate preparations, especially if renal function is compromised, or from rapid shifts of phosphate out of cells, most often provoked by mechanical or metabolic injury. Most cases of hyperphosphatemia associated with intestinal phosphate loads have involved children who received phosphate-containing laxatives or enemas. Hyperphosphatemia due to cytolytic release of intracellular phosphate can be quite dramatic, with serum phosphate concentrations of 20 mg/dL or greater. This problem was described initially as a complication of rapid-induction chemotherapy for certain hematologic malignancies, although it also may occur from cellular injury associated with trauma, hyperthermia, overwhelming infection, hemolysis, rhabdomyolysis, or metabolic acidosis.

Most often, hyperphosphatemia is mild and asymptomatic, although chronic hyperphosphatemia is an important factor in the development of secondary hyperparathyroidism in progressive renal failure. The clinical manifestations of acute, severe hyperphosphatemia are related mainly to those of the accompanying hypocalcemia, caused by formation of insoluble calcium phosphate precipitates. Thus, tetany, muscle cramps, paresthesias, and seizures may occur, compounded by
other metabolic disturbances (hyperkalemia, acidosis, hyperuricemia) that frequently coexist. Generalized precipitation of calcium phosphate into soft tissues may produce renal failure.

Therapeutic options for hyperphosphatemia are limited. Volume expansion may be helpful to improve GFR in acute syndromes. Identification and removal of any exogenous sources of phosphate are important, and phosphate-binding aluminum hydroxide antacids may be useful in limiting intestinal phosphate absorption and chelating phosphate secreted into the intestine. Hemodialysis is the most effective approach and should be considered early in severe hyperphosphatemia, especially in the tumor lysis syndrome and particularly if symptomatic hypocalcemia cannot be adequately treated for fear of inducing widespread soft tissue calcification.
Hypophosphatemia

Etiology

Hypophosphatemia may result from one or more of three general mechanisms: increased urinary losses due to decreased net renal tubular phosphate reabsorption; rapid shifts of phosphate from extracellular fluid into the intracellular space or the mineral phase of bone; or, rarely, severe and selective deprivation of dietary phosphate, as may occur with chronic ingestion of large amounts of nonabsorbable aluminum-containing antacids. Fasting or starvation does not lead directly to hypophosphatemia, apparently because phosphate is mobilized from catabolized bone and soft tissue in amounts sufficient to maintain serum phosphate, even during prolonged caloric deprivation. Starvation does induce phosphate deficiency and therefore predisposes the affected person to subsequent hypophosphatemia.

Chronic hypophosphatemia usually can be traced to ongoing renal phosphate wasting. Elevation of serum PTH for any reason (other than renal failure), as in primary hyperparathyroidism or secondary hyperparathyroidism due to vitamin D or calcium deficiency, results in inhibition of tubular phosphate reabsorption and fasting hypophosphatemia. Phosphate clearance also is increased by PTHrP in patients with humoral hypercalcemia of malignancy.

As noted previously, hypercalcemia exerts a direct phosphaturic effect on the renal tubules that is most evident in hypoparathyroid persons but that also may contribute to the hypophosphatemia in primary hyperparathyroidism. Nevertheless, in severely hypercalcemic patients, hypophosphatemia may be masked acutely by accompanying dehydration and reduced GFR. When severe hypocalcemia and hypomagnesemia coexist, selective correction of hypomagnesemia may provoke hypophosphatemia.

---

**TABLE 26-8 -- Causes of Hypophosphatemia**

<table>
<thead>
<tr>
<th>Reduced Renal Tubular Phosphate Reabsorption</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTH/PTHrP-Dependent</td>
</tr>
<tr>
<td>Primary hyperparathyroidism</td>
</tr>
<tr>
<td>PTHrP-dependent hypercalcemia of malignancy</td>
</tr>
<tr>
<td>Secondary hyperparathyroidism</td>
</tr>
<tr>
<td>Vitamin D deficiency/resistance</td>
</tr>
<tr>
<td>Calcium starvation or malabsorption</td>
</tr>
<tr>
<td>Rapid, selective correction of severe hypomagnesemia</td>
</tr>
<tr>
<td>PTH-Independent</td>
</tr>
<tr>
<td>Familial hypophosphatemic rickets (X-linked hypophosphatemic rickets)</td>
</tr>
<tr>
<td>Fanconi syndrome(s), other renal tubular disorders</td>
</tr>
<tr>
<td>Cystinosis</td>
</tr>
<tr>
<td>Amyloidosis</td>
</tr>
<tr>
<td>Hemolytic uremic syndrome</td>
</tr>
<tr>
<td>Following renal transplantation</td>
</tr>
<tr>
<td>Rewarming or hyperthermia</td>
</tr>
<tr>
<td>Oncogenous osteomalacia syndrome</td>
</tr>
<tr>
<td>Idiopathic hypercalciuria</td>
</tr>
<tr>
<td>Poorly controlled diabetes, alcoholism</td>
</tr>
<tr>
<td>Hyperaldosteronism</td>
</tr>
<tr>
<td>Drugs or toxins</td>
</tr>
<tr>
<td>Ethanol</td>
</tr>
<tr>
<td>Acetazolamide, other diuretics</td>
</tr>
<tr>
<td>High-dose glucocorticoids</td>
</tr>
<tr>
<td>Bicarbonate</td>
</tr>
<tr>
<td>Toluene</td>
</tr>
<tr>
<td>Heavy metals (lead, cadmium)</td>
</tr>
<tr>
<td>Calcitonin</td>
</tr>
<tr>
<td>High-dose estrogens</td>
</tr>
<tr>
<td>Ifosfamide</td>
</tr>
<tr>
<td>Cisplatin</td>
</tr>
<tr>
<td>Suramin</td>
</tr>
<tr>
<td>Foscarnet</td>
</tr>
<tr>
<td>N-methyl formamide</td>
</tr>
<tr>
<td>Pamidronate</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Impaired Intestinal Phosphate Absorption</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aluminum-containing antacids</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Shifts of Extracellular Phosphate into Cells or Bone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute intracellular shifts</td>
</tr>
<tr>
<td>Intravenous glucose, fructose, glycerol</td>
</tr>
<tr>
<td>Insulin therapy for hyperglycemia, diabetic ketoacidosis</td>
</tr>
<tr>
<td>Catecholamines (epinephrine, albuterol, terbutaline, dopamine)</td>
</tr>
<tr>
<td>Acute respiratory alkalosis, salicylate intoxication, acute gout</td>
</tr>
<tr>
<td>Gram-negative sepsis, toxic shock syndrome</td>
</tr>
<tr>
<td>Recovery from acidosis, starvation</td>
</tr>
<tr>
<td>Rapid cellular proliferation</td>
</tr>
<tr>
<td>Leukemic blast crisis</td>
</tr>
</tbody>
</table>
as hypocalcemia persists and PTH secretion increases.

Renal phosphate clearance may be increased in a variety of conditions that do not involve increases in PTH or PTHrP. These conditions include the proximal tubular dysfunction associated with Fanconi's syndrome, severe hypokalemia or hypomagnesemia, acute metabolic alkalosis induced by bicarbonate infusion, and exposure to certain drugs or toxins (see Table 26-7). Excessive phosphaturia also may occur in patients with idiopathic hypercalcemia or primary hyperaldosteronism, with severe glycosuria in poorly controlled diabetes, following renal transplantation, and in Rey's syndrome.

Two syndromes familial X-linked hypophosphatemic rickets (XLH) and oncogenic osteomalacia result from renal tubular phosphate wasting that appears to be driven by humoral mechanisms. Involvement of a humoral mediator is clear-cut in oncogenic osteomalacia, as the syndrome resolves following extirpation of the responsible neoplasm. The renal lesion in XLH is still not well characterized by studies in a relevant animal model. Both syndromes include renal tubular phosphate wasting, impaired renal synthesis of 1,25(OH)₂D₃, borderline hypocalcemia, hypocalcuria, and rickets or osteomalacia. Serum PTH generally is normal or only slightly elevated, and 1,25(OH)₂D₃ is inappropriately normal. The bone disease dominates the clinical picture (see Chapter 27).

Rapid egress of extracellular phosphate into cells is the cause of hypophosphatemia that develops acutely during administration of intravenous glucose, insulin therapy for hyperglycemia, or administration of catecholamines (pressors or bronchodilators) or in profound respiratory alkalosis or leukemic blast crisis. Hypophosphatemia in these situations is more pronounced when the underlying phosphate depletion, as in hyperparathyroidism or vitamin D deficiency, or following prolonged malnutrition or with alcoholism or glycosuria. Accelerated uptake of phosphate into cells, principally into muscle and bone, is particularly likely in postsurgical or trauma patients, in whom it may be promoted by high levels of circulating catecholamines and exacerbated by concurrent respiratory alkalosis, fever, volume expansion, sepsis, and hypokalemia. Similar mechanisms may pertain in nonsurgical illnesses such as acute myocardial infarction. Hypophosphatemia complicating administration of hematopoietic growth factors such as erythropoietin or G-CSF is due to the high demand for new intracellular phosphate imposed by rapid cellular proliferation in the bone marrow.

Clinical Features

The clinical significance of hypophosphatemia has been the subject of some controversy and probably depends on the presence and severity of underlying phosphate depletion. Unfortunately, the status of the total-body phosphorus pool is reflected only indirectly by the concentration of phosphate in the extracellular fluid, which contains less than 5% of body phosphorus. Thus, although serum phosphate concentrations generally are used to characterize hypophosphatemia as severe (<11.5 mg/dL, or <0.30.5 mM), moderate (1.52.2 mg/dL, or 0.50.7 mM), or mild (2.23.0 mg/dL, or 0.751.0 mM), the serum phosphate level may be normal or even high (depending on renal function) in the presence of profound intracellular phosphate deficiency. Conversely, it may be low when intracellular phosphate is relatively normal, as following a sudden movement of extracellular phosphate into cells.

The prevalence of severe hypophosphatemia among hospitalized patients is less than 1%, whereas mild or moderate hypophosphatemia may be detected in 2% to 5% of patients. Hypophosphatemia is recognized most often in critically ill patients, alcoholics or other malnourished individuals, diabetics, and those with acute infectious or pulmonary disorders.

The clinical manifestations of severe hypophosphatemia are protean. Among the most common are various neuromuscular symptoms, ranging from progressive lethargy, muscle weakness, and paresthesias to paralysis, coma, and even death, depending on the severity of the phosphate deficiency. Confusion, profound weakness, paralysis, seizures, and other major sequelae generally are limited to patients with serum phosphate concentrations below 0.8 to 1.0 mg/dL.

Biochemical evidence of muscle injury is observed within 1 or 2 days in over a third of patients whose serum phosphate concentrations falls to less than 2 mg/dL. Overt rhabdomyolysis also may occur, especially in the setting of chronic alcoholism with underlying malnutrition and phosphate depletion. However, by the time this problem is recognized, the serum phosphate often has been raised by the large amounts of cellular phosphate released from damaged muscle.

Reversible respiratory failure due to respiratory muscle weakness may preclude successful weaning from ventilatory support. Left ventricular dysfunction, heart failure, and ventricular arrhythmias may result from profound hypophosphatemia but may not be significant if serum phosphate is greater than 1.5 mg/dL. In one study, correction of moderate hypophosphatemia (phosphate level of <2 mg/dL) in patients with septic shock led to a significant increase in blood pressure as well as improvement of left ventricular function and arterial pH. Hematologic sequelae of severe hypophosphatemia include hemolyis, platelet dysfunction with bleeding, and impaired leukocyte function (phagocytosis and killing). Erythrocytes demonstrate increased fragility, altered membrane composition, rigidity, and microspherocytosis; and reduced levels of ATP and 2,3-diphosphoglycerate (2,3-DPG). The reduction in erythrocyte 2,3-DPG impairs oxyhemoglobin dissociation, thereby potentially reducing oxygen delivery to tissues. This problem, together with accelerated hemolysis, may provoke a substantial increase in cardiac output. The blockade in cellular glycolysis becomes demonstrable at levels of serum phosphate between 1 and 2 mg/dL. Glucose intolerance and insulin resistance also have been demonstrable in affected patients.

Treatment

Hypophosphatemia appears most often in acutely or critically ill persons. Accordingly, it is often difficult to discern whether hypophosphatemia is responsible for features of the multiple organ dysfunction commonly encountered in this population. For example, although depression of intracellular high-energy organophosphates has been demonstrated during treatment of diabetic ketoacidosis, and phosphate repletion leads to more rapid recovery of erythrocyte 2,3-DPG concentrations, opinion is divided as to whether phosphate therapy in this setting hastens recovery, prevents complications, or reduces mortality. Nevertheless, because severe hypophosphatemia has been associated in a variety of clinical settings, with serious neuromuscular, cardiovascular, and hematologic dysfunction that is at least partially reversible with phosphate repletion, most authorities now agree that a relatively low threshold for treatment should be adopted.

The decision to correct hypophosphatemia urgently should be guided by the estimated severity of the cellular phosphate deficit, the presence of signs or symptoms suggestive of phosphate depletion, and the overall clinical status of the patient. The presence of renal insufficiency (a risk for iatrogenic hyperphosphatemia), concomitant administration of intravenous glucose (alone or as a component of hyperalimentation solutions), and the potential for aggravating coexistent hypocalcemia also should be considered.

Limited data are available from clinical trials to predict the appropriate dose and rate of phosphate administration. In patients without severe renal insufficiency or hypocalcemia, administration of intravenous phosphate at rates of 2 to 8 mmol/hour of elemental phosphate over 4 to 8 hours frequently corrects hypophosphatemia without provoking hyperphosphatemia or hypocalcemia. Suggested guidelines based on serum phosphate concentration are shown in Table 26-9. It is essential that serum calcium and phosphate be monitored every 6 to 12 hours during and after phosphate therapy, both to detect untoward consequences and because many patients require additional infusions for recurrent hypophosphatemia within 24 to 48 hours of apparently successful repletion. Less acute or severe hypophosphatemia should be managed with oral (or enteral) phosphate supplements if possible, generally given as a total of 1.0 to 2.0 g/day (as elemental...
phosphate) of neutral sodium or potassium phosphate in divided doses three to four times a day (see Table 26-6). In many patients, however,

**TABLE 26-9 -- Urgent Therapy of Hypophosphatemia**

<table>
<thead>
<tr>
<th>Consider</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Severity of hypophosphatemia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Likelihood of underlying phosphate depletion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical condition of the patient</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Renal function</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serum calcium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concurrent parenteral therapy (glucose, hyperalimentation)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Guidelines**

<table>
<thead>
<tr>
<th>Serum Phosphate (mg/dL)</th>
<th>Rate of Infusion (mmol/hr)</th>
<th>Duration (hr)</th>
<th>Total Phosphate (mmol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;2.5</td>
<td>2.0</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>&lt;1.5</td>
<td>4.0</td>
<td>6</td>
<td>24</td>
</tr>
<tr>
<td>&lt;1.0</td>
<td>8.0</td>
<td>6</td>
<td>48</td>
</tr>
</tbody>
</table>

*Rates shown are normalized for a 70-kg person. Most formulations available in the United States provide 3 mMol/mL of sodium or potassium phosphate.*

oral phosphate therapy is limited by gastrointestinal symptoms such as nausea or diarrhea.
DISORDERS OF MAGNESIUM METABOLISM

The fourth most abundant extracellular cation, magnesium, like calcium, plays a critical physiologic role, particularly in neuromuscular function. The importance of intracellular magnesium in energy metabolism, as a cofactor for ATP and a wide variety of enzymes and transporters, is reflected in the fairly global clinical effects that accompany disorders of magnesium homeostasis. Hypomagnesemia and hypermagnesemia are among the most common electrolyte disturbances; one or the other of these abnormalities is observed in as many as 20% of hospitalized patients and even more frequently (i.e., in 30% to 40%) among those admitted to intensive care units.

Hypermagnesemia

Magnesium homeostasis is achieved mainly through highly efficient regulation of tubular magnesium reabsorption in the loop of Henle. As normal kidneys can readily excrete even large amounts of magnesium (i.e., 500 mEq/day), high filtered loads of magnesium rarely cause hypermagnesemia except in patients with severe acute or chronic renal failure. Increased magnesium loads in such cases may arise from ingestion of large amounts of oral magnesium salts, typically given as cathartics or antacids, or from extensive soft tissue ischemia or necrosis in patients with trauma, sepsis, cardiopulmonary arrest, burns, or shock (Table 26-10). Hypermagnesemia may result from parenteral administration of magnesium salts, as when magnesium is used to treat preeclampsia. The infants of such hypermagnesemic mothers may manifest transient hypermagnesemia as well, along with parathyroid suppression and neurobehavioral symptoms. The use of oral magnesium preparations as laxatives may lead to hypermagnesemia if absorption is increased by intestinal ileus, obstruction, or perforation.

The most prominent clinical manifestations of hypermagnesemia are vasodilatation and neuromuscular blockade, which may involve both pre- and postsynaptic inhibition of neuromuscular transmission. Signs and symptoms generally do not appear unless the serum magnesium concentration exceeds 4 mEq/L.

<table>
<thead>
<tr>
<th>TABLE 26-10 – Causes of Hypermagnesemia</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Excessive Magnesium Intake</strong></td>
</tr>
<tr>
<td>Cathartics, antacids, enemas</td>
</tr>
<tr>
<td>Dead Sea drowning</td>
</tr>
<tr>
<td>Parenteral magnesium administration</td>
</tr>
<tr>
<td>Magnesium-rich urologic irrigants</td>
</tr>
<tr>
<td>Intestinal obstruction or perforation following magnesium ingestion</td>
</tr>
<tr>
<td><strong>Rapid Mobilization from Soft Tissues</strong></td>
</tr>
<tr>
<td>Trauma</td>
</tr>
<tr>
<td>Shock, sepsis</td>
</tr>
<tr>
<td>Cardiac arrest</td>
</tr>
<tr>
<td>Burns</td>
</tr>
<tr>
<td><strong>Impaired Magnesium Excretion</strong></td>
</tr>
<tr>
<td>Renal failure</td>
</tr>
<tr>
<td>Familial hypocalciuric hypercalcemia</td>
</tr>
<tr>
<td><strong>Other</strong></td>
</tr>
<tr>
<td>Adrenal insufficiency</td>
</tr>
<tr>
<td>Hypothyroidism</td>
</tr>
<tr>
<td>Hypothermia</td>
</tr>
</tbody>
</table>

4 mEq/L Hypotension, often refractory to pressors and volume expansion, may be one of the earliest signs of progressive hypermagnesemia. Lethargy, nausea, and weakness, accompanied by reduction in or loss of deep tendon reflexes, may progress to stupor or coma with respiratory insufficiency or quadriparesis at serum concentrations in excess of 8 to 10 mEq/L. Gastrointestinal hypomotility or ileus is common. Facial flushing and pupillary dilatation may be observed. Hypotension may be complicated by a paradoxical relative bradycardia, and other cardiac effects may be evident, including prolongation of the PR, QRS, and QTc intervals and appearance of heart block and, ultimately, asystole as serum concentrations approach 20 mEq/L.

Hypermagnesemia also causes hypocalcemia and increased urinary calcium excretion, the result of both a direct suppression ofPTH secretion and a PTH-independent inhibition of renal tubular calcium reabsorption. Severe hypocalcemia opposes the effect of hypermagnesemia on PTH secretion, so that serum PTH typically remains within the normal range but is still inappropriate for the serum calcium level.

Successful treatment of hypermagnesemia requires identification and interruption of the source of magnesium, together with measures to increase clearance of magnesium from the extracellular fluid. Use of magnesium-free cathartics or enemas to accelerate clearance of ingested magnesium from the gastrointestinal tract, together with vigorous intravenous hydration, generally has been successful in reversing hypermagnesemia. Refractory cases, especially those in patients with advanced renal insufficiency, may require hemodialysis. Intravenous calcium (100 to 200 mg) infusions have been advocated as an effective antidote to hypermagnesemia, and there are examples in which this approach has apparently been successful, at least temporarily.
Hypomagnesemia

Hypomagnesemia may occur because of impaired intestinal absorption of magnesium or defective renal tubular reabsorption of magnesium, or a combination of these (Table 26-11). Because only 1% of the body's magnesium content is present in extracellular fluid, measurements of serum magnesium concentration typically do not adequately reflect total-body magnesium or the magnesium status of the intracellular compartment in critical tissues such as muscle. Thus, patients with deficiency of tissue magnesium may fail to manifest overt hypomagnesemia but will exhibit abnormal retention (i.e., >50% in 24 hours) of infused magnesium, a maneuver that may be employed to assess magnesium status.

Etiology

Intestinal Causes of Hypomagnesemia

Selective dietary magnesium deficiency does not occur, and it is remarkably difficult, in fact, to induce magnesium depletion experimentally by feeding magnesium-deficient diets, probably because renal magnesium conservation is so efficient. Large amounts of magnesium may be lost in chronic diarrheal states (this fluid may contain more than 10 mEq of magnesium per liter), or via intestinal fistulae or prolonged gastrointestinal drainage. More commonly, magnesium becomes trapped within fatty acid "soaps" in disorders associated with chronic malabsorption. In a rare but informative genetic syndrome

<table>
<thead>
<tr>
<th>Impaired Intestinal Magnesium Absorption</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary infantile hypomagnesemia</td>
</tr>
<tr>
<td>Malabsorption syndromes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Increased Intestinal Magnesium Losses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protracted vomiting or diarrhea</td>
</tr>
<tr>
<td>Intestinal drainage</td>
</tr>
<tr>
<td>Intestinal fistulas</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Impaired Renal Tubular Magnesium Reabsorption</th>
</tr>
</thead>
<tbody>
<tr>
<td>Congenital magnesium-wasting syndromes</td>
</tr>
<tr>
<td>Bartter's syndrome</td>
</tr>
<tr>
<td>Gitelman's syndrome</td>
</tr>
<tr>
<td>Magnesuria with nephrocalcinosis</td>
</tr>
<tr>
<td>Acquired renal disease</td>
</tr>
<tr>
<td>Tubulointerstitial disease</td>
</tr>
<tr>
<td>Postobstruction, acute tubular necrosis (diuretic phase)</td>
</tr>
<tr>
<td>Renal transplantation</td>
</tr>
<tr>
<td>Drugs and toxins</td>
</tr>
<tr>
<td>Ethanol</td>
</tr>
<tr>
<td>Diuretics (loop, thiazide, osmotic)</td>
</tr>
<tr>
<td>Cisplatin</td>
</tr>
<tr>
<td>Pentamidine</td>
</tr>
<tr>
<td>Cyclosporine</td>
</tr>
<tr>
<td>Aminoglycosides</td>
</tr>
<tr>
<td>Foscarnet</td>
</tr>
<tr>
<td>Amphotericin B</td>
</tr>
<tr>
<td>Endocrine and metabolic abnormalities</td>
</tr>
<tr>
<td>Extracellular fluid volume expansion</td>
</tr>
<tr>
<td>Hyperaldosteronism (primary, secondary)</td>
</tr>
<tr>
<td>Inappropriate antidiuretic hormone secretion</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
</tr>
<tr>
<td>Hypercalcinia</td>
</tr>
<tr>
<td>Phosphate depletion</td>
</tr>
<tr>
<td>Metabolic acidosis</td>
</tr>
<tr>
<td>Hyperthyroidism</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rapid Shifts of Magnesium out of Extracellular Fluid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intracellular redistribution</td>
</tr>
<tr>
<td>Recovery from diabetic ketoacidosis</td>
</tr>
<tr>
<td>Refeeding syndrome</td>
</tr>
<tr>
<td>Correction of respiratory acidosis</td>
</tr>
<tr>
<td>Catecholamines</td>
</tr>
<tr>
<td>Accelerated net bone formation</td>
</tr>
<tr>
<td>Following parathyroidectomy</td>
</tr>
<tr>
<td>Osteoblastic metastases</td>
</tr>
<tr>
<td>Treatment of vitamin D deficiency</td>
</tr>
<tr>
<td>Calcitonin therapy</td>
</tr>
<tr>
<td>Other losses</td>
</tr>
</tbody>
</table>
termed primary hypomagnesemia, a defect in the saturable component of intestinal magnesium absorption causes hypomagnesemia that can be partially overcome by administering large amounts of oral magnesium. 

Renal Causes of Hypomagnesemia

Renal magnesium wasting may result from a primary tubular transport defect, as occurs in Bartter's syndrome and a number of other rare inherited magnesium-wasting renal tubular disorders (see Table 26-11). Most often, however, it is attributable to an acquired abnormality in tubular magnesium reabsorption. In normal persons, magnesium reabsorption is virtually complete within several days of instituting experimental dietary magnesium deficiency, even before serum magnesium has declined substantially. Thus, the finding of more than 1 mEq/day of urinary magnesium in a frankly hypomagnesemic patient indicates a defect in renal tubular magnesium reabsorption. Acquired primary renal tubular magnesium wasting occurs in various tubulointerstitial disorders, recovery from acute tubular necrosis or obstruction, renal transplantation, various endocrinopathies, alcoholism, and exposure to certain drugs (see Table 26-11).

Hypomagnesemia or magnesium depletion due to subtotal renal reabsorption may complicate a variety of endocrinopathies, including hyperaldosteronism, hyperthyroidism, and disorders associated with hypercalcemia, hypercalciuria, or phosphate depletion. In primary hyperparathyroidism, PTH stimulates increased tubular magnesium reabsorption, but this increase is opposed by a direct tubular effect of hypercalcemia. As a result, the serum magnesium level in primary hyperparathyroidism generally is normal or only slightly reduced. In hypoparathyroidism, serum and urinary magnesium levels are low. The magnesium depletion in hypoparathyroidism is consistent with loss of both the magnesium-retaining action of PTH and the stimulatory effect of 1,25(OH)\(_2\)D on intestinal magnesium absorption.

Diabetes is among the most common medical disorders associated with hypomagnesemia. The severity of the hypomagnesemia in diabetics correlates with indices of glycosuria and poor glycemic control, which suggests that urinary losses of magnesium on the basis of glycosuria may partly explain the magnesium depletion. Rapid correction of hypoglycemia with insulin therapy causes magnesium to enter cells and may further lower the extracellular magnesium concentration during treatment.

Alcoholism is another very common clinical setting in which hypomagnesemia occurs. Magnesium depletion in alcoholism may result in part from nutritional deficiency of magnesium, overall caloric starvation and ketosis, and gastrointestinal losses due to vomiting or diarrhea, but an acute magnesium effect of alcohol ingestion probably plays the major role. This effect of alcohol is most evident when blood alcohol levels are rising and may be related to transient suppression of PTH secretion. Other factors that may contribute to hypomagnesemia in alcoholism include pancreatitis, malabsorption, secondary hyperaldosteronism, respiratory alkalosis, and elevation in plasma catecholamines, which increase intracellular sequestration of magnesium.

Numerous drugs have been identified as causes of defective renal tubular magnesium reabsorption and hypomagnesemia. These agents include diuretics of all classes (especially loop diuretics), cisplatin, pentamidine, cyclosporine, aminoglycosides, foscarnet, and amphotericin. Most often, drug-induced hypomagnesemia is mild and reversible, particularly that associated with diuretic therapy. In over half of patients undergoing cisplatin therapy, hypomagnesemia is noted within days or weeks and roughly half of those who develop the abnormality exhibit persistent hypomagnesemia many months or even years later. The median duration of hypomagnesemia in cisplatin-treated patients is about 2 months, but recovery has been observed up to 2 years after treatment. Cisplatin may induce a more generalized nephropathy and azotemic renal failure, but the magnesium wasting appears to be an isolated functional abnormality. There is some evidence that the renal magnesium-wasting syndrome can be prevented by intravenous magnesium administration (24 to 40 mEq) before or during cisplatin infusion. Such findings suggest that cisplatin may selectively impair magnesium reabsorption by binding competitively to sites or cells involved in binding and transport of magnesium.

A syndrome very similar to that seen with cisplatin therapy is observed frequently in transplant recipients who receive cyclosporin A. Cisplatin may selectively impair magnesium reabsorption by binding competitively to sites or cells involved in binding and transport of magnesium. This abnormality appears to be due mainly to impaired renal 1-hydroxylation of 25(OH)D, although tissue resistance to 1,25(OH)\(_2\)D may also play a role. The serum 1,25(OH)\(_2\)D concentration usually is low during hypomagnesemia, which may result from magnesium depletion per se, parathyroid insufficiency, or coexistent vitamin D deficiency. Deficiency of 1,25(OH)\(_2\)D is probably not the main cause of hypocalcemia in these patients, however,
because hypocalcemia can be rapidly corrected (within hours to days) by magnesium therapy alone, well in advance of any increase in the serum 1,25(OH)₂D₃ concentration.

Therapy of Hypomagnesemia

Mild, asymptomatic hypomagnesemia may be treated with oral magnesium salts (MgCl₂, MgO, or Mg(OH)₂) usually given in divided doses totaling 40 to 60 mEq (480 to 720 mg) per day (see Table 26-6). Diarrhea sometimes occurs with larger doses but generally is not a problem. The gluconate form (supplying 58 mg of magnesium per gram) is said to cause less diarrhea. Patients with malabsorption or ongoing urinary magnesium losses may require chronic oral therapy to avoid recurrent magnesium depletion. Although intestinal magnesium absorption is severely impaired in renal failure, oral magnesium must be administered with great caution in this setting, especially in patients receiving concomitant therapy with 1,25(OH)₂D₃.

Symptomatic or severe hypomagnesemia (magnesium concentration <1 mEq/L), especially if complicated by hypocalcemia, usually signifies magnesium deficits of at least 1 to 2 mEq/kg and is best treated promptly with parenteral magnesium salts. The use of intramuscular magnesium sulfate (MgSO₄) is to be discouraged, as the injections are painful and provide relatively little magnesium (2 mL of 50% MgSO₄ supplies only 8 mEq of magnesium, as compared with typical magnesium deficits in excess of 100 mEq). Moreover, because unretracted sulfate ions also may increase urinary calcium excretion, administration of intravenous magnesium chloride or gluconate probably is the most logical approach to initial parenteral therapy for patients who also may be hypocalcemic.

In adult hypomagnesemic patients with normal renal function, rates of infusion of 2 to 4 mEq/hour (i.e., 50 to 100 mEq/day) generally are needed to maintain serum magnesium concentrations in the range of 2 to 3 mEq/L. Up to 100 mEq/day for 2 days can be safely administered without elevating serum magnesium concentration above 4 mEq/L, whereas doses of 200 mEq/day may increase serum magnesium to 4.5 to 5.5 mEq/L and thus are excessive. In patients with active seizures or other urgent indications, the infusion may be preceded by a slowly administered bolus of 10 to 20 mEq, followed by a higher rate of infusion (i.e., 10 to 15 mEq/hour) for the first 1 or 2 hours only. Patients with normal renal function can readily excrete over 400 mEq/day of magnesium in the urine without becoming hypermagnesemic, but even mild renal failure may greatly limit magnesium excretion. Therefore, doses of magnesium supplements should be reduced two- to threefold and careful serial monitoring of serum magnesium performed in patients with compromised renal function.

It is important to appreciate that a large fraction of parenterally administered magnesium may be excreted in the urine, even in patients with profound magnesium deficiency. Many such patients excrete as much as 50% to 75% of infused magnesium; in normal subjects, this fraction approaches 100%. Moreover, because equilibration of the intracellular and extracellular magnesium pools is relatively slow, it is generally necessary to continue magnesium therapy for 3 to 5 days to achieve adequate repletion of the typical deficit of 1 to 2 mEq/kg. Because serum magnesium may become normal well before tissue stores are repleted, monitoring of urinary magnesium excretion is a more reliable measure of the approach to full repletion, especially after patients are switched to oral therapy.

The need for calcium, potassium, and phosphate supplementation should be considered in the usual clinical setting of hypomagnesemia. Vitamin D deficiency also frequently coexists and should be treated with oral or parenteral vitamin D or 25(OH)D. Use of 1,25(OH)₂D₃ is not necessary, does not hasten recovery, and may actually worsen hypomagnesemia by suppressing PTH secretion and thereby promoting renal magnesium excretion. Initial parenteral magnesium therapy in hypocalcemic patients may produce dramatic hypophosphatemia via the rapid stimulation of PTH secretion. This effect is most likely to be problematic in patients with underlying phosphate depletion (as in malabsorption, alcoholism, or diabetes), in whom it may provoke acute neuromuscular dysfunction, and may be avoided by concomitant intravenous calcium therapy.
References


37. Russell J, Shewen LM. The effects of 1,25(OH)\(_2\)D \(_3\) and high calcium on transcription of the pre-proparathyroid hormone gene are direct. Trans Assoc Am Physicians 1987; 100:256262.


108. Rubin DA, Jüppner H, Zabriskie express the common parathyroid hormone/parathyroid hormonerelated peptide receptor (PTH1R) and a novel receptor (PTH3R) that is preferentially activated by mammalian and fugu parathyroid hormonerelated peptide. J Biol Chem 1999; 274:88758879.


111. A. 483494.


118. Rubin DA, Jüppner H. Zebrafish express the common parathyroid hormone/parathyroid hormonerelated peptide receptor (PTH1R) and a novel receptor (PTH3R) that is preferentially activated by mammalian and fugu parathyroid hormonerelated peptide. J Biol Chem 1999; 274:88758879.


126. Hoare SRJ, Bonner TI, Usdin TB. Comparison of rat parathyroid hormone-related peptide 2 (PTHrP2) receptor activation: PTH is a low potency partial agonist at the rat PTH2 receptor. Endocrinology 1999; 140:44144426.


142. Marx UC, Ackermann K, Bayer P, et al. Solution structures of human parathyroid hormone fragments hPTH(134) and hPTH(139) and bovine parathyroid hormone fragment bPTH(137). Biochem Biophys Res Commun 2000; 267:213220.


References


References


309. de Boland AR, Namere I, Norman AW. Ca2+-channel agonist BAY K8444 mimics 1,25(OH)2-vitamin D3 rapid enhancement of Ca2+ transport in chick perfused duodenum. Biochem Biophys Res Commun 1990; 166:217222.


References


1366
23. Tambyah PA, Ong BKC, Lee KO. Reversible parkinsonism and asymptomatic hypocalcemia with basal ganglia calcification from hyperparathyroidism 26 years after thyroid surgery. Am J Med 1993; 94:4445.


References

1369


Reilkin R. Hypocalcemia resulting from calcium accretion by a chondrosarcoma. Cancer 1974; 34:18341837.


Reilkin R. Hypocalcemia resulting from calcium accretion by a chondrosarcoma. Cancer 1974; 34:18341837.


References


Chapter 27 - Metabolic Bone Disease

Lawrence G. Raisz
Barbara E. Kream
Joseph A. Lorenzo

STRUCTURE AND FUNCTION OF THE SKELETON

The skeleton is one of the largest organ systems in the body. It consists of a large mineralized matrix and a small but highly active cellular fraction. The skeleton serves the dual functions of maintaining the structure of the body and providing a storehouse of minerals and protein. Imbalance between the structural and the storage functions of the skeleton can be important in the pathogenesis of metabolic bone disease.

Embryology and Anatomy

In the embryo, skeletal development begins with the condensation of mesenchyme into cartilage. Bone formation then occurs through one of two distinct pathways. The growth of long bones and the vertebrae involves endochondral bone formation (Fig. 27-1). The cartilage growth plate cells proliferate and then undergo hypertrophy; this is followed by partial degradation of the matrix, which then mineralizes. The cartilage is invaded by vessels, and the spicules of mineralized cartilage are covered by osteoblasts to form a cancellous or trabecular bone mass (the primary spongiosa). These structures are then resorbed and replaced by trabecular plates made up entirely of bone, termed the secondary spongiosa. This bone undergoes rapid remodeling. In the adult, it is most abundant at the ends of the long bones and in the bodies of the vertebrae.

Intramembranous bone formation occurs adjacent to the cartilage template, typically in flat bones, such as the skull, scapula, and ileum, and on the outer surfaces of long bones. Initially, relatively disorganized woven bone is formed, but this rapidly converts to more organized lamellar bone produced by oriented layers of osteoblasts. The main difference between endochondral and intramembranous bone formation is that the latter does not use calcified cartilage as a direct template for osteoblasts.

Cortical bone is the dense bone that is found, for example, in the shafts of long bones. It makes up 80% of the mass of the skeleton, determines its shape, and provides much of its strength. Modeling of cortical bone involves not only new bone formation but also resorption to alter skeletal shape (see Fig. 27-1). The wide cortex at the epiphyseal plate of long bones must be resorbed because these bones elongate to maintain the narrow tubular structure of the diaphysis.

Although remodeling begins early in cancellous bone, remodeling of cortical bone begins only after the cortex reaches a critical thickness. In smaller animals such as rodents, cortical bone can remain lamellar. In large animals and humans, lamellar cortical bone is gradually replaced through haversian remodeling to form cylindrical osteons.
Bone mineral consists of fibers of type I collagen laid down in layers that have various orientations, which may add to the strength of the matrix. The matrix contains many additional proteins, including small amounts of other collagen types that may be important in the interaction of type I collagen with noncollagen proteins. The noncollagen proteins represent about 10% of the total protein in bone and may serve to direct the formation of fibers, enhance mineralization, strengthen the attachment between cells and matrix, and provide signals for bone remodeling (Table 27-1).

The protein composition of matrix may vary particularly between woven and lamellar bone. These proteins range from the large cell-attachment proteins (e.g., thrombospondin and fibronectin) with molecular masses higher than 400 kd, to the small, vitamin K-dependent gamma-carboxylated proteins, matrix Gla protein and osteocalcin, which are 6-kd calcium-binding proteins. A number of the noncollagen proteins (e.g., biglycan, decorin, bone sialoprotein, osteopontin, and osteoadherin) are highly acidic. In addition to cell-attachment functions of matrix proteins, many noncollagen proteins represent about 10% of the total protein in bone and may serve to direct the formation of fibers, enhance mineralization, strengthen the attachment between cells and matrix, and provide signals for bone remodeling.

### Table 27-1 -- Noncollagen Proteins of Bone

<table>
<thead>
<tr>
<th>Name</th>
<th>Possible Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Osteocalcin</td>
<td>May regulate bone mineral maturation; negative regulator of bone formation and osteoblast function</td>
</tr>
<tr>
<td>Matrix Gla protein</td>
<td>Inhibits mineralization</td>
</tr>
<tr>
<td>Thrombospondins</td>
<td>Cell attachment; binds to heparin, type I collagen, thrombin, laminin</td>
</tr>
<tr>
<td>Vitronectin</td>
<td>Cell attachment; binds to collagen and heparin</td>
</tr>
<tr>
<td>Fibronectin</td>
<td>May regulate elastic fiber formation</td>
</tr>
<tr>
<td>Osteopontin</td>
<td>Binds to cells; mediates effect of mechanical stress on osteoclasts and osteoblasts</td>
</tr>
<tr>
<td>Bone sialoprotein</td>
<td>Binds cells, binds calcium; may initiate mineralization</td>
</tr>
<tr>
<td>Alkaline phosphatase</td>
<td>Hydrolyzes mineralization inhibitors</td>
</tr>
<tr>
<td>Osteonectin</td>
<td>Binds to growth factors; may regulate mineralization</td>
</tr>
<tr>
<td>Tetracitin</td>
<td>May regulate mineralization</td>
</tr>
<tr>
<td>Biglycan</td>
<td>May bind to collagen and TGF-</td>
</tr>
<tr>
<td>Decorin</td>
<td>Binds to collagen and may regulate fibril diameter; binds to TGF-</td>
</tr>
<tr>
<td>Fibromodulin</td>
<td>Binds to collagen and may regulate fibril diameter; binds to TGF-</td>
</tr>
<tr>
<td>Osteoadherin</td>
<td>Promotes integrin-mediated cell attachment</td>
</tr>
<tr>
<td>Hyaluronan</td>
<td>Synthesized during early bone formation; may capture space destined to become bone</td>
</tr>
</tbody>
</table>


sequences, these proteins contain varying amounts of carbohydrate and may be termed glycoproteins or proteoglycans. Noncollagen proteins of bone are often highly phosphorylated, which enables them to bind calcium, and may regulate mineralization. The matrix is also a storehouse for growth factors and their binding proteins (see later).

### Collagen Synthesis

Type I collagen, a rigid, rodlike, insoluble molecule composed of two 1 chains and one 2 chain, consists of repeating triplets of amino acids with glycine in every third position and a high content of proline and lysine residues, which requires ascorbic acid. Collagen is initially synthesized as a soluble propeptide with large nonhelical extensions at both the COOH and NH₂ termini.プロコラーゲン also contains COOH-terminal interchain disulfide bonds that help initiate triple-helix formation. Procollagen is released into the cisternae of the rough endoplasmic reticulum, packaged in the Golgi vesicles, and secreted extracellularly. The procollagen peptides are then removed by specific peptidases.

### Chemistry of Matrix and Mineral

Bone mineral is made up of small, imperfect hydroxyapatite crystals, which contain carbonate, magnesium, sodium, and potassium. When bone forms in the presence of fluoride ions, the fluorapatite crystal is larger and less soluble than hydroxyapatite and may increase bone fragility.

---

**Figure 27-2** Synthesis and assembly of collagen fibrils. A, Intracellular post-translational modifications of pro alpha chains, association of propeptide domains, and folding into triple-helical conformation. Gal, galactose; Glc, glucose; Glc Nac, N-acetylglucosamine; (Man)n, mannosae. 8, Enzymatic cleavage of propeptide to collagen, self-assembly of collagen monomers into fibrils, and cross-linking of fibrils. (Modified from Prockop DJ, Kivirikko K. Heritable diseases of collagen. N Engl J Med 1984; 311: 377.)

To produce mature insoluble collagen molecules, which are further stabilized by intramolecular and intermolecular cross-links. The major cross-links are formed by lysine and hydroxylysine residues that ultimately form pyridinium ring structures (see later).

**Mineralization**

Bone mineral is made up of small, imperfect hydroxyapatite crystals, which contain carbonate, magnesium, sodium, and potassium. When bone forms in the presence of fluoride ions, the fluorapatite crystal is larger and less soluble than hydroxyapatite and may increase bone fragility.
Mineralization occurs by two distinct mechanisms. The initial mineralization of calcified cartilage and woven bone probably occurs by means of matrix vesicles. These membrane-bound bodies are released from chondrocytes and osteoblasts, contain alkaline phosphatase, and can form a nidus for crystallization. In contrast, in lamellar bone, the collagen fibers are tightly packed and matrix vesicles are rarely seen. Mineralization does not occur immediately after collagen deposition, and there is a layer of 10 to 100 µm of unmineralized osteoid between the mineralization front and the osteoblast. Perhaps changes in the packing of the fibrils and in the composition of the noncollagen proteins are required for mineralization. Mineralization of collagen fibrils begins in the hole zones, where there is more room for inorganic ions to accumulate (Fig. 27-3). Mineralization requires calcium, phosphate, and alkaline phosphatase. This process is impaired in vitamin D deficiency and hypophosphatasia.
Osteoblast Differentiation and Function

Bone is formed by osteoblasts, \(^1\) which are highly differentiated cells that have many unique features (Fig. 27-4). They are initially derived from mesenchymal stem cells, which are termed inducible osteoprogenitors. \(^1\) Adult bone also contains less pluripotent bone cell precursors, or committed osteoprogenitors.

Osteoprogenitor cells (or preosteoblasts) replicate and differentiate into active osteoblasts that may exhibit varying phenotypic characteristics. \(^1\) For example, osteoblasts in early development and during repair produce woven bone, whereas more mature osteoblasts produce lamellar bone.

The distribution of noncollagen proteins in cortical and trabecular bone may also differ. The activity of osteoblasts can vary during bone formation. Some cells are tall and closely packed and produce a large amount of matrix in a small area; others are flatter and produce matrix at a slower rate over a larger area. Nevertheless, all differentiated osteoblasts share certain features. They are connected by gap junctions and contain a dense network of rough endoplasmic reticulum and a large Golgi complex. They secrete collagen and noncollagen proteins in an oriented fashion and produce more type I collagen and alkaline phosphatase than other mesenchymal cells. Some products, such as osteocalcin, are produced almost uniquely by osteoblasts; others, such as osteopontin, are produced by a limited number of cell types.

Mature osteoblasts have a finite capacity to produce matrix, and bone formation is sustained by the arrival of new populations of cells at the bone surface. After osteoblasts have completed their matrix formation, they may become embedded in the matrix as osteocytes, may undergo apoptotic cell death, or may be converted to flattened lining cells, which cover a large percentage of the surface of bone with a thin cytoplasmic layer.

The conversion of osteoblasts to osteocytes involves a reduction of, but not a complete loss of, metabolic activity. \(^1\) A critical feature is the development of an extensive network of cytoplasmic connections. The osteoblasts have multiple cell processes that are connected to underlying osteocytes through small canaliculi. After mineralization is complete, the processes persist as connections between osteocytes (Fig. 27-5). This extended syncytium is probably important in maintaining the viability of the osteocytes, which otherwise would be separated from the extracellular fluid.

Initially, osteocytes may continue to synthesize collagen and play a role in mineralization. Later, the major role of the osteocyte-osteoblast syncytium may be to sense mechanical forces. \(^1\) One hypothesis is that small strains produce fluid shear stress in the canaliculi between osteocytes. This effect may result in intracellular signaling through changes in ion channels or in the production of biologically active molecules (e.g., prostaglandins) from the phospholipid membrane.

Cells of the osteoblastic lineage are important not only in forming bone but also in initiating resorption. \(^1\) It is still not clear which cells of the osteoblast lineage activate resorption. However, the nature of the factors that osteoblast-lineage cells produce, which regulate osteoclast formation and activity, is now more clearly understood (see later). Most of the hormonal factors that stimulate bone resorption act on osteoblasts or their precursor cells. Osteoblast-lineage cells also produce factors that regulate resorption, including cytokines, prostaglandins, and growth factors. In cell culture, contact between osteoblastic cells and hematopoietic cells appears to be necessary for osteoclast formation (Fig. 27-6). Osteoblasts may also play a role in initiating bone resorption by releasing enzymes that prepare the bone surface. Hormones and local factors that stimulate bone resorption increase the production of collagenase, other metalloproteinases, and plasminogen activator by osteoblasts. These enzymes may remove the surface proteins of bone, which block the access of osteoclasts to the mineralized matrix. Osteoblasts may also influence the development and maintenance of the marrow because they are sources of growth factors, cytokines, and chemokines that may act on hematopoietic cells.
Osteoclast Differentiation and Function

Although osteoclasts are derived from hematopoietic progenitors, their relation to other hematopoietic lineages is not fully defined. Hematopoietic stem cells under the direction of granulocyte-macrophage colony-stimulating factor (GM-CSF) and colony-stimulating factor 1 (macrophage colony-stimulating factor [M-CSF]) may differentiate into either monocyte-macrophages or preosteoclasts. The latter fuse to form highly differentiated multinucleated osteoclasts that resorb bone (see Fig. 27-6). Progression through the osteoclast pathway probably involves multiple local and systemic hormones that may include 1,25-dihydroxyvitamin D (1,25(OH)₂ D), prostaglandins, and the cytokines interleukin-1 (IL-1), IL-6, and tumor necrosis factor (TNF).

The nature of the osteoblast-lineage cell products, which directly regulate osteoclast formation and function, has been clarified. The principal Stimulator of osteoclast formation is a member of the TNF protein family, termed receptor activator of nuclear factor B ligand (RANKL). This protein was originally identified as a product of activated T lymphocytes but is also recognized as a critical Stimulator of osteoclastogenesis.

Production of RANKL in osteoblast-lineage cells is stimulated by essentially all agents that enhance osteoclast formation, including parathyroid hormone (PTH), 1,25(OH)₂ D, prostaglandins, and many cytokines. Mice deficient in RANKL do not form osteoclasts and have osteopetrosis. In contrast, injection of RANKL into mice stimulates osteoclast formation and bone resorption. RANKL is produced as a membrane protein. In activated T lymphocytes, however, RANKL is cleaved from the cell membrane and is released as a soluble factor. It is unclear whether similar events occur in osteoblast-lineage cells, although there is some evidence that cleavage and release of soluble RANKL occur in malignant cells that metastasize to bone.

In addition to RANKL, osteoblast-lineage cells produce an inhibitor of osteoclastogenesis, called osteoprotegerin (OPG). OPG is a soluble receptor for RANKL that binds this ligand and prevents interaction of RANKL with its bioactive receptor RANK. OPG is produced widely. In bone marrow cultures, a number of stimulators of resorption, including PTH, 1,25(OH)₂ D, and prostaglandin E₂ (PGE₂), inhibit OPG production. Hence, for these factors there is a reciprocal relationship between RANKL stimulation and OPG inhibition that causes activation of osteoclastogenesis and enhanced resorption. Mice deficient in OPG have osteoporosis, whereas mice that overexpress OPG have increased bone mass. These results, together with those for RANKL-deficient mice or mice injected with RANKL, demonstrate that osteoclast-mediated bone resorption is tightly regulated by the combined actions of RANKL and OPG.

The active receptor for RANK is RANK, a member of the TNF receptor family. Osteoclasts and their immediate precursor cells express RANK. Binding of RANKL to RANK activates a series of intracellular pathways that activate nuclear factor B and an N-terminal kinase (JNK). The TNF receptor-associated factors (TRAFs) bind RANK intracellularly and are involved in RANK responses. Mice deficient in TRAF-6, like those deficient in RANK, develop osteopetrosis. In addition to its effects on bone, the RANK-RANKL system is involved in lymphocyte function as well as breast and lymph node development. Mature osteoclasts express RANK, and treatment of these cells with RANKL inhibits programmed cell death (apoptosis) and stimulates resorptive activity.

In addition to RANKL, M-CSF is essential for osteoclast formation. Mice deficient in M-CSF have osteopetrosis and few osteoclasts. In cultures of isolated osteoclast precursor cells, both M-CSF and RANKL must be present for mature osteoclasts to form. M-CSF enhances RANK production in osteoclast precursors and inhibits apoptosis of both osteoclast precursors and mature osteoclasts. Binding of c-fms by M-CSF activates tyrosine kinase activity in the receptor, which initiates a series of intracellular events.

The formation of multinucleated osteoclast-like cells in vitro requires both hematopoietic precursors and cells of the osteoblastic lineage. In vivo or in cultures with devitalized bone, mononuclear preosteoclasts attach to the bone surface and form multinucleated osteoclasts by fusion. The accumulation of additional nuclei into osteoclasts by fusion probably continues while the cell is actively resorbing. The life span of the osteoclast is limited. As osteoclasts become inactive, they die by apoptosis. Hormones that enhance bone resorption may delay apoptosis, and inhibitors of resorption can accelerate it. The mechanisms that limit the extent of osteoclastic resorption are incompletely understood and may involve inhibition by calcium ions, which accumulate under the osteoclast resorbing surface, or by local inhibitory factors, such as transforming growth factor (TGF-β), which are released and activated during resorption.

The mature osteoclast is a unique and highly specialized cell. It usually contains 10 to 20 nuclei, but giant osteoclasts with up to 100 nuclei can be seen in Paget's disease or giant cell tumors of bone. The large size of osteoclasts is probably essential for their resorptive function, which depends on their ability to isolate a region of the bone surface from the extracellular fluid and produce a local environment that can dissolve bone mineral and degrade matrix. The resorbing apparatus consists of a central ruffled border area, which secretes hydrogen ions and proteolytic enzymes, surrounded by a clear or sealing zone that anchors the cell to the bone surface by a ring of contractile proteins linked to integrin receptors. The osteoclast may secrete adhesion proteins, such as osteopontin and bone sialoprotein, that anchor the cell to bone by binding both to mineral and to cell-surface integrins.

Acidification of the ruffled border area requires that osteoclasts have proton pumps. These pumps are similar to the vascular proton pumps that acidify intracellular organelles, but in the osteoclast they are exteriorized to increase the extracellular hydrogen ion concentration. The hydrogen ions dissociate from carbonic acid, which is synthesized by carbonic anhydrase; the bicarbonate generated by this dissociation is removed from the cell by chloride-bicarbonate exchange. Ion pumps can transport the dissolved calcium from the bone surface through the cell to the extracellular fluid.

However, calcium can also reach the extracellular fluid directly if the sealing zone is disrupted. The proteolytic enzymes produced by the osteoclast include lysosomal enzymes and metalloproteinases. Lysosomal proteases can degrade collagen at the low pH present in the ruffled border area. Cathepsin K is probably the most important of these. Metalloproteinases, which are active at neutral pH, have also been detected at the resorption site. In trabecular bone, osteoclasts characteristically resorb to a limited depth and then move laterally to produce irregular, plate-like resorption areas that are termed Howship's lacunae. In cortical remodeling, the path of directed resorption is longer, possibly because of renewal of osteoclasts from hematopoietic cells brought to the site through the haversian canal.
BONE REMODELING AND ITS REGULATION

After peak bone mass has been achieved, the cellular activity in the skeleton is largely devoted to an orderly sequence of bone resorption and formation, termed remodeling. This process produces plate-like structures on trabecular surfaces and cylindrical structures (osteons) in cortical bone called basic multicellular or basic structural units. The remodeling cycle can be divided into four steps: (1) activation, (2) resorption, (3) reversal, and (4) formation (Fig. 27-8 and Fig. 27-9).

Similar sequences are seen in trabecular and cortical remodeling. In young adults, this cycle is tightly coupled and the amount of new bone formed by osteoblasts is equal to the amount that is resorbed by osteoclasts. Postmenopausal and age-related bone loss begins when resorption increases and formation no longer keeps pace. The activation of new remodeling sites may also increase with age.

Calcitonin inhibits bone resorption by acting directly on the osteoclast but appears to play a small role in the regulation of bone turnover in adults. Bone mass is not increased in patients with medullary thyroid carcinoma, who have an excess of calcitonin production, or in athyreotic patients receiving adequate thyroid hormone replacement, who have low calcitonin levels.

Calcium-Regulating Hormones

Parathyroid Hormone

PTH acts on bone to stimulate resorption but does not act on osteoclasts in the absence of cells of the osteoblastic lineage; moreover, PTH receptors are abundant on osteoblasts but not on osteoclasts. PTH acts on osteoblasts to cause cell contraction; to induce immediate-early response genes, including c-fos and the inducible form of prostaglandin G/H synthase (cyclooxygenase); and to increase the synthesis of local mediators, including insulin-like growth factor I (IGF-I) and IL-6. High concentrations of PTH in vitro inhibit expression of type I collagen, but intermittent administration of PTH in vivo or in vitro can stimulate bone formation. PTH induces production of RANKL by cells of the osteoblast lineage and thereby increases osteoclastogenesis and the activity of osteoclasts. In some settings, PTH increases proliferation of cells of the osteoblast lineage and decreases their death by apoptosis.

Vitamin D

The hormonal form of vitamin D1,25(OH)2D3 is necessary for intestinal calcium and phosphorus absorption and, therefore, for mineralization. This form of vitamin D also has effects on the skeleton, but its physiologic role in bone remodeling is not clear. By increasing RANKL production by osteoblasts, it is a potent stimulator of osteoclast formation in cell culture, and high concentrations increase osteocalcin synthesis by osteoblasts and inhibit collagen synthesis. Lower concentrations may increase bone formation but not to the extent seen with intermittent administration of PTH.

Calcitonin

Calcitonin inhibits bone resorption by acting directly on the osteoclast but appears to play a small role in the regulation of bone turnover in adults. Bone mass is not greatly altered in patients with medullary thyroid carcinoma, who have an excess of calcitonin production, or in athyreotic patients receiving adequate thyroid hormone replacement, who have low calcitonin levels. In fact, bone turnover is increased in patients with medullary thyroid carcinoma. Nevertheless, subtle alterations in calcitonin production or response may play a role in metabolic bone disease.
Other Systemic Hormones

Growth Hormone

Deficiency and excess of growth hormone have marked effects on skeletal growth. Growth hormone increases both circulating and local levels of IGF-I, which mediates the skeletal effects of growth hormone. Both exogenous growth hormone and IGF-I increase bone remodeling. Growth hormone also stimulates cartilage growth, probably through an increase in local IGF production and direct stimulation of cartilage cell proliferation. Whether systemic IGF plays a role in skeletal growth is not known, but low levels of growth hormone receptors are present in bone cells and administration of IGF-I together with its major binding protein, IGFBP-3, can increase skeletal growth.

Glucocorticoids

Glucocorticoids exert biphasic effects on bone formation and resorption. In vivo, glucocorticoids may increase bone resorption indirectly by diminishing calcium absorption and producing secondary hyperparathyroidism. Low levels of glucocorticoids increase osteoclastic activity in organ culture, whereas high levels inhibit it, perhaps by decreasing the production of cytokines and prostaglandins. Glucocorticoids enhance differentiation of osteoblast precursors in cell culture and cause a transient increase in collagen synthesis in bones in organ culture, possibly as a result of increased sensitivity to endogenous IGF-I. However, glucocorticoids act in the long term in vitro and in vivo to inhibit bone formation, to decrease osteoblast replication and differentiation, to decrease IGF-I synthesis, and to increase osteoblast apoptosis.

Thyroid Hormones

In children, hyperthyroidism is associated with increased skeletal growth and hypothyroidism results in decreased growth. Thyroid hormones are crucial for cartilage growth and differentiation and enhance the response to growth hormone. Thyroid hormones increase bone turnover, although their effects on bone formation are less clear. Increased resorption as a result of hyperthyroidism may result in a coupled increase in bone formation, but thyroid hormone may also directly stimulate bone cell replication.

Insulin

Normal skeletal growth depends on an adequate amount of insulin. Excess insulin production by the fetuses of mothers with uncontrolled diabetes results in excessive growth of the skeleton and other tissues, and undertreated diabetes mellitus impairs skeletal growth and mineralization. In vitro, insulin at physiologic concentrations selectively stimulates osteoblastic collagen synthesis by a pretranslational mechanism. Insulin can mimic the effects of IGF-I but only at supraphysiologic levels. Insulin does not appear to affect bone resorption.

Gonadal Hormones

Both estrogens and androgens are critical for skeletal development and maintenance. Bone cells contain estrogen and androgen receptors, but it is difficult to demonstrate direct effects of gonadal steroids on bone formation or resorption in cell and organ culture. Gonadal hormones are crucial for the pubertal growth spurt, and estrogen is necessary for epiphyseal closure. Deficiency of estrogen or androgen increases bone resorption in vivo, possibly by increasing local synthesis or sensitivity to cytokines, such as IL-1, TNF, or IL-6, and to prostaglandins (see later). Androgens can increase bone formation in vivo. The effect of estrogens on bone formation is less clear. The absolute rate of bone formation is increased in estrogen deficiency states, possibly because of an increase in the activation frequency and the number of remodeling sites in bone. However, the fact that estrogen deficiency causes bone loss implies a relative deficiency in bone formation. Estrogens may also stimulate osteoblasts directly.
Local Regulators

Characterization of local regulators produced within the bone itself represents a major advance in bone biology. These local factors can be synthesized by bone cells or by adjacent hematopoietic cells and can interact both with each other and with systemic hormones. They are critical in the repair of skeletal damage and in the response to mechanical forces.

Cytokines

IL-1, IL-1, TNF-, and TNF- are potent stimulators of bone resorption and inhibitors of bone formation and may mediate bone loss after estrogen withdrawal. IL-6 increases osteoclastogenesis in cell cultures and may mediate some of the resorbing activity of PTH. IL-6 is produced by osteoblasts, and its production is stimulated by PTH, PGE$_2$, and other factors that increase resorption. IL-11, another member of the IL-6 cytokine family, also stimulates resorption. As noted previously, colony-stimulating factors are probably important in the early stages of osteoclast formation.

IL-4 and IL-13 inhibit resorption and prostaglandin synthesis in bone cells, and leukemia inhibitory factor has biphasic effects on bone formation. IL-7 stimulates B lymphopoiesis, which may be involved in osteoclastogenesis. IL-10 is an inhibitor of osteoclastogenesis and bone resorption. IL-15 and IL-17 stimulate IL-6, whereas IL-18 is inhibitory through its ability to increase production of granulocyte-macrophage colony-stimulating factor.

Interferon inhibits resorption by inhibiting osteoclast responses to RANKL. In addition to direct effects, responses to cytokines can be blocked by inhibitors, such as the IL-1 receptor antagonist and the soluble TNF receptor, or they can be enhanced by activators such as the soluble IL-6 receptor.

Transforming Growth Factor and Epidermal Growth Factor

These peptides stimulate bone resorption through the same receptor and act by both prostaglandin-dependent and prostaglandin-independent pathways. TGF- and epidermal growth factor (EGF) are potent mitogens in bone that probably act on both mesenchymal and hematopoietic precursors. TGF- is produced by neoplasms and may play a role in the increased bone resorption that occurs in certain malignancies.

Prostaglandins

Prostaglandins are potent regulators of bone cell metabolism and are synthesized by many cell types in the skeleton. Prostaglandin production in bone is regulated by the effects of local and systemic hormones and mechanical forces on the inducible cyclooxygenase (COX-2). Increased prostaglandin production may contribute to the increase in bone resorption with immobilization, the increase in bone formation with impact loading, and the changes after estrogen withdrawal. Many of the hormones, cytokines, and growth factors that stimulate bone resorption also increase prostaglandin production.

Prostaglandins have biphasic effects on bone formation. Stimulation of bone formation is seen in vivo, and inhibition of collagen synthesis occurs in osteoblast cultures. Bone cells produce PGE$_2$, PGF$_2$, prostacyclin, and lipoxygenase products (e.g., leukotriene B$_4$), which may also stimulate bone resorption.
Growth Factors

A large number of growth factors have been identified in bone, and their production by bone cells has been established by measurements of messenger ribonucleic acid (mRNA) and protein. Bone cells also produce binding proteins for growth factors, which may regulate their activity and storage in bone.

Insulin-like Growth Factors

IGFs increase bone cell replication, matrix synthesis, and bone formation. Both IGF-I and IGF-II are synthesized by bone cells and stored in bone matrix. More IGF-II is stored in human bone, but IGF-I is a more potent stimulator of osteoblasts. Binding of IGF-I and IGF-II to matrix may be mediated by specific IGF-binding proteins. Five of the six known binding proteins have been identified in bone, and these both inhibit and enhance IGF responses. Because PTH and PGE₂ increase and glucocorticoids decrease skeletal IGF-I synthesis, IGFs may mediate the effects of these hormones on bone growth. IGF-I and its binding proteins may also stimulate osteoclast formation.

Transforming Growth Factor and Bone Morphogenetic Proteins

This family of peptides has many complex actions on bone metabolism. TGF-β acts to stimulate bone formation and to inhibit bone resorption. Because it is stored in bone matrix and released in an active form by bone-resorbing hormones, it may play a role in the reversal phase of bone remodeling when resorption stops and formation starts. The bone morphogenetic proteins (BMPs), like TGF-β, stimulate mitosis and the growth of bone cells and have similar effects on other mesenchymal cells, particularly cartilage. TGF-β and BMPs accelerate the healing of bone defects in vivo.

Fibroblast Growth Factors

Acidic and basic fibroblast growth factors (FGF-1 and FGF-2, respectively) are crucial for skeletal development. FGFs inhibit collagen synthesis in cell culture but increase bone formation in vivo. FGF can stimulate bone resorption by both prostaglandin-dependent and prostaglandin-independent pathways. FGFs are heparin-binding growth factors, and this binding can influence their cellular activity and their sequestration in extracellular matrix. Mice deficient in FGF-2 have decreased bone mass.

Platelet-Derived Growth Factor

Platelet-derived growth factor (PDGF) was originally isolated from blood platelets and can stimulate bone formation, and it is also present in bone. The two gene products, PDGF-A and PDGF-B, form dimers. PDGF-AA is the major product of bone cells, but PDGF-BB is a more active stimulator of osteoblast replication. PDGF also stimulates bone resorption by prostaglandin-dependent and prostaglandin-independent mechanisms.
CLINICAL EVALUATION OF METABOLIC BONE DISEASE

Bone Densitometry

The most widely used procedure for measuring bone mass is dual-energy x-ray absorptiometry (DXA).\textsuperscript{64} Other methods include quantitative computed tomography (QCT), quantitative radiography, single-energy x-ray absorptiometry, and ultrasonography.\textsuperscript{65,66} The correlations among these methods are quite variable.\textsuperscript{67,68,69}

Both densitometry and ultrasound data are reported in terms of T-scores (standard deviations from the young adult norm for that instrument) or Z-scores (standard deviations from the expected value for individuals of the same sex, age, and body size). These values depend on the normative data that have been obtained for each specific instrument. Moreover, the normative data are likely to be different for different populations, not only for men and women but also for members of different racial and ethnic groups. Methods of assessing microarchitecture using magnetic resonance imaging (MRI) and QCT are being developed but are not yet available for clinical use.\textsuperscript{70}

Dual-Energy X-Ray Absorptiometry

DXA can provide accurate and reproducible values for bone mineral content (BMC) and bone mineral density (BMD) in the lumbar spine, the proximal femur, the distal radius, and the whole body. BMD is calculated from the BMC and the area of bone scanned (g/cm\textsuperscript{2}); hence, it represents areal bone density rather than true volumetric density.

DXA has many advantages. Radiation exposure is minimal (<10 mrem), and scanning time is short (5 to 20 minutes). If quality control is maintained, variability of repeated readings is less than 1% for phantom standards; less than 2% for lumbar spine, total body, and radius; and less than 3% for proximal femur.

The major disadvantages of DXA are as follows:

1. Changes with disease progression or therapy are small in relation to the variability of the measurement.\textsuperscript{71}
2. The test is moderately expensive.
3. Anteroposterior measurements of the lumbar spine in older patients are subject to errors caused by aortic calcification and osteoarthritic changes.\textsuperscript{72}

Quantitative Computed Tomography

QCT, which employs instruments available in most radiology departments, can be used to assess true bone density (g/cm\textsuperscript{3}) and to separate cancellous and cortical bone in the vertebral body. QCT has also been used to measure cortical and trabecular bone density in the appendicular skeleton. The radiation exposure (100 to 300 mrem) is larger than for DXA, and the precision and accuracy are lower but within the acceptable range. A major disadvantage may be cost, but this varies widely.

Peripheral Densitometry

A number of methods to measure bone mass and density in the appendicular skeleton have been developed that are less expensive, faster, and more portable than DXA or QCT.\textsuperscript{73} Measurement of cortical bone in the shaft of the radius and ulna and trabecular bone in the distal radius by radiography, photon absorptiometry, or CT scanning is precise and can be used to predict fracture risk in populations, but it cannot predict BMD of the spine and hip in individual patients. The advantages of ultrasonography, particularly of the calcaneus, are that (1) it does not use x-rays, (2) it is rapid and portable, and (3) it has the capability of predicting fracture risk. These measurements may be particularly useful for large-scale screening programs.
Biochemical Measurements

One of the most important advances in metabolic bone disease has been the development of more accurate biochemical measurements that can assess rates of bone formation and resorption. In population studies, these methods have been used to show that increased turnover (i.e., high rates of both resorption and formation) correlates inversely with bone mass and may predict a high rate of bone loss and an increased risk of fracture. However, markers currently available are characterized by a wide normal range and considerable variability, which limit their use in individual patients. The most common current clinical use of these assays in care of patients is to obtain a more rapid assessment of the response to antiresorptive agents, which can be detected at 3 to 6 months, before changes in BMD.
Markers of Bone Formation

Alkaline Phosphatase

Total serum alkaline phosphatase is measured to assess osteoblastic activity in Paget's disease, primary hyperparathyroidism, osteomalacia, and rickets. An immunoassay that selectively measures the bone isoenzyme may increase the usefulness of this test in osteoporosis, in which changes in osteoblastic activity are smaller. High bone-specific alkaline phosphatase values have been shown to predict bone loss and fractures.\(^8\)

Osteocalcin

Osteocalcin, a bone carboxyglutamic acid-containing protein, is one of the few proteins that are relatively specific for skeletal tissue. A fraction of the osteocalcin synthesized by osteoblasts is released into the circulation. Carboxyl-terminal cleavage of the molecule may occur after release, but both the intact and amino-terminal portions can be measured by specific immunoassays. Serum osteocalcin correlates with skeletal growth rates in childhood and puberty and is increased when bone turnover is accelerated (e.g., in hyperparathyroidism, hyperthyroidism). In Paget's disease, osteocalcin is elevated to a lesser degree than alkaline phosphatase.

Because osteocalcin production is increased by 1,25(OH)\(_2\)D, the levels may be low in osteomalacia and rickets even when alkaline phosphatase is elevated. Conversely, osteocalcin levels may be selectively reduced in patients given glucocorticoids to a greater degree than other formation markers. Undercarboxylated osteocalcin is present in vitamin K and vitamin D deficiency, increases with age, and is associated with increased fracture risk.

Procollagen Peptides

The amino-terminal and carboxyl-terminal extension peptides of procollagen (see Fig. 27-2), which are removed during processing of collagen, are released into the circulation. The N-terminal propeptide assay may be more sensitive and reliable than the C-terminal assay.\(^8\) Their measurement is an index of total-body synthesis of collagen, the bulk of which is derived from bone. Procollagen peptide levels correlate with histologic measures of bone formation. Levels of procollagen peptides are high in infants and may provide a clinically useful index of growth.
Markers of Bone Resorption

Calcium

Measurement of fasting urinary calcium excretion is convenient but shows wide variation, reflecting the net result of intestinal absorption, bone resorption, and mineralization as well as renal tubular handling of calcium. Markedly increased urinary calcium occurs with a marked increase in osteoclastic activity with little change in formation, for example, in some patients with osteolytic bone metastases.

Hydroxyproline

Collagen degradation releases hydroxyproline into the circulation in both free and peptide-bound forms. Because bone resorption is by far the largest contributor to collagen breakdown, urinary hydroxyproline excretion has been used as a measure of bone resorption. However, 80% to 90% of the released hydroxyproline is metabolized, and hydroxyproline from collagen or gelatin in the diet is excreted in urine. Therefore, the sample should be obtained after a 12-hour fast or while the patient is receiving a gelatin-free diet. The hydroxyproline assay has been most useful in conditions in which resorption is markedly increased, such as Paget's disease, hyperparathyroidism, and malignancy. Because cross-link excretion can also be used in these conditions, hydroxyproline assays are used less frequently in clinical assessment.

Collagen Cross-Links

Unlike hydroxyproline, the pyridinoline and deoxypyridinoline cross-links that stabilize collagen in the extracellular matrix (Fig. 27-11) are not metabolized but are excreted in the urine in either a free or peptide-bound form. The deoxypyridinoline cross-link is almost entirely derived from skeletal tissue and therefore is a more sensitive indicator of bone resorption than pyridinoline, which is also found in skin and other connective tissues.

Measurement of total urinary pyridinoline and particularly deoxypyridinoline by high-performance liquid chromatography (HPLC) probably provides the best measure of bone resorption but is expensive and time-consuming. Immunoassays have been developed for free pyridinoline and deoxypyridinoline as well as for peptides that include these cross-links and are released during resorption. These assays can now be carried out in serum as well as urine. Measurements correlate with bone turnover and change in response to agents that affect resorption. Hence, they are useful in assessing changes in resorption in the course of disease or in response to therapy. They may also be valuable in identifying patients with high bone turnover, who not only have low bone mass but also lose bone rapidly and are more likely to develop osteoporosis.

All of these assays have shown diurnal variation and may also be affected by meals. For urine assays, a fasting, second-voided morning urine sample is probably the most reliable.

Other Assays

Tartrate-resistant acid phosphatase is secreted by osteoclasts into serum and may be useful as a measure of bone resorption. Other collagen breakdown products, such as glycosylated hydroxylysines, may be used to assess resorption. Bone sialoprotein is a product of osteoblasts but apparently is not released during bone formation, it correlates best with other measures of bone resorption.
Bone Biopsy

Transiliac bone biopsy can provide direct information about cancellous bone volume, the density of connections between trabecular plates (connectivity), and the function of bone cells. The rate of bone formation and mineralization can be measured by this technique with the use of dynamic histomorphometry after tetracycline labeling (Fig. 27-12), but bone resorption is more difficult to assess by bone biopsy. Bone biopsy necessitates the use of a large needle with a 7- to 9-mm internal bore and the technical skill to obtain a sample that is not crushed or distorted. The sample must be processed without decalcification and stained appropriately. Unstained sections are needed in order for one to see fluorescent tetracycline labels. Special stains may be used to identify mast cells in mastocytosis or aluminum in renal osteodystrophy.

Bone biopsies are rarely indicated in the clinical care of patients with osteoporosis but may be indicated for patients with unusual skeletal lesions or for young men or women who have osteoporotic fractures with no evident secondary cause. However, therapeutic decisions can generally be made without biopsies. Biopsies may be indicated more frequently in renal osteodystrophy because the different forms of this disorder are managed differently.
Radiographs and Bone Scans

The use of radiographs and bone scans in diagnosis is covered under the discussions of specific disorders. Radiographs are very important in detecting fractures. High-resolution radiographs and CT images have also been used to assess cortical porosity.

Bone scans using technetium 99m linked to a bisphosphonate are useful in localization of bone lesions. Uptake is a function of blood flow to the region and the amount of mineralizing bone. The test does not give information about the nature of the lesion but may serve as a guide for further studies. The current systems for MRI may be used to assess bone structure. Moreover, the type of tissue in the marrow or surrounding the bone can be identified. MRI is particularly useful in the detection of small soft tissue masses or vascular lesions in bone.
OSTEOPOROSIS

Primary Osteoporosis

Definition

Osteoporosis is by far the most common metabolic bone disease. One in two white and Asian postmenopausal women and at least one in eight older men and women of other racial backgrounds are likely to have an osteoporotic fracture at some time during their lifetime (Fig. 27-13). Osteoporosis has been better defined by the Consensus Development Conference (1) as "a disease characterized by low bone mass and microarchitectural deterioration of bone tissue, leading to enhanced bone fragility and a consequent increase in fracture risk."

Currently, diagnostic categories for postmenopausal women are based on measurements of BMD (Table 27-2). These categories are clearly arbitrary but do give some indication of fracture risk. However, the risk of fracture at any given BMD increases markedly with age and can be affected by a number of other factors.

A more rational approach to diagnosis and management might be to obtain an estimate of fracture risk based on all factors in individual patients. The use of T-scores to categorize BMD measurements as indicating the presence or absence of osteoporosis is complicated by the fact that the estimation of fracture risk is both site and method specific. Despite great advances, there are still many unresolved questions in defining and diagnosing osteoporosis.

Epidemiology

Osteoporosis has been considered a disorder of postmenopausal women of Northern European descent because they have high rates of fractures. However, the frequency of osteoporotic fractures is also high in other populations and is likely to increase further as life expectancy increases. Moreover, the age-adjusted incidence of hip fractures around the world is rising, possibly related to increasing industrialization and decreasing physical activity. Most of the epidemiologic data are for hip fractures, but vertebral fractures are equally common. In one study, the lifetime risk of osteoporotic fractures of the hip, spine, or wrist after age 50 years was about 40% in women and 13% in men (Table 27-3). The temporal pattern of the increase in fracture incidence differs for the hip, spine, and wrist (Fig. 27-14). The incidence of osteoporotic fractures varies among populations, possibly because of variations in skeletal architecture and turnover and in bone mass. For example, in South Africa the incidence of hip fracture in Bantus is only a fraction of that in whites, although the Bantus have lower bone densities.

Pathogenesis

Understanding of the pathogenesis of primary osteoporosis remains largely descriptive. Decreased bone mass and increased fragility can occur because of (1) failure to achieve optimal peak bone mass, (2) bone loss caused by increased bone resorption, or (3) inadequate replacement of lost bone as a result of decreased bone formation. Moreover, an analysis of the pathogenesis of osteoporosis must take into account the heterogeneity of clinical expression.

TABLE 27-2 - Diagnostic Categories for Osteoporosis Based on Measurements of Bone Mineral Density and Bone Mineral Content

<table>
<thead>
<tr>
<th>Category</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>A value for BMD or BMC ± 1 SD of the young adult reference mean</td>
</tr>
<tr>
<td>Low bone mass (osteopenia)</td>
<td>A value for BMD or BMC &gt; 1 SD and &lt; 2.5 SD lower than the young adult mean</td>
</tr>
<tr>
<td>Osteoporosis</td>
<td>A value for BMD or BMC &gt; 2.5 SD lower than the young adult mean</td>
</tr>
<tr>
<td>Severe osteoporosis (established osteoporosis)</td>
<td>A value for BMD or BMC &gt; 2.5 SD lower than the young adult mean in the presence of one or more fragility fractures</td>
</tr>
</tbody>
</table>

BMC, bone mineral content; BMD, bone mineral density; SD, standard deviation.

Inadequate Peak Bone Mass

Studies of twins suggest that genetic determinants are responsible for up to 85% of the variation in peak bone mass. However, genetic factors may be less important in determining fracture incidence in elderly people. Polymorphisms of candidate genes, including vitamin D and estrogen receptors, collagen, cytokines, apolipoprotein E, and growth factors, have been analyzed to assess their possible roles in determining peak bone mass, remodeling, and fracture risk. The results generally show small effects or are inconsistent. This problem may reflect the fact that it is difficult to determine the appropriate control population or that these polymorphisms may reflect effects of linked genes. Moreover, gene effects may be influenced by environmental factors.

A broader search for quantitative trait loci associated with differences in bone mass has identified a number of chromosomes that may be involved. This approach is likely to lead to the identification of a number of genes that influence the skeleton and are determinants not only of peak bone mass but also of microarchitecture and turnover. Furthermore, microarchitectural features of the skeleton, such as the length of the femoral neck, are genetically determined and can influence fracture incidence. High levels of physical activity and good calcium intake during childhood and puberty can help achieve maximal peak bone mass.
Skeletal development involves a number of systemic hormones, including glucocorticoids, growth hormone, thyroid hormones, and particularly gonadal steroids. These hormones are responsible for the initiation of the pubertal growth spurt, and a delay in puberty is associated with a decrease in peak bone mass. Estrogen plays the primary role in epiphyseal closure and the decrease in bone remodeling after puberty in both men and women. Thus, men with defects in the estrogen receptor or in the enzyme aromatase, which converts testosterone to estrogen, show failure of epiphyseal closure, low bone mass, and high turnover. These changes can be reversed by estrogen administration in patients who lack aromatase.

**Pathogenetic Factors**

Bone mass achieved during puberty is maintained throughout life, with Evidence of bone mass remaining stable for years because resorption and formation are equal. However, bone loss in women, particularly in the distal radius, may begin well before menopause. The incidence of Colles’ fractures increases in women in their 40s. In contrast, the increase in vertebral crush fractures is greatest 5 to 15 years after menopause and hip fractures show a logarithmic increase later in life. These differences indicate that rates of bone loss and changes in fragility, although related to menopause and age, vary in different sites.

Excessive bone resorption probably reflects an increased number of sites of resorption rather than an increase in the amount of bone resorbed at each site. However, prolonged osteoclastic activity may play a role, increasing the depth of each resorption cavity and causing trabecular perforation. This activity may be due to loss of estrogen and other factors that normally cause programmed cell death (apoptosis) in osteoclasts. The finding that markers of bone turnover are increased in postmenopausal women, even in older people, suggests that excessive bone resorption is the predominant cause of osteoporosis. Moreover, both increased resorption and increased formation correlate inversely with BMD and directly with fracture risk in postmenopausal women and men as well. Biopsies of both men and women show loss of connectivity of trabeculae, presumably related to excessive resorption. Yet biopsies of patients with established osteoporosis often show decreased bone formation.

**Biochemical Abnormalities**

Classically, osteoporosis is differentiated from disorders, such as osteomalacia and osteogenesis imperfecta, by the fact that there is no obvious defect in mineralization or the structure of collagen. Subtle abnormalities may occur, however. For example, a polymorphism in the collagen gene COL1A1 can result in differences in the ratio of the 1 to 2 chains, which might alter collagen strength. Differences in crystal structure and alignment have also been described. However, these may simply be the consequence of differences in turnover. The possibility that one or more of the many noncollagen proteins in bone is abnormal in osteoporosis has not been adequately explored.

**Decreased Bone Formation**

Skeletal bone mass increases during puberty and young adult life, even though the rate of resorption is high. Thus, menopausal and age-related bone loss must involve relative impairment of bone formation. With age, the amount of bone formed decreases with each bone structural unit, as evidenced by a decrease in mean wall thickness. This decrease may be due to an age-related decline in skeletal growth factors.

### TABLE 27-3 — Estimated Lifetime Fracture Risk in Women and Men from Rochester, Minnesota, at Age 50 Years

<table>
<thead>
<tr>
<th>Fracture Site</th>
<th>Women (%) [95% Confidence Interval]</th>
<th>Men (%) [95% Confidence Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proximal femur</td>
<td>17.5 (16.8–18.2)</td>
<td>6.0 (5.6–6.5)</td>
</tr>
<tr>
<td>Vertebral</td>
<td>15.6 (14.8–16.3)</td>
<td>5.0 (4.6–5.4)</td>
</tr>
<tr>
<td>Distal forearm</td>
<td>16.0 (15.7–16.7)</td>
<td>2.5 (2.2–3.1)</td>
</tr>
<tr>
<td>Any of the above</td>
<td>39.7 (38.7–40.6)</td>
<td>13.1 (12.4–13.7)</td>
</tr>
</tbody>
</table>

*Clinical diagnosed fractures.


**Local Factors**

Other systemic hormones may play a role in age-related bone loss. PTH levels increase with age. The increase is probably due to decreased dietary intake and impaired intestinal absorption of calcium, often associated with vitamin D deficiency, but estrogen deficiency may also play a role. Increased PTH can accelerate cortical bone loss. However, PTH levels in patients with vertebral fractures are not different from those in age-matched control subjects. 25(OH)D levels are often decreased in osteoporotic patients but levels of 1,25(OH)2D are not consistently altered, although some older patients with hip fractures do have low levels. Calcitonin deficiency does not appear to play a role in osteoporosis, although pharmacologic doses of calcitonin can prevent bone loss or increase bone mass in patients with high bone turnover.

Glucocorticoid excess can produce secondary osteoporosis but does not appear to play a major role in primary osteoporosis. Growth hormone secretion and circulating IGF-I decrease with age, and decreased IGF-I levels have been reported in osteoporotic patients, particularly in males. Thyroid hormone excess may exacerbate osteoporosis.
Two features of osteoporosis suggest a role for local factors in pathogenesis:

1. Systemic hormones that influence the skeleton, including estrogen and PTH, alter the production of local factors (e.g., cytokines, prostaglandins, and growth factors).
2. Differential bone loss occurs in different parts of the skeleton.

Production of IL-1, TNF-α, and IL-6 may be increased in estrogen-deficient and osteoporotic patients. However, there have been negative results in studies of the role of cytokines in osteoporosis in humans. The most striking evidence for the role of cytokines in the bone loss of estrogen deficiency comes from rodent models. The loss of bone after ovariectomy in rats can be blocked by inhibiting the activity of IL-1 and TNF-α. Ovariectomy does not cause bone loss in mice lacking the IL-1-activating receptor. PGE2 production is increased in bones from oophorectomized animals and decreased by estrogen administration. Estrogen levels decrease and PTH levels increase IL-6 production. Studies in rodent models suggest the existence of an important interaction between cytokines produced by cells in the marrow, possibly both hematopoietic and mesenchymal, and osteoblasts. These local factors act through the RANKL-RANK system, including OPG, but defects in this system have not yet been identified in osteoporotic patients.

It is likely that skeletal as well as systemic production of IGF or IGF-binding proteins plays a role in osteoporosis. TGF-β and BMPs may regulate bone remodeling, and TGF-β levels are decreased in bones from oophorectomized animals. Decreased TGF-β levels may not only impair bone formation but also increase bone resorption.

**Nutrition and Lifestyle**

Calcium deficiency and decreased physical activity in early life can lead to a lower peak bone mass and may accelerate bone loss later in life. Calcium supplementation, particularly when given with vitamin D, can slow bone loss in elderly people. Vitamin D deficiency is common in elderly persons, particularly those with disabilities.

Protein deficiency may also be associated with bone loss. Vitamin K deficiency is associated with increased hip fracture risk. There are positive correlations between body fat, lean body mass, and bone density. One mechanism of protection of bone density in overweight people may be conversion of adrenal androgens to estrogens in fat. Another pathway may be the decrease in sex hormone globulin associated with increased body mass index. Finally, increased fat and muscle mass would lead to increased impact loading and mechanical stress on bone.

**Nonskeletal Factors**

Low body weight and weight loss are important risk factors for osteoporotic fractures. This association may be due in part to nonskeletal effects, such as decreased padding of the hip and decreased muscle strength. Neuromuscular factors, such as loss of muscle strength, impaired balance, and impaired vision, are important, particularly in increasing the risk of falls that result in hip fracture.

Drugs that affect the central nervous system or that decrease vascular volume and cause postural hypotension are likely to increase this risk, particularly in older adults.

**Clinical Features**

**Vertebral Crush Fractures**

Compression fractures of the vertebrae, which occur spontaneously or with minimal trauma, are the most common manifestation of osteoporosis. The terms postmenopausal osteoporosis and type I osteoporosis have been applied to vertebral crush fractures at a younger age and mainly in women, whereas senile osteoporosis and type II osteoporosis have been used for hip fractures in older women and men. These distinctions may not be helpful clinically. The disorders certainly are not separate, because patients with any type of osteoporotic fracture are more likely to have subsequent fractures of either the spine or the hip.

The clinical course of the vertebral crush fracture syndrome varies. Some patients exhibit compression of only one vertebra, and others show collapse of multiple vertebrae. Nevertheless, the risk of additional fractures is high. Vertebrae may show extensive loss of trabecular structure before they collapse. Radiologically, fractures can vary from mild end-plate deformities or anterior wedging to complete vertebral collapse. The most frequent fractures are in the thoracic vertebrae below T6 and in the lumbar vertebrae.

Patients with vertebral crush fractures usually have back pain that leads to radiologic assessment. Pain in the lumbar or sacral area, compared with pain in the thoracic area, is less likely to be associated with vertebral compression. Height loss is a sensitive indicator of compression, but height loss can occur without fractures as a result of narrowing of vertebral discs and postural changes.

Many patients are asymptomatic, particularly those who have only anterior wedging. Anterior wedging in the upper thoracic vertebrae (T5 to T8) is common in older men and women and is not necessarily associated with severe compression fractures. Bone density measurements can help predict fracture risk in these patients. Anterior wedging may also develop early in life, probably during pubertal growth. This condition, called Scheuermann’s disease, may be transmitted as an autosomal dominant disorder.

Multiple vertebral crush fractures cause severe impairment. Kyphosis and loss of the lumbar lordosis are deforming and can exacerbate back pain. Impairment of chest wall function may reduce vital capacity. Compression of abdominal contents may be disfiguring and uncomfortable. Ultimately, impingement of the ribs on the iliac crest is another source of pain. Many patients have additional spinal abnormalities, including spondylolisthesis, intervertebral disc disease, and osteoarthritis.
particularly in the spinal facets. Osteoporosis itself rarely compresses nerve roots or the spinal cord. Although patients with severe osteoarthritides are somewhat less likely to have osteoporosis, these two common disorders commonly occur in the same patient. 164

Hip Fracture

Fractures of the proximal femur are a major cause of morbidity and mortality in older people. Most fractures are in the femoral neck or at the base of the greater trochanter and are associated with trauma, although the trauma may be minimal. The risk is influenced by factors that increase the risk of falling and by the type of fall as well as the structure of the skeleton and surrounding soft tissue. 165 The increased incidence of hip fractures with age is caused both by increased falls and by continued bone loss. Vitamin D deficiency may play a role. 166

Hip fracture is usually treated surgically, and the costs are substantial. In addition, perioperative and postoperative complications are associated with a 5% to 20% mortality rate. Many elderly patients cannot return to their previous level of activity after hip fracture and require long-term nursing home care. It is important to perform a diagnostic evaluation and to develop a prevention plan for these patients because a second hip fracture or a fragility fracture at another site is likely to occur. Unfortunately, most patients with hip fractures do not undergo evaluation or treatment to prevent progression of osteoporosis and additional fractures. 167

Colles’ Fracture

Colles’ fractures of the distal radius, which is composed largely of trabecular bone, are caused by falling on the outstretched hand. The incidence in women begins to increase after age 40 years and may be associated with premenopausal and perimenopausal bone loss 168 and with genetic factors. 169 Unlike that of vertebral and hip fractures, the incidence of Colles’ fractures in men does not increase with age. Colles’ fractures usually heal well and only occasionally result in long-term morbidity. Women with a Colles’ fracture should be assessed for so that an appropriate treatment plan can be provided.

Other Fractures

Fractures at any site, with the possible exception of the face and skull, can be associated with osteoporosis. Measurements of bone mass and further diagnostic work-up are indicated for all fractures that occur with minimal trauma.

Osteoporosis in Men

The incidence of hip and spine fractures in men increases with age and is about one third that in women. 162 Men often have vertebral deformities associated with trauma earlier in life. In men, the increase in hip fractures tends to occur later in life, and a higher proportion of men have definable secondary causes. 169 Bone histomorphometry shows increased resorption in most patients. 169

Osteoporosis in men is associated with low estrogen and high SHBG levels. 164 Abnormalities of the IGF system are also implicated. 165 A diagnostic work-up and therapeutic plan should be provided for men with fragility fractures, but this is rarely carried out in practice. Screening for osteoporosis in older men who do not have fractures has not yet been evaluated but may be justified now that preventive therapy is available.

Juvenile Osteoporosis

Juvenile osteoporosis 160 is a rare, self-limiting disease that can begin between the ages of 8 and 14 years with back pain and vertebral compression. If bone turnover is high, antiinflammatory or anti-inflammatory agents may be beneficial. 164 However, deficient bone formation may be the critical defect leading to fractures in these children. 164 A mutation in type I collagen has been reported in one family with this disorder. 164 Spontaneous remission usually occurs, and the disorder usually does not lead to permanent deformity.

139

Idiopathic Osteoporosis

Osteoporosis with no obvious secondary cause in premenopausal women or younger men is called idiopathic osteoporosis. The term is not used consistently, and patients so defined include individuals with both high and low bone turnover. 164 Some patients have a transient, self-limited condition, whereas others have a progressive and disabling disease. Idiopathic osteoporosis can be associated with nonspecific inflammatory changes, and these cases may be caused by abnormal cytokine activity. A careful evaluation, including consideration of bone biopsy, should be made to search for secondary causes.

Osteoporosis in Pregnancy

Osteoporosis in pregnancy is rare and may represent bone disease that has been present before pregnancy. Ordinarily, maternal bone loss is modest, 164 presumably because the high levels of estrogen protect the skeleton and high levels of 1,25(OH)2 D stimulate calcium absorption. Bone loss can occur, however, particularly in adolescents. 164 Lactation is associated with transient bone loss but is not a risk factor for osteoporosis. 167

Localised Osteoporosis

Immobilization is the most common cause of localized osteoporosis (see later). Transient osteoporosis of the hip has been reported in middle-aged men and in pregnancy. 162 Regional migratory osteoporosis can occur without immobilization, particularly in the lower extremities. 163 This phenomenon may be associated with local inflammation or autonomic dysfunction with vasomotor changes and hyperesthesia, a syndrome called reflex sympathetic dystrophy. 166
Secondary Osteoporosis

The division of osteoporosis into primary and secondary forms is somewhat arbitrary. For example, patients with diseases that lead to hypogonadism early in life are considered to have secondary osteoporosis, whereas osteoporosis in women with natural menopause and older men with low sex hormone levels is termed primary. Moreover, many patients have a combination of primary and secondary causes. Although most postmenopausal women and older men do not have a definable secondary cause, the few who do can be treated more effectively. Therefore, this possibility should be considered in every patient. There are many causes of secondary osteoporosis (Table 27-4), only a few of which are discussed here.

Glucocorticoid-Induced Osteoporosis

The most common form of secondary osteoporosis is that induced by exogenous glucocorticoids. Cushing's syndrome, caused by an excess of endogenous glucocorticoids, is less common but may also involve osteoporosis at presentation. Patients with rheumatoid arthritis, chronic pulmonary disease, or gastrointestinal disease who receive exogenous glucocorticoids are at additional risk because disease-associated inflammation, poor nutrition, and immobilization can worsen bone loss. Glucocorticoid-induced osteoporosis is particularly common in postmenopausal women, presumably because they also have primary osteoporosis. However, fragility fractures can occur in any patient receiving moderate to high doses of glucocorticoids. The increased fracture risk appears within a few months of initiating therapy and rapidly declines after cessation of treatment.

Glucocorticoid-induced osteoporosis is a result of both increased bone resorption and decreased bone formation. Increased
inhibition of osteoblasts, which are highly sensitive to glucocorticoids. For example, as little as 2.5 mg of prednisone given at bedtime can block the normal nocturnal rise in osteoclastion. Glucocorticoids decrease osteoblast formation and increase osteoblast and osteocyte apoptosis. Glucocorticoids also increase urinary calcium and phosphate and can inhibit gonadal hormone production by blocking gonadotropin release. Levels of testosterone are low in men receiving prednisone, 20 mg/day or more. Exogenous glucocorticoids decrease secretion of corticotropin, thereby decreasing the production of adrenal androgens, which are the major precursors for estrogen formation in postmenopausal women.

Clinically, glucocorticoid-induced osteoporosis is similar to primary osteoporosis. Initial bone loss is predominantly trabecular and is best assessed in the spine or distal radius. However, rib fractures and aspecific necrosis of the femoral or humeral heads or the vertebralae are common in glucocorticoid-induced osteoporosis, although they are rare in primary osteoporosis. Glucocorticoid-induced osteoporosis can be reversible, particularly in young patients who are cured of Cushing’s syndrome, in patients who cannot discontinue glucocorticoid therapy, early preventive therapy may be effective. Bisphosphonates and PTH prevent bone loss in patients with glucocorticoid-induced osteoporosis. Bisphosphonates also decrease fracture risk; they may act in part by suppressing osteoblast and osteocyte apoptosis.

Hypogonadism

Hypogonadism can occur in either men or women and has multiple causes. Patients with primary hypogonadism related to ovarian or testicular failure or secondary hypogonadism related to hypothalamic or pituitary disease lose bone rapidly and often have fragility fractures. The hypogonadotropic group includes patients with anorexia nervosa, athletic amenorrhea, prolactinoma, or lesions of the pituitary gland or hypothalamus, including tumors. Undernutrition and hypercortisolism may also contribute to bone loss in anorexia nervosa and athletic amenorrhea. Loss of growth hormone may play a role in the osteoporosis of pituitary tumors. Patients with prolactinomas can secrete parathyroid hormone-related protein (PTHrP), which may cause bone loss.

Drug-induced hypogonadism is increasing in frequency. The drugs implicated include long-acting progestins used for contraception in young women and gonadotropin-releasing hormone (GnRH) analogues used to block hormone production in women with endometriosis and men with prostate cancer.

Other Endocrine Causes

Hyperthyroidism can produce bone loss; however, the increase in formation in young persons is usually adequate, and, if the disease is treated early, changes in bone mass are small. Although there is no general increase in fracture risk in hyperthyroid patients, long-term thyroid hormone excess may lead to bone loss and increased risk of fractures. In individuals at risk for osteoporosis, primary hyperthyroidism may be missed or excessive amounts of exogenous thyroid hormone may be administered for many years. Osteoporosis has been seen in patients with growth hormone deficiency, acromegaly, and prolactinoma and may be caused by gonadotropin deficiency and the loss of gonadal hormones with any large pituitary tumor. Patients with insulin-dependent diabetes mellitus, especially men, often have low bone mass and diminished bone formation, but osteoporotic fractures are not a major problem. The role of noninsulin-dependent diabetes mellitus in the pathogenesis of osteoporosis is unclear. Bone loss may occur, particularly in patients who are not obese, but the data on fracture incidence are inconsistent.

Malignancy

Multiple myeloma and other lymphoproliferative malignancies can produce a clinical picture that resembles primary osteoporosis. It is particularly important to exclude myeloma in patients with rapidly progressive vertebral crush fracture syndrome. Metastases to the spine may also cause vertebral compression and should be considered in the differential diagnosis, particularly for patients with normal bone density. These lesions can usually be detected by MRI.

Other Diseases

The incidence and severity of osteoporosis are increased in patients with chronic hepatic and intestinal disorders, probably for nutritional reasons and because these patients often receive glucocorticoids or other drugs that affect the skeleton. Abnormal cytokine production may be a factor in inflammatory bowel disease. Although it was initially thought that impairment of vitamin D function in hepatic and intestinal disease would cause osteomalacia, the most common lesion in such patients is osteoporosis. Unfortunately, these patients often do not respond to vitamin D supplementation. People with severe alcoholism can also have osteoporosis; however, low intakes of ethanol may be associated with increased bone mass and a decreased fracture risk.

Mastocytosis causes both osteoporosis and osteosclerosis, and the number of mast cells may be increased in the marrow of patients with primary osteoporosis. The functional significance of the mast cells is not known, although mast cells produce heparin, which can cause bone loss. Hyperplastic anemias, such as thalassemia, can also cause bone loss, partly because of bone erosion by the marrow and partly because of hypogonadism associated with transfusion-induced hemochromatosis. Osteoporosis after organ transplantation is common and results both from the underlying disease and from the drugs used to prevent graft rejection.

Drugs

A number of drugs can produce osteoporosis. Heparin stimulates bone resorption and inhibits bone formation and can cause osteoporosis. Patients receiving anticonvulsants, including phenytoin, barbiturates, and carbamazepine, often have low bone mass. Impairment of vitamin D metabolism has been described, but most patients have osteoporosis with normal mineralization. Immunosuppressive agents, such as cyclosporine and FK506, are associated with bone loss. GnRH analogues, which decrease production of gonadal hormones, can lead to osteoporosis. Some agents used in cancer chemotherapy probably act both by inhibiting osteoclasts and by suppressing gonadal hormones.
Diagnosis

As indicated by the World Health Organization, osteoporosis can be diagnosed before fracture occurs by measuring bone density. The frequency of diagnosis, therefore, depends on the frequency, site, and timing of bone density measurements. There is no general agreement about who should be screened or when screening should be done. A suggested approach is illustrated in Figure 27-17.

Screening at menopause is recommended for women who have multiple risk factors, such as low body weight or a personal or family history of fragility fractures (appendicular or axial fractures after a fall from standing height or less). Universal screening of postmenopausal women after age 65 years has been recommended as cost-effective. Bone density should also be measured in men and premenopausal women with fragility fractures. BMD measurements may also be useful in early postmenopausal women under consideration for hormone replacement therapy. Bone density measurements not only establish the severity of bone loss but also provide a baseline for monitoring the patient’s therapeutic response. The test may also enhance health-related behavior. The subsequent work-up should be the same whether osteoporosis is diagnosed on the basis of screening or after the finding of a fragility fracture.

The history should include a detailed analysis of calcium intake and nutrition, changes in height or weight, physical activity and lifestyle, smoking history, menstrual and reproductive history, and personal or family history of fragility fractures or other metabolic or endocrine disorders that may affect the skeleton. Physical examination should include a careful height measurement, assessment of the spine, and evaluation for thyroid or adrenal disease. Radiologic assessment of fractures may be possible using DXA as well as ordinary radiographs. In addition, MRI or CT may be indicated if there are neurologic changes or if fractures are associated with normal bone density, raising the possibility of malignancy.

A minimal laboratory screen should include measurement of serum calcium, preferably as ionized calcium or with albumin to permit correction for protein-bound calcium and fasting calcium excretion (most easily measured as the calcium/creatinine ratio in the second-voided morning specimen). Appropriate tests to exclude secondary causes of osteoporosis should be based on the history and physical findings. Serum phosphorus and alkaline phosphatase are useful in ruling out hyperparathyroidism and osteomalacia. Measurements of vitamin D metabolites, urine phosphorus, and PTH are indicated if the screening test results are abnormal. Serum electrophoresis, blood count, and erythrocyte sedimentation rate can help to rule out myeloma, and thyroid function should be assessed. Laboratory studies for Cushing's syndrome are indicated in patients with suggestive clinical features. Measurements of gonadal and pituitary hormones are indicated for younger patients with osteoporosis. Gluten-sensitive enteropathy should be ruled out in patients with weight loss or frequent bowel movements.

Despite the inverse correlation between markers of bone resorption and formation and bone mass, these measurements have wide variations and cannot substitute for measurements of BMD in the diagnosis of osteoporosis. Because elevated values of both resorption and formation markers do indicate increased risk for bone loss and fractures, these measurements may become useful in determining the need for therapy, particularly if they can be made more accurate and less expensive.
Prevention and Therapy

Although it is important to relieve pain and to limit the impact of deformities in established osteoporosis, the primary goal of treatment is to prevent fractures. Therefore, prevention and therapy are considered together.

Nutrition and Calcium Supplementation

The calcium intakes recommended for prevention and treatment of osteoporosis range from 1 to 2 g/day. In children and adolescents, intakes of 1000 to 1200 mg/day are recommended. Most studies indicate that calcium supplementation slows bone loss, and there is limited evidence that calcium supplementation alone can decrease fracture risk. Moreover, low calcium intakes in the presence of low calcium absorption increase the risk of hip fractures. High calcium intakes are generally safe, although it may be worthwhile to check urinary calcium levels.

There is no clear advantage for any particular calcium formulation. Calcium carbonate is inexpensive and, when taken with meals, is usually well absorbed, even in patients with achlorhydria. Calcium citrate and other salts may be absorbed better than calcium carbonate in the fasting state. It is also worthwhile to include foods high in calcium in the diet. Patients should be informed about the calcium content of the major food sources, such as dairy products.

Vitamin D intake should be at least 400 U/day, and up to 2000 U/day is probably safe; higher levels may produce hypercalcemia or hypercalciuria. Calcium and vitamin D increase bone mass, decrease seasonal bone loss, and decrease the incidence of fractures, particularly in populations likely to have deficient intakes or limited sun exposure. Other forms of vitamin D have been used, including calcitriol (25(OH)D), calcitriol (1,25(OH)2D), and 1(OH)D, but there is no direct evidence that these are superior to ordinary vitamin D, which is less expensive. Dietary intakes of other minerals and of vitamins C and K, which are important for bone matrix synthesis, as well as protein should be adequate.

Exercise, Lifestyle, and Prevention of Falls

The role of exercise in treatment of osteoporosis has not been defined. On the basis of limited data, ¼ hour of weight-bearing exercise per day is recommended for patients who can tolerate it. Epidemiologic data suggest that lifetime leisure exercise is associated with higher BMD at the hip but may have no effect on fracture incidence. Patients are often better able to develop and maintain a suitable exercise program under the supervision of a physical therapist. Patients should also be instructed in body mechanics and posture in order to minimize musculoskeletal damage and the likelihood of falls. They should stop smoking and limit their intake of alcohol.

Medicines that cause prolonged sedation or postural hypertension should be avoided in elderly people. Older patients at high risk may be able to avoid hip fractures by wearing a hip protector. Help should be provided for coping with osteoporosis and for designing a lifestyle that maintains function and minimizes fracture risk. Excessive sodium intake should be avoided because it can increase urinary calcium excretion.

Management of Fractures

Fractures of the hip as well as other appendicular fractures are generally treated surgically. Vertebral fractures may require transient bed rest. A careful but intensive program of rehabilitation is critical in patients with fractures of the hip and spine. Pain relief for patients with vertebral crush fractures can usually be achieved with mild analgesics and local physical therapy. Calcitonin has analgesic effects and may be useful in patients with severe pain.

Surgical treatment of individual vertebral fractures by injection of methylacrylate has been used to relieve pain and expand the compressed vertebral body. Kyphoplasty employs a balloon to create a space for the resin, whereas in vertebroplasty the resin is injected directly.

Hormone Replacement Therapy

Postmenopausal women with and without osteoporosis should be informed about the benefits and risks of hormone replacement therapy. Many different regimens and forms of estrogen have been shown to slow bone loss; the choice is based on tolerance and convenience. Estrogen induces a decrease in bone resorption in postmenopausal women at all ages. Bone formation also decreases, but bone mass usually increases. Although the increase may result primarily from filling in of the remodeling space, estrogen may also have some anabolic effect, particularly in high doses. The rate of increase in bone mass tends to decrease with time.

Patients with an intact uterus should be given progestagens to prevent endometrial hyperplasia and an increased risk of endometrial cancer. Intermittent progestagen use reduces uterine bleeding. Low-dose, continuous progestagen may be more acceptable in older patients because it usually causes only transient breakthrough bleeding. Older women may respond to lower doses of estrogen, particularly if they combine calcium and vitamin D. At the lowest biochemically effective doses (0.25 mg of micronized estradiol per day), side effects of breast tenderness, menstrual bleeding, and endometrial hyperplasia have been minimal. All patients taking estrogen should have annual mammograms and should undergo appropriate evaluation if atypical uterine bleeding occurs.

The most important contraindication to the use of estrogen is breast cancer. A strong family history of breast cancer is a relative contraindication. Estrogen use is associated with a small but significant increase in breast cancer incidence. The incidence may be increased when progesterin is added. Hormone replacement therapy increases the incidence of thromboembolic disease and may exacerbate headaches or hypertension. Thus, the presence of thromboembolic disease, hypertension, or migraine headaches may also be a relative contraindication. Nevertheless, in epidemiologic studies, patients receiving hormone replacement therapy had lower mortality rates.

Data concerning the overall benefits of estrogen in prevention of cardiovascular disease and Alzheimer's disease are conflicting. Moreover, whereas prevention of bone loss has been adequately demonstrated, data on reduction of fractures are limited, in part because no large prospective trials have been completed. Although significant bone loss occurs in fewer than 15% of women taking estrogen, rapid bone loss can occur after estrogen is discontinued.

Bisphosphonates

Bisphosphonates are pyrophosphate analogues that bind to bone mineral and inhibit resorption. The first compound available for clinical use, etidronate, inhibits bone mineralization at high doses but increases bone mass without impairing mineralization when given intermittently. Second-generation bisphosphonates, such as alendronate and risedronate, do not impair mineralization at therapeutic doses and can be given continuously. Bisphosphonates bind to bone mineral and are then taken up by osteoclasts and rapidly inhibit bone resorption. The inhibitory effects on osteoclast function differ for the first-generation compounds, which lower adenosine triphosphate levels at high concentrations, and the second-generation compounds, which impair isoprenylation of proteins at low concentrations.

Alendronate and risedronate are approved for prevention and treatment of osteoporosis on the basis of evidence that they decrease bone resorption, increase bone mass in the spine and hip, and decrease the incidence of fractures. Bisphosphonates can prevent bone loss in patients receiving glucocorticoids and...
in men. There is no consensus on the duration of therapy, but continued benefit has been observed in patients treated for up to 7 years.

Bisphosphonates are poorly absorbed orally and must be taken on an empty stomach with no food or other medication. The major side effects are gastrointestinal, particularly esophageal irritation. This problem can be circumvented by using parenteral bisphosphonates. Pamidronate is available, although not approved for osteoporosis therapy, and other still more potent intravenous bisphosphonates are being developed.

Gastrointestinal side effects may also be reduced by giving bisphosphonates weekly instead of daily. Because they act by different mechanisms, bisphosphonates can be added to estrogen for patients who are responding poorly or have severe disease. However, it is not known whether combined therapy further reduces fracture risk.

Calcitonin
Calcitonin, an inhibitor of bone resorption, can increase bone mass, particularly in association with high turnover rates. Calcitonin also has some analgesic properties and may be particularly useful in patients with recent painful vertebral fractures. It is available either for subcutaneous injection or as a nasal spray. The former preparation is probably more effective but is less well tolerated, often producing gastrointestinal side effects. In a 3-year randomized trial, nasal calcitonin at 200 U/day was found to decrease fracture incidence significantly, although the effects on bone turnover and bone mass were diminished by the end of the study. Doses of 100 or 400 U/day did not significantly decrease fractures. Nasal calcitonin is less effective in increasing BMD than alendronate.

Selective Estrogen Receptor Modulators
A number of compounds have had effects similar to those of estrogen on bone, but they act as antagonists in the breast.

Tamoxifen has been shown to diminish bone loss in women with breast cancer. Another selective estrogen receptor modulator, raloxifene, has not only prevented bone loss but has also reduced the risk of fracture in osteoporotic patients. Its effects on bone turnover and mass are somewhat less than those of estrogen. Raloxifene does not stimulate the breast or uterus and may even decrease the risk of breast cancer. It reduces low-density lipoprotein levels but does not increase high-density lipoprotein. It is associated with an increased risk of thromboembolism and may produce hot flashes.

Other Therapeutic Approaches
Many years ago, PTH given intermittently in low doses was shown to increase bone mass in animals.

Parathyroid Hormone
The use of intermittent low-dose PTH in humans has produced a substantial increase in trabecular bone mass with little loss or even a gain in cortical bone in the femur and has reduced the incidence of fractures. Treatment with PTH alone or in combination with an antiresorptive agent is likely to be the most effective approach in patients who lose bone or continue to have fractures with estrogen or bisphosphonates. PTH may be particularly useful in glucocorticoid-induced osteoporosis or in patients given GnRH analogues. PTH must be given by injection, and patients must be monitored carefully for hypercalcemia and hypercalciuria.

Fluoride
Fluoride increases bone formation and has been used in the treatment of vertebral crush fractures. However, in a large prospective controlled trial of monofluorophosphate, the incidence of vertebral fractures was not decreased. A slow-release fluoride preparation was reported to decrease fractures in a smaller number of osteoporotic patients. Fluoride is not approved for treatment of osteoporosis in the United States.

Anabolic Steroids
Anabolic androgenic steroids (e.g., testosterone) and analogues (e.g., nandrolone) may increase bone as well as muscle mass. However, high doses produce unacceptable androgenic side effects in many women. Testosterone therapy is effective in hypogonadal men. Testosterone can also increase bone and muscle mass in older men, particularly those with low levels of bioavailable testosterone.

Thiazides
Thiazides can decrease urinary calcium excretion and increase bone mass in patients with hypercalciuria and may reduce cortical bone loss in normal postmenopausal women. The incidence of hip fractures may be lower in patients receiving thiazides, but the data are inconsistent. Thiazides can be combined with amiloride to minimize potassium loss. This therapy is particularly appropriate in patients with hypercalciuria and osteoporosis.
RICKETS AND OSTEOIMALACIA

Rickets and osteomalacia are disorders of the mineralization of newly synthesized organic matrix. In adults, the disorder involves only bone; in children, however, defects also occur in the growth plate and in the mineralization of cartilage, leading to characteristic deformities.

Pathogenesis

To understand the pathogenesis of rickets and osteomalacia, we should recognize that vitamin D is a prohormone that can be synthesized in the skin or supplied in the diet. Vitamin D deficiency is usually the combined result of deficient sun exposure and decreased dietary intake or intestinal malabsorption. Rickets and osteomalacia can also be caused by metabolic defects in the vitamin D hormone system, including inadequate activation in the liver and kidney and abnormalities of the vitamin D receptor (see Chapter 28). Mineralization can also be impaired when the supply or transport of mineral is inadequate in renal, intestinal, or bone cell disorders.

Nutritional and Gastrointestinal Disorders

Inadequate vitamin D intake is less common in the United States than in other countries because many foods are supplemented with this vitamin. However, the combination of inadequate sunlight or lack of the appropriate ultraviolet wavelengths, which occurs during the winter in the northern half of the United States, and failure to provide vitamin D supplements in the diet can lead to rickets in infants and osteomalacia in older persons. Individuals with darker pigmentation of the skin are more susceptible to vitamin D deficiency because they are less efficient in converting 7-dehydrocholesterol to vitamin D. For example, nutritional rickets occurs in black infants who are breast-fed without vitamin D supplementation. Adult Asian Indians in the United States and Europe have low 25(OH)D levels and are more likely to have osteomalacia. Intestinal malabsorption of fat can also cause deficiency of vitamin D and of other fat-soluble vitamins. Inability to produce adequate amounts of 25(OH)D can occur in advanced liver disease and with use of antiepileptic drugs.

Calcium deficiency rickets may differ from other forms of rickets and osteomalacia, particularly in adolescents, who may have genu valgum without end-plate deformities. In contrast, phosphate deficiency causes typical rickets. Because most foods contain phosphate, this form of rickets requires markedly unbalanced nutrition, such as can occur with prolonged intravenous feeding, removal of phosphate by dialysis with a low-phosphate solution, or use of aluminum-containing antacids, which bind phosphate in the intestine.

Renal Defects

Impairment of 1-hydroxylase can occur because of loss of renal mass or in renal tubular disorders such as the Fanconi syndrome. A hereditary deficiency of 1-hydroxylase, termed vitamin D-dependent rickets type I or pseudovitamin D deficiency, is a rare autosomal recessive disorder in which rickets develops during the first year of life. It occurs in rodents with homogenously inactivating mutations of the 1-hydroxylase gene and responds to physiologic doses of calcitriol.

Hereditary Resistance to Vitamin D

This severe form of rickets occurs in members of families who are homozygous for defects in the vitamin D receptor gene. These patients also often have alopecia. They may show a skeletal response to high doses of calcium and phosphorus, but this does not reverse the alopecia.

Familial X-Linked Hypophosphatemia

Originally termed vitamin D-resistant rickets, this syndrome is caused by a defect in phosphate transport. Although the most apparent abnormality is decreased renal tubular reabsorption of phosphate, phosphate transport may be impaired in other cells, particularly osteoblasts. Mutations in a gene (PEX) for a membrane-bound endopeptidase have been found in this disorder. This enzyme is thought to inactivate a hormone, phosphatonin, that inhibits phosphate transport. These patients may also have some impairment in 1-hydroxylase activity, and treatment involves a combination of calcitriol and phosphate. However, it is difficult to achieve normal growth rates in this disorder.

The condition exhibits genetic and phenotypic heterogeneity. Autosomal dominant and recessive as well as X-linked transmission has been described. The autosomal dominant form has been linked to a mutation in a new member of the FGF family, FGF-23; hence, FGF-23 may be the phosphatonin hormone. Although most cases are familial, some severe sporadic cases have their onset at puberty.

Oncogenic Osteomalacia

This severe form of osteomalacia may be caused by production of phosphatonin by fibrous and mesenchymal tumors.


Figure 27-19 Active osteomalacia in a patient (a sibling of the patient in Fig. 27-18) with hereditary tissue resistance to 1,25-dihydroxyvitamin D at age 18 with pseudofracture of the left tibia. (From Marx SJ, Spiegel AM, Brown EM, et al. Familial syndrome of decrease in sensitivity to 1,25-dihydroxyvitamin D. J Clin Endocrinol Metab 1978; 47:1303-1310. Copyright © 1978, by The Endocrine Society.)

that are often small and difficult to identify. These tumors express FGF-23. Removal of the tumor causes rapid reversal of the osteomalacia.

Hypophosphatasia

A rare autosomal recessive disorder, hypophosphatasia is characterized by mutations in the gene for the tissue-nonspecific (liver-bone-kidney) isoenzyme of alkaline phosphatase.
Clinical manifestations vary from death in utero related to severe deformities through infantile and childhood rickets to adult-onset osteomalacia. Premature loss of deciduous teeth and impaired dentition in adults are common. Levels of organic phosphate compounds, such as pyridoxal 5-phosphate, in plasma are increased.

Fluoride and Bisphosphonates

High doses of sodium fluoride or of first-generation bisphosphonates (e.g., etidronate) can produce osteomalacia. Anticonvulsant therapy in patients with a marginal vitamin D supply can cause osteomalacia by decreasing 25(OH)D levels.
Clinical Features

Rickets differs from osteomalacia in that it occurs before closure of the epiphyses. Enlargement of cartilage at the growth plate causes the so-called rachitic rosary at the costochondral junctions of the ribs and widening of the cartilaginous ends of the long bones. Impaired mineralization results in bowing of long bones. Radiologically, widening, cupping, and fraying of the metaphyses are seen (Fig. 27-18). Severe vitamin D deficiency causes muscle weakness, and this weakness, combined with the deformity of the chest wall, causes an increased incidence of pneumonia. The clinical expression of osteomalacia in adults varies widely. The most common deformity is bowing of the legs, and in severe cases bone pain and weakness may cause the patient to be bedridden. 

Radiographic changes include subperiosteal erosions caused by marked secondary hyperparathyroidism and a virtually pathognomonic but relatively uncommon lesion, the so-called pseudofracture (Looser’s zones or Milkman’s syndrome) of the long bones, ribs, scapulae, or public rami (Fig. 27-19). Coarsening of the trabecular pattern in the spine may be present, but it is also seen in osteoporosis.
Diagnosis

Although clinical features may point to rickets or osteomalacia, the diagnosis depends on laboratory studies. The biochemical picture can vary with different pathogenetic mechanisms and with different stages of disease. In infants with vitamin D deficiency, serum calcium may be low and the serum phosphorus concentration may be normal initially; as secondary hyperparathyroidism develops, however, calcium concentrations usually return to the low-normal range and serum phosphorus levels fall further. In advanced stages, the serum calcium concentration may fall again. This fall has been attributed to the inability of secondary hyperparathyroidism to maintain the serum calcium level when the bone surface is covered by osteoid and is resistant to attack by osteoclasts.

In adults, the characteristic picture is a low-normal or slightly decreased serum calcium level, a decreased urinary calcium level, and a low serum phosphate level. Increased alkaline phosphatase levels reflect the activity of the osteoblasts, which form unmineralized matrix. PTH levels may be markedly increased. The key diagnostic test in vitamin D deficiency is demonstration of a decreased serum 25(OH)D value. The 25(OH)D levels may also be decreased in hepatic disease or with drugs that impair 25-hydroxylation. The 1,25(OH)$_2$D levels may be normal in vitamin D deficiency, presumably because of a maximal stimulation of 1-hydroxylase by the low serum phosphorus and high PTH levels. Nevertheless, the amount of this hormone is inadequate to activate the receptors in intestine and bone. Because of the high activity of 1-hydroxylase, administration of vitamin D to these patients causes a further increase in 1,25(OH)$_2$D to supranormal levels.

The diagnosis of other forms of rickets and osteomalacia can be made by measuring vitamin D metabolite levels. For example, low values of 1,25(OH)$_2$D and normal levels of 25(OH)D suggest a defect in 1-hydroxylase that may be genetic or acquired as a result of loss of renal function or tumor-induced osteomalacia. High levels of 1,25(OH)$_2$D and normal levels of 25(OH)D are seen in patients with vitamin D receptor defects. In X-linked hypophosphatemia, serum phosphorus levels are low and levels of serum calcium and vitamin D metabolites are normal. If alkaline phosphatase levels are low, a definitive diagnosis of hypophosphatasia should be sought by measuring pyridoxal 5-phosphate; elevated levels are relatively specific for hypophosphatasia and correlate with clinical severity.

Although the diagnosis of osteomalacia can usually be made on the basis of clinical findings and laboratory studies, a bone biopsy is sometimes needed for a definitive diagnosis. The characteristic finding is markedly widened osteoid seams and impaired mineralization with diffuse or absent tetracycline labeling. The bone also shows great variation in trabecular width and resorption lacunae resulting from secondary hyperparathyroidism. A modest increase in osteoid width can occur in a high-turnover state, such as hyperparathyroidism, thyrotoxicosis, or Paget's disease; however, tetracycline labeling shows a normal mineralization front in these conditions.
Therapy

Vitamin D is effective in the treatment of nutritional and malabsorptive rickets and osteomalacia. High doses (50,000 to 100,000 U/day) may be given initially, but it is important not to overtreat patients because vitamin D is stored in the fat and excessive amounts can cause prolonged hypercalcemia and hypercalciuria. Monitoring of urinary calcium excretion is useful to determine when the vitamin D dose should be decreased. Patients with malabsorption may require large doses or parenteral vitamin D. Patients with defects in 1-hydroxylase can be treated with calcitriol. If the cause is tumor-induced osteomalacia, the treatment is to find and remove the lesion, but oral or intravenous phosphate and calcitriol can be used to heal the skeletal lesions.

Severe defects in the vitamin D receptor are most difficult to treat. Massive doses of calcium and phosphorus have been given to these patients, but normal growth is rarely achieved and alopecia persists. Intravenous calcium therapy may be necessary and is effective in restoring bone mineralization. Similarly, normal growth may not occur despite repletion of phosphorus and administration of calcitriol in patients with X-linked hypophosphatemia, although bone mineralization can be restored. Growth hormone has been used to help achieve normal height. Careful monitoring of levels of calcium, phosphorus, and vitamin D metabolites is important to prevent impairment of renal function. At present, there is no effective therapy for hypophosphatasia.
HYPERPARATHYROID BONE DISEASE

In the past, severe cases of primary hyperparathyroidism showed osteitis fibrosa cystica generalisata, manifested by generalized bone loss with increased bone resorption, including both subperiosteal and endosteal surfaces. The formation of fibrotic cystic lesions (brown tumors) in the long bones and jaw caused swelling, pathologic fractures, and bone pain. This bone disease is now rarely seen in primary hyperparathyroidism but may occur in poorly managed secondary hyperparathyroidism.

With the common forms of hyperparathyroidism, the major finding is an increased rate of remodeling in bone. Bone mass is decreased in the cortex, largely because of endosteal resorption, but trabecular bone is preserved. Bone density is low in the cortical bone of the radius, but bone density in the metaphyses and vertebrae, which represent largely trabecular bone, is normal or only moderately decreased. Patients with mild to moderate disease do not have progressive bone loss; however, when these patients are cured surgically, bone density in the spine can increase by as much as 15%, even in post-menopausal women who are at the highest risk for fracture. Bone density also increases in the radius and hip. Moreover, bone loss is attenuated in postmenopausal women with hypoparathyroidism. On the basis of these results, patients with primary hyperparathyroidism and low levels of bone density who are at risk for fractures are candidates for parathyroid surgery.
RENAL OSTEODYSTROPHY

In view of the central role of the kidney in regulating mineral metabolism, it is not surprising that patients with chronic renal failure frequently have skeletal abnormalities. In renal failure, the decreased capacity to synthesize 1,25(OH)_2 D and to excrete phosphate causes secondary hyperparathyroidism. This results from the lowering of serum calcium by phosphate, the impairment of calcium absorption in the intestine, and the loss of the feedback inhibitory effect of 1,25(OH)_2 D on PTH production. The increased secretion of PTH ultimately leads to osteitis fibrosis cystica.

Development of bone disease can be slowed or prevented by phosphate restriction or by treatment with phosphate binders and calcitriol. In the past, aluminum hydroxide was used to bind phosphate, a therapy that sometimes caused aluminum-induced adynamic bone disease. This condition is less common with the use of calcium salts to decrease phosphate absorption, but low-turnover bone disease still occurs, particularly in patients in whom secondary hyperparathyroidism has been reversed. Osteomalacia can occur in patients receiving dialysis if they have an inadequate supply of vitamin D and calcium, but this is unusual.

Renal osteodystrophy causes growth retardation and skeletal deformities in children, and both children and adults have bone pain and muscle weakness. Soft issue calcifications are particularly dangerous when they occur in blood vessels and lead to ischemia and gangrene. Calcification may be the result of a high calcium-phosphorus product and of vessel wall changes secondary to renal failure or direct effects of PTH. To prevent calcification, it is important to avoid a high serum calcium-phosphorus ion product and to minimize secondary hyperparathyroidism.

Diagnosis

The diagnosis of a specific form of renal osteodystrophy can often be made on biochemical grounds. Levels of PTH are high in patients with severe secondary hyperparathyroidism. Plasma aluminum levels may be elevated in patients with aluminum-induced osteodystrophy but do not necessarily reflect the stores in bone. Deferoxamine chelates aluminum, and this agent can be used to measure the body burden and to treat aluminum overload. The interpretation of biochemical markers of bone turnover is difficult in renal disease because their clearance may be altered.

A bone biopsy may clarify the pathogenesis of renal osteodystrophy. Using double tetracycline labeling, it is possible to determine whether mineralization is impaired. Sections that have not been decalcified show the extent of osteoid seams and resorption surfaces. Aluminum can be identified by special stains. Amyloid deposits, which consist largely of β2-microglobulin, may be seen in the bone. Amyloidosis of the bone is associated with cystic lesions but not necessarily with bone pain.
Therapy

The treatment of renal osteodystrophy can be highly successful if it is correctly focused on specific pathogenetic mechanisms. The goal should be to maintain normal serum calcium and phosphorus levels and minimize exposure to aluminum. Phosphate restriction should be instituted relatively early in renal failure, but diets low in phosphorus are difficult to achieve. Therefore, after the filtration rate is reduced below 25% of normal, it is usually necessary to administer phosphate-binding salts such as calcium carbonate, calcium acetate, or calcium citrate. The ability of citrate to increase aluminum absorption is a concern. Correction of acidosis is also important in preventing bone disease.

Early in renal failure, modest supplementation with vitamin D may be sufficient to maintain 1,25(OH)₂D levels, but eventually calcitriol itself should be administered. Low doses (0.25 to 0.5 µg/day) are well tolerated, but higher doses can lead to hypercalcemia and hypercalciuria. In severe secondary hyper-parathyroidism, low doses are often not sufficient to suppress PTH secretion and intravenous calcitriol may be used.

In some cases, parathyroidectomy may be required. Persistent hypercalcemia in patients with renal failure, intractable pruritus, extracellular calcifications, and severe skeletal lesions are all indications for surgery. However, parathyroidectomy should be avoided in patients with aluminum-induced bone disease because symptoms may be worsened. Hemodialysis patients have a high incidence of fractures of the distal forearm and ribs.

Renal transplantation corrects many of the biochemical disturbances that lead to renal osteodystrophy, but bone disease may progress. Usually, secondary hyperparathyroidism slowly resolves, but patients with persistent hypercalcemia or autonomy of the parathyroid glands (tertiary hyperparathyroidism) may require surgery. A major concern, particularly in older patients, is progressive osteoporosis; glucocorticoids and immunosuppressants worsen bone loss in these patients. Finally, osteonecrosis or avascular necrosis, particularly of the proximal femur, is common after renal transplantation.
PAGET’S DISEASE

Paget's disease may affect as many as 3% of adults older than 40 years of age; it is often asymptomatic and usually progresses slowly.

Pathogenesis

The primary abnormality in Paget's disease is the localized, uncontrolled formation of large, highly active osteoclasts. The initial lesion is an increase in bone resorption. The response to this resorption, particularly in bones that are subject to mechanical force, is an intense but chaotic increase in osteoblastic activity. The characteristic histologic appearance is of focal lesions with many giant osteoclasts and active osteoblasts. The bone that forms in the lesions is disorganized and has a mosaic pattern with loss of the usual lamellar structure. The marrow shows a pattern of fibrosis and increased vascularity.

The concept of a viral origin of Paget's disease is based on the finding of nuclear inclusion bodies in osteoclasts and the detection of viral transcripts in hematopoietic cells from patients with the disease. Several paramyxoviruses have been suggested, including measles and canine distemper virus. Expression of the measles virus nucleocapsid protein gene in osteoclast precursor cells was found to enhance the ability of these cells to form osteoclasts. However, further work is needed to establish pathogenetic links between viral sequences and production of abnormal osteoclasts in this condition. Pagetic osteoclasts differ from normal osteoclasts not only in their greater size and the presence of viral inclusions but also because they express IL-6, which may play a role in pathogenesis. Expression of resorption stimulators by osteoblast-lineage cells is probably also involved in the development of Paget's disease.

Pagetic bone often distributes in a heterogeneous pattern throughout the skeleton. In addition, when lesions appear, they remain stationary over many years. Hence, abnormalities only in osteoclast precursor cells, which are believed to circulate freely in blood, cannot explain all the pathology that is seen in this condition. A description of the pathogenetic mechanisms in Paget’s disease must elucidate how the spotty distribution of this condition develops. Some insight into this puzzle was provided by studies of bone marrow stromal cells from pagetic lesions. These cells, believed to be stationary in bone, have been shown to produce enhanced RANKL mRNA compared with marrow stromal cells from uninvolved areas in the same patient.

There may also be a genetic component in Paget's disease. As many as 15% to 30% of patients have a positive family history, and first-degree relatives of patients with Paget's disease have a sevenfold greater relative risk of having the disorder than individuals with no affected relatives. There is also ethnic and geographic clustering. The incidence is high in some areas of northern Europe, particularly in northern England, but low in Norway and Sweden. A rare related illness, familial expansile osteolysis, which is seen in young adults as a generalized increase in osteoclastic activity and bone turnover, has been found to result from an activating mutation of RANK, the receptor for RANKL.
Clinical Features

Paget's disease affects men and women almost equally, but men tend to be more symptomatic. The disease is usually not clinically apparent until age 50 to 60 years. It usually progresses slowly and does not develop in new sites. Many different bones can be affected, and the lesions can vary from single, monostotic lesions to involvement of almost the entire skeleton. The pelvis, femur, spine, skull, and tibia are most commonly involved, whereas hands and feet are rarely affected.

Paget's disease is often discovered in asymptomatic patients because of an elevated serum alkaline phosphatase measurement obtained on routine screening or because of a radiograph taken for an unrelated problem. The most common symptom is bone pain at the site of pagetic involvement. Pain also commonly occurs in adjacent joints as a result of secondary degenerative arthritis. Bowing of the legs is common (Fig. 27-21), and pathologic fractures can occur. Vertebral involvement can cause kyphosis and compression of the spinal cord. Neural changes can also result from vascular steal because of the high blood flow to the lesion. The most common consequence of Paget's disease of the skull is hearing loss, which can be both conductive and neurosensory. Extensive involvement of the base of the skull can produce basilar impression and, rarely, brain stem compression (Fig. 27-22). Facial and skull deformities and dental problems are common.

Figure 27-21 Paget's disease of the tibia. Note the bowing, marked irregularity of the anterior cortex and the flame-shaped lytic lesion of the posterior cortex. (Courtesy of Dr. Ethel S. Siris.)

The incidence of osteosarcoma is increased but is less than 1%. When osteosarcoma does occur, it is highly malignant. Most patients do not live longer than 1 to 3 years. Fibrosarcomas, chondrosarcomas, and benign giant cell tumors are also occasionally seen. The giant cell tumors, termed reparative granulomas, may represent an extension of pagetic tissue outside the skeleton. These tumors are sensitive to antipagetic therapy and may also respond to glucocorticoids.

Patients with Paget's disease may have an increased incidence of primary hyperparathyroidism. Angioid streaks are often seen in the fundus. Pseudogout, gout, and osteoarthritis occur. Patients with heart disease may show worsening of heart failure, which has been attributed to the increase in blood flow in pagetic lesions.
Diagnosis

As noted previously, the diagnosis of Paget's disease may be made by the finding of an elevated alkaline phosphatase concentration or after a routine radiograph. In older persons with deformities or bone pain, the diagnosis should be considered and a careful family history and review of the musculoskeletal system by both history and physical examination should be obtained. A bone scan should be carried out to localize possible pagetic sites. Positive scans do not necessarily indicate Paget's disease, and radiographs should be obtained to confirm that Paget's disease is the cause of the increased uptake. Rarely, pagetic sites in bone are not evident on bone scan because there is a minimal formation response in the lesion. Such pagetic lesions in the skull are termed osteoporosis circumscripta.

An audiogram should be obtained in patients with involvement of the petrous bone or in those with complaints of hearing loss. Because of the possible increased incidence of hyperparathyroidism, ionized calcium levels should be measured in the initial work-up. Monostotic Paget's disease, particularly in the vertebrae, may be difficult to distinguish from metastatic disease. In addition, some patients with vertebral disease may have impingement on the spinal canal. In these individuals, the area should be examined by CT or MRI. Bone biopsies can be useful in atypical cases. An ordinary aspiration biopsy sometimes yields the giant osteoclasts that are pathognomonic of Paget's disease. Samples of bone that show the irregular marble bone pattern can also be diagnostic.

After the initial evaluation has been completed, the patient can usually be monitored biochemically by serial measurements of total or bone-specific alkaline phosphatase and a marker of bone resorption. Urinary hydroxyproline measurements may be used, as may serum or urinary levels of type I collagen breakdown products such as C-telopeptide or N-telopeptide or measurements of pyridinolone or deoxypyridinolone in the urine. Although 24-hour urine measurements of hydroxyproline are often recommended, a morning fasting measurement avoids the dual problems of collecting a 24-hour sample and maintaining a diet free of gelatin products.
Therapy

In the past, patients with Paget’s disease were often simply observed until symptoms were clear-cut or until there was evidence of progression in critical areas of the skeleton. With the newer bisphosphonates, treatment is instituted earlier. Pamidronate is available for IV use. The recommended dose for Paget's disease is 30 mg infused over 4 hours on 3 consecutive days for a total of 90 mg. Patients with more extensive disease may require retreatment and should be evaluated by measuring levels of bone turnover markers at regular intervals. IV pamidronate is generally safe, although transient fever and a transient increase in bone pain may occur. Rare idiosyncratic reactions include uveitis.

Other bisphosphonates are given orally. Alendronate (40 mg/day for 6 months), risedronate (30 mg/day for 2 months), and tiludronate (400 mg/day for 3 months) are approved in the United States. As with pamidronate, patients may require repeated therapy with oral bisphosphonates after a drug-free period that varies with each agent (6 months for alendronate, 2 months for risedronate, and 3 months for tiludronate). Oral bisphosphonates need to be taken in a manner that minimizes the development of esophagitis or interactions with food and other therapeutics in the stomach. These treatments have largely replaced therapy with calcitonin, etidronate, or plicamycin. All of these agents act by inhibiting osteoclastic activity, and the earliest indication of therapeutic response is a drop in resorption markers followed by a decrease in formation markers.

The indications for treatment are pain that can be attributed to Paget’s disease and deformities that might produce neurologic changes or are likely to lead to fracture, such as the osteolytic flame lesion or blade of grass lesion in weight-bearing bones (see Fig. 27-20). Hearing loss may be an indication for therapy, although most patients do not show major improvement after treatment.

Patients with heart disease and extensive Paget's disease should be treated in the hope that decreased pagetic activity will improve management. With the advent of safe and effective therapy, patients with mild to moderate disease, particularly those with the potential for complications (i.e., those with lesions in weight-bearing bone, the vertebral bodies, or the base of the skull), are being treated before symptoms develop. Early treatment is logical in young patients with Paget's disease because it is hoped that therapy may prevent progression. However, proof of this hypothesis is not conclusive.

Many patients with Paget's disease have pain associated with joint damage that does not respond to antipagetic therapy. These patients may respond to anti-inflammatory drugs. If osteoarthritis is advanced, knee and hip replacement may be appropriate but biochemical remission should be obtained before surgery.

A high calcium intake may be useful in Paget's disease. Bisphosphonate therapy can lower the serum calcium level and cause secondary hyperparathyroidism, which is probably not advantageous; increased calcium intake may prevent this development. Moreover, calcium loading can produce an increase in endogenous calcitonin secretion that may have beneficial effects. It is also important to monitor serum 25(OH)D levels periodically in patients with Paget's disease who are to be treated with bisphosphonates, as osteomalacia is not uncommon in the elderly population who are at risk for this condition and low serum vitamin D levels may exacerbate the potential of bisphosphonates to cause hypocalcemia. Urinary calcium should be checked before calcium or vitamin D supplementation is given because an increase in the incidence of renal stones has been reported in pagetic patients.
Hereditary Hyperphosphatasia

Although hereditary hyperphosphatasia has been termed juvenile Paget's disease, it involves all of the skeleton and develops in infants. The serum alkaline phosphatase levels are very high. There are severe bone deformities, and the histologic appearance resembles that of Paget's disease with high bone turnover, although the osteoclasts are not enlarged. Treatment with bisphosphonates or calcitonin may be effective in reducing bone turnover and improving bone lesions. A new familial form with expansile long bone lesions has been described.
OSTEOGENESIS IMPERFECTA

Osteogenesis imperfecta, or brittle bone disease, is a heterogeneous, congenital disorder in which increased bone fragility leads to fractures and deformity. It ranges in severity from a lethal perinatal form to a mild disorder that results only in increased fractures.

Pathogenesis

Most patients with osteogenesis imperfecta have defects in the genes for type I collagen. Bones, ligaments, skin, sclerae, and teeth are affected. The incidence of osteogenesis imperfecta is estimated to be 1 in 200,000 to 500,000. The heterogeneity of the features is caused by the variety of genetic defects, although phenotypic variation occurs even with the same genetic abnormality (see Table 27-5). The more severe forms, type II and type III, involve mutations at the helical portion of the collagen molecule that prevent normal assembly and produce unstable triple helices. Point mutations in this portion of the collagen gene can be associated with mild disease (type IV).

Type I osteogenesis imperfecta differs from the other forms in that there is usually a deletion of one allele of the 1(I) procollagen gene, resulting in decreased collagen production but a normal molecular structure.

Bone biopsies show decreased cortical width and trabecular bone volume, increased turnover, and decreased bone formation in patients with type I disease. The disorder in a subgroup of patients with low turnover and ligamentous calcifications has been designated type V osteoporosis imperfecta.
Classification and Clinical Features

The classification devised by Sillence and modified by Byers is summarized in Table 27-5. In addition to the bone involvement, there may be ligament laxity, joint hypermobility, and easy bruising. Dentin formation is often abnormal, and the teeth are fragile and discolored. Blue sclerae are a variable manifestation and do not correlate with severity. Because of the thoracic deformities, patients with severe manifestations are predisposed to pulmonary infections and usually have a shortened life span. Intelligence is not affected, and individuals with marked deformities can be highly productive if appropriate conditions are provided.
Diagnosis

In patients with moderate to severe disease, the clinical features make the diagnosis relatively straightforward; in patients with the milder forms, however, the diagnosis may be missed. In children without deformities, multiple fractures are usually attributed to trauma; in infants, the presence of such fractures may lead to an accusation of parental abuse. In the absence of typical clinical features, the diagnosis can be made only biochemically. Culture of fibroblasts from skin biopsies and analysis of the collagen by gel electrophoresis can point to a defect, and the techniques of molecular biology can identify the mutation more specifically. This analysis is useful for families because specific deoxyribonucleic acid (DNA) polymorphisms may allow prenatal diagnosis if the mutation has already been identified in other affected family members.

In children and adolescents with multiple fractures but no deformity, measurements of bone density and turnover may point toward the diagnosis. In the type I disorder, both bone density and serum type I procollagen peptide levels are likely to be decreased. However, because excretion of collagen cross-links is increased in most types of osteogenesis imperfecta,
Therapy

Antiresorptive therapy with IV pamidronate has been shown to decrease fractures in children with severe osteogenesis imperfecta, even before 3 years of age. Supportive treatment is important. The Osteogenesis Imperfecta Foundation works with patients and families to improve the quality of life. Orthopedic and rehabilitation services can be helpful in dealing with deformities. Genetic counseling and prenatal diagnosis, including ultrasound examination and testing for informative DNA polymorphisms, are important for the family. Gene therapy is being explored.
Other Connective Tissue Disorders Affecting the Skeleton

Other inherited disorders of connective tissue with impairment of skeletal development or increased bone fragility include Ehlers-Danlos syndrome, Menkes' disease, lysinuric protein intolerance, and homocystinuria. In these disorders, abnormalities of collagen cross-linking can affect bone and other connective tissues. In Ehlers-Danlos syndrome the cross-linking enzyme lysyl oxidase is deficient, and in Menkes' disease copper deficiency impairs the function of the enzyme. Lysinuric protein intolerance and homocystinuria probably also impair cross-linking of collagen.
OSTEOPETROSIS

Osteopetrosis, or marble bone disease, is a heterogeneous group of disorders characterized by a generalized increase in bone density caused by defective osteoclastic bone resorption. These syndromes are part of a larger group of sclerosing bone dysplasias (see later) that vary in severity and sites of skeletal involvement. Most cases are inherited, but some are acquired.

Osteopetrotic rats and mice with specific defects in osteoclasts have been found as a result of spontaneous mutations and developed using gene knockout techniques. The op/op mouse has a genetic defect in the production of M-CSF that results in failure of osteoclast formation and can be corrected by treatment with M-CSF. Knockouts of the proto-oncogenes c-src and c-fos; of TRAF-6, a mediator of RANK action; and of the critical osteoclast protease cathepsin K also cause osteopetrosis. Similar mutations may cause human osteopetrosis, but to date only defects in the vascular proton pump and in carbonic anhydrase, which affects acidification by osteoclasts, have been identified.

Infantile Osteopetrosis

Infantile osteopetrosis is a rare, autosomal recessive disorder in which failure to resorb bone and calcified metaphyseal cartilage causes near obliteration of the marrow spaces. Extramedullary hematopoiesis occurs in the liver and spleen. The cranial nerve foramina do not form normally, causing optic atrophy and other cranial nerve defects. The bones, although dense, are brittle, and pathologic fractures can occur. The impaired function of the hematopoietic system causes death in the first decade from hemorrhage or infection. On the basis of studies in animal models in which transfer of hematopoietic tissue resulted in cure, patients have been treated with total body radiation and grafting of marrow from human leukocyte antigen-identical donors. Some of these patients have been shown to have genetic defects of subunits of the vascular-type H^+-adenosine triphosphatase that acts as a proton pump in osteoclasts, but other patients show no abnormalities in these genes.
Carbonic Anhydrase II Deficiency

Carbonic anhydrase II deficiency, a nonlethal autosomal recessive disorder, is associated with a complete deficiency of the type II carbonic anhydrase that provides carbonic acid for hydrogen ion secretion by osteoclasts and by the distal tubules. Hence, osteopetrosis is accompanied by renal tubular acidosis that may involve both distal and proximal lesions. Affected individuals are shorter than their siblings and may have calcification of the basal ganglia.
Autosomal Dominant Osteopetrosis

Two forms of autosomal dominant osteopetrosis (ADO) have been described. ADO type II is the classical Albers-Schönberg disease in which there is generalized osteosclerosis with thickening of the vertebral end plates (sandwich vertebrae) and bone within bone in the pelvis. ADO type I is characterized by generalized osteosclerosis and thickening of the cranial vault.

In ADO type I, the number of osteoclasts is reduced; in ADO type II, the osteoclasts are large but apparently nonfunctional. Interestingly, bone resorption can be stimulated by triiodothyronine in patients with both types. The term benign osteopetrosis has been used to describe these disorders, but fractures, orthopedic problems, and cranial nerve involvement do occur, particularly in ADO type II.
FIBROUS DYSPLASIA

Clinical Features

Fibrous dysplasia is characterized by expanding lesions within the bone that contain both fibroblastic and osteoblastic elements. The disorder can occur as a monostotic lesion without any associated abnormalities or in a polyostotic form, which may occur as part of the McCune-Albright syndrome, associated with functional abnormalities of one or more endocrine glands and irregular hyperpigmented macules called café-au-lait spots.

The most common endocrine manifestation is precocious puberty, particularly in girls. The molecular defect in the McCune-Albright syndrome is somatic mosaicism for an activating mutation of the Gs subunit of the nucleotide-binding regulatory protein that couples receptors to adenylyl cyclase. Similar defects have been found in bone lesions in the absence of McCune-Albright syndrome. Local production of PTHrP may also be involved in pathogenesis. The bone lesions may show increased expression of the c-fos oncogene.
Diagnosis

Monostotic fibrous dysplasia is usually diagnosed in the second or third decade of life as an expanding bone lesion that can cause fracture, deformity, or nerve entrapment. Sarcomatous degeneration can occur. In fibrous dysplasia, any skeletal site can be affected but the femur, tibia, ribs, and face are most often involved. Histologically, the lesions contain many spindle-shaped fibroblasts and islands of woven bone. Bone lesions can worsen during pregnancy, and estrogen receptors have been identified in the bone lesions of patients with McCune-Albright syndrome.
Therapy

The course of both monostotic and polyostotic fibrous dysplasia is variable. Patients who show progression, nerve compression, or pathologic fractures may require surgery. Careful assessment of the endocrine system is critical in the McCune-Albright syndrome because early intervention can prevent irreversible changes resulting from precocious puberty. The bone lesions may respond to bisphosphonate. \[356\] \[357\] Treatment with 1,25(OH)\(_2\)D may reduce PTHrP production and decrease activity of the bone lesions. \[353\]
SCLEROSING BONE DYSPLASIAS

A number of rare, congenital skeletal disorders cause irregular bone structure and varying degrees of sclerosis. Acquired osteosclerosis related to increased bone formation occurs in fluorosis and in rare cases of hepatitis C infection, and a localized form is seen in certain metastatic malignancies, particularly cancers of the prostate and breast. Osteosclerosis with increased bone formation occurs in mice overexpressing the transcription factor Fra-1.

Pyknodysostosis

Pyknodysostosis is an autosomal recessive disorder characterized by short stature, a large cranium, and small facies. Unlike patients with osteopetrosis, these patients do not demonstrate loss of the marrow cavity and are not anemic. Histologically, trabecular bone volume is increased despite an increase in the number of osteoclasts, suggesting a defect in osteoclast function. This defect has been attributed to a mutation of the cathepsin K gene.
Progressive Diaphyseal Dysplasia

Known as Camurati-Engelmann disease, progressive diaphyseal dysplasia consists of patchy thickening of the bone on both periosteal and endosteal surfaces and is associated with gait abnormalities and muscle wasting. The histologic picture is one of increased bone formation rather than decreased resorption. The disease can be inherited as an autosomal dominant disorder. The gene has been localized to chromosome 19q13, and the disorder may be due to a defect in the latency-associated peptide of TGF-1. Glucocorticoid therapy relieves bone pain and reverses the histologic abnormalities in some cases.
Endosteal Hyperostosis

The term van Buchem's disease is applied to both severe and mild forms of hyperostosis. The disorder begins in infancy, and progressive enlargement of the jaw commences at puberty. Some subjects have cranial nerve deficits caused by impingement of the foramina. This disorder may be related to sclerosteosis, in which there is also syndactyly. Subjects with both disorders are usually tall and heavy. Both disorders have been localized to chromosome 17q12-q21.
Other Disorders

Unusual disorders that affect bone structure include (1) osteopoikilosis, which is characterized by many small foci of sclerosis in cancellous bone; (2) osteopathia striata, in which there are linear striations of the ends of the long bones and associated cranial sclerosis; and (3) melorheostosis, in which patchy areas of hyperostosis affect both the cortex and the medullary canal, sometimes associated with thickening of the overlying skin. Mixed pictures involving several different forms of sclerosis have been described.

Axial osteomalacia and fibrogenesis imperfecta ossium are rare disorders characterized by a coarse and dense appearance of trabecular bone. These disorders may be forms of osteomalacia because they show defective mineralization, presumably because of a defect in the matrix produced by osteoblasts.

Pachydermoperiostosis is a disorder characterized by periostitis, sclerosis of the distal portions of the tubular bones, and thickening of the skin. It may be inherited as an autosomal recessive disorder. It is also termed primary hypertrophic osteoarthropathy and must be distinguished from the secondary hypertrophic osteoarthropathy that occurs in patients with chronic pulmonary disease.

Infantile cortical hyperostosis is characterized by overgrowth of cortical bone. The similarity of this disorder to the changes seen after prolonged infusion of PGE in neonates has suggested a possible cause, and the disorder has been treated with glucocorticoids and nonsteroidal anti-inflammatory drugs.
EXTRASKELETAL CALCIFICATION AND OSSIFICATION

Mineral deposition in soft tissues is a common consequence of tissue damage and of a local elevation of the extracellular calcium-phosphate product. Ectopic or heterotopic bone formation can occur at sites of injury or surgical trauma and may be related to the presence of an inductive protein matrix. This form of bone induction stimulated the search for BMPs. The frequency of heterotopic ossification after hip replacement or spinal cord injury can be decreased by treatment with nonsteroidal anti-inflammatory drugs that inhibit prostaglandin synthesis or by local irradiation. Extensive subcutaneous calcium deposition may be crippling in inflammatory disorders such as dermatomyositis.

The term myositis ossificans is used when bone formation occurs in traumatized muscle, but similar masses can be formed in tendon, ligaments, joint capsules, and fascia without trauma.

Tumoral Calcinosis

Primary tumoral calcinosis is an inherited disorder characterized by periarticular calcification and hyperphosphatemia. It is associated with defective phosphate transport and sometimes with excessive activity of renal 1-hydroxylase. The primary disorder must be differentiated from secondary tumoral calcinosis, which occurs in association with renal failure and with hypercalcemic disorders.

Treatment by phosphate depletion using phosphate-binding antacids or diuretics, parathyroidectomy, and dialysis has been attempted.
Fibrodysplasia Ossificans Progressiva

This rare congenital disorder is most often sporadic but can be transmitted as an autosomal dominant disorder. Characteristic short phalanges and soft tissue swelling can be detected at birth. Abnormal regulation of a BMP has been implicated, and linkage analysis has localized the abnormality to chromosome 4q27. Painful, tender lesions caused by true ectopic bone can develop in connective tissues. Progressive deformities may include scoliosis and ankylosis of the spine and rib cage. There is no known therapy.

Progressive osseous heteroplasia is another genetic disorder in which ossification occurs largely in the skin.
SKELETAL ABNORMALITIES ASSOCIATED WITH LDL RECEPTOR-RELATED PROTEIN 5 (LRP5)

Recently a gene defect was identified that produced a rare form of osteoporosis associated with ocular changes, termed the osteoporosis-pseudoglioma syndrome.\[385\] The mutation is in the LDL receptor-related protein 5 (LRP5), a receptor that interacts with the frizzled receptor to mediate signaling by members of the wnt family. In addition, studies of a remarkable family who have unusually high bone mass inherited as an autosomal dominant trait, but who are otherwise phenotypically normal, also show an abnormality in the LRP5.\[386\] In this case, the mutation appears to be one that results in activation of LRP5 and hence could enhance signaling. These new observations may indicate another important pathway in regulation of the skeleton, with relevance to the pathogenesis and/or treatment of osteoporosis.
References


100. Ralston SH. The genetics of osteoporosis. JQAM 1997; 90:247251.


References

162. Johansson AG, Eriksson EN, Lindh E, et al. Reduced serum levels of the growth hormone dependent insulin-like growth factor binding protein and a negative bone balance at the level of individual remodeling units in idiopathic osteoporosis in men. J Clin Endocrinol Metab 1997; 82:27652798.


References


Chapter 28 - Kidney Stones

Rebeca D. Monk
David A. Bushinsky

Nephrolithiasis is a common disorder with an incidence greater than one case per 1000 patients per year. The incidence peaks in the third and fourth decades. The prevalence in industrialized nations is close to 10% and increases with age until approximately age 70 years. In general, stones may be composed of calcium oxalate, calcium phosphate, uric acid, magnesium ammonium phosphate (struvite), or cystine, or a combination thereof. A variety of pathogenetic mechanisms determine the type of stone formed.

Stones tend to localize in the renal tubules and collecting system but are also commonly found within the ureters and bladder. Nephrolithiasis rarely results in renal insufficiency or life-threatening illness but is responsible for substantial morbidity. The severe pain of renal colic can lead to frequent hospitalization, shock wave lithotripsy, or invasive surgical procedures. Insight into the mechanisms involved in stone formation can help direct appropriate therapy, which is known to decrease the incidence of stone disease and its associated morbidity.
EPIDEMIOLOGY OF STONE FORMATION

Numerous factors determine the prevalence of stones, including sex, age, race, and geographic distribution. Nephrolithiasis is more common in men than women at a ratio of 2:1 to 4:1. In the United States, blacks, Hispanic Americans, and Asian Americans are much less likely to have stones than whites. Geography also appears to influence stone formation in the United States, with a decreasing prevalence from south to north and, to some degree, from east to west.

The greater exposure to sunlight in the southeastern United States may be responsible for the increased rates of nephrolithiasis in that area. Sun exposure can lead to more concentrated urine by increasing insensible fluid losses due to sweating; in addition, it may result in intestinal calcium absorption and urinary calcium excretion by stimulating vitamin D production.

Geographic location can also influence the type of stone formed. Uric acid stones, for example, predominate in Mediterranean and Middle Eastern countries, where they constitute up to 75% of all the stones formed. In the United States, however, fewer than 10% are pure uric acid stones, and more than 70% of stones formed are composed of calcium and an associated anion. Less common are magnesium ammonium phosphate (struvite or infection) stones, which account for about 10% to 15% of stones formed, and cystine stones, which are due to an autosomal recessive disorder and constitute only about 1% of all stones formed. (Table 28-1)

---

[1][2][3][4][5][6][7][8][9]
PATHOGENESIS OF STONE FORMATION

Kidney stones form when urine becomes oversaturated with respect to the specific components of the stone. Saturation is dependent on chemical free ion activities of the stone constituents. Factors that affect chemical free ion activity include urinary ion concentration, pH, and complexing with other substances. For example, an increase in the quantity of urinary calcium or a decrease in urine volume increases the free ion activity of calcium ions in the urine. Urinary pH can also modify chemical free ion activity. A low urinary pH increases the free ion activity of uric acid ions but decreases the activity of calcium and phosphate ions. Citrate combines with calcium ions to form soluble complexes and can thereby decrease their free ion activity. When the chemical free ion activities are increased, the urine becomes oversaturated and new stones may form or established stones grow. In the setting of decreased free ion activity, urine becomes undersaturated and stones do not grow and may even dissolve. The equilibrium solubility product is the degree of chemical free ion activity of stone components in a solution in which the stone neither grows nor dissolves.

Formation of stones occurs through either homogeneous or heterogeneous nucleation:

1. Homogeneous. Progressive oversaturation can eventually result in formation of small clusters of ions; these clusters grow in size to form a permanent solid phase.

2. Heterogeneous. This category refers to crystal formation on the surface of a different crystal type or on other foreign substances, such as cells. In vivo, this type of nucleation is more common than homogeneous nucleation because crystals form at a lower level of oversaturation in the presence of a solid phase.

### TABLE 28-1 -- Percentage of Patients with Various Stone Types in the United States

<table>
<thead>
<tr>
<th>Type</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mixed calcium oxalate and calcium phosphate</td>
<td>37</td>
</tr>
<tr>
<td>Calcium oxalate</td>
<td>26</td>
</tr>
<tr>
<td>Calcium phosphate</td>
<td>7</td>
</tr>
<tr>
<td>Uric acid</td>
<td>5</td>
</tr>
<tr>
<td>Struvite</td>
<td>22</td>
</tr>
<tr>
<td>Cystine</td>
<td>2</td>
</tr>
</tbody>
</table>


The crystals must then aggregate into clinically significant stones, a process that takes longer than the passage of urine through the renal tubules. For stone formation to occur before the crystals are swept away in the urine, they must adhere to the renal tubular cells in order to allow more time for growth. This anchoring of crystals to tubular cells is a process that is undergoing careful study.

An important factor in the development of kidney stones may be the absence of adequate levels or activity of crystallization inhibitors in patients with stones. Uropontin, pyrophosphate, and nephrocalcin are endogenously produced substances that have been shown to inhibit calcium crystallization. Differences in the amount or activity of inhibitors may account for the variability in stone formation among people with similar degrees of urinary oversaturation. The influence of such inhibitors on stone formation is also an area of ongoing research.

Clinically, most physicians evaluate the lithogenic potential of the urine from stone formers by measuring the rate of excretion of the principal stone-forming elements in mass per unit time (e.g., milligrams per 24 hours). It is clear, however, that the lithogenic potential of urine is better determined by the degree of oversaturation. Computer programs that calculate saturation from concentrations of various elements in the urine are now available and more accurately determine the risk of stone formation. Any calculation of mean saturation under-estimates the maximum oversaturation because of hourly variations in water and solute excretion throughout the day.
PATHOGENESIS OF IDIOPATHIC HYPERCALCIURIA

Idiopathic hypercalciuria (IH) is the most common cause of calcium-containing kidney stones. IH is defined as excessive urinary calcium excretion (hypercalciuria) in the setting of normocalcemia and in the absence of secondary causes of hypercalciuria. The disorder is familial, was initially thought to exhibit an autosomal dominant pattern of inheritance, but may be polygenic. The mechanism by which IH leads to hypercalciuria is not known. It has been postulated that IH may comprise three distinct disorders:

1. Excessive intestinal calcium absorption.
2. Decreased renal tubular calcium reabsorption.
3. Enhanced bone demineralization.

Recent observations have led many to believe that IH consists of a spectrum of these disorders, with considerable overlap among these potential mechanisms.

Calcium Homeostasis

Urinary calcium homeostasis is regulated by the gastrointestinal (GI) tract, the kidneys and bone, and the hormones parathyroid hormone (PTH) and 1,25-dihydroxyvitamin D₃ (1,25(OH)₂ D₃). Approximately 99% of the calcium in the body is contained within the bone mineral. Daily bone resorption and bone formation, which in healthy, nonpregnant adults should be equal, allow less than 1% of bone calcium to be exchanged with that in the extracellular fluid.

Both PTH and 1,25(OH)₂ D₃, at high concentrations, stimulate release of calcium from the bone mineral through osteoclast-mediated bone resorption. Net calcium influx into the extracellular fluid is achieved principally by absorption from the GI tract, which occurs through 1,25(OH)₂ D₃-dependent and -independent mechanisms. Although PTH appears to have no direct effect on GI calcium absorption, increased levels of the hormone can stimulate production of 1,25(OH)₂ D₃, which in turn leads to enhanced absorption.

The roughly 60% of calcium in the extracellular fluid that is not protein-bound is freely filtered by the renal glomeruli. Approximately 80% to 85% of this amount is passively reabsorbed in the proximal tubule. Most of the remaining calcium is reabsorbed in the distal cortical tubules under PTH stimulation. Ultimately, these reabsorptive mechanisms result in a urinary calcium concentration that is less than 2% of the daily filtered load of calcium. Except during pregnancy and lactation, urinary calcium excretion equals intestinal calcium absorption in healthy adults.
Potential Mechanisms for Development of Idiopathic Hypercalciuria

Dysregulation of calcium flux at any of these sites may lead to hypercalciuria. For example, excessive calcium absorption by the GI tract would lead to a transient increase in the serum calcium. This increase in serum calcium would suppress secretion of PTH and 1,25(OH)\(_2\)D\(_3\), which, along with the increased filtered load of calcium to the kidneys, would result in hypercalciuria. Excessive 1,25(OH)\(_2\)D\(_3\) would have a similar effect on intestinal calcium absorption but would also result in influx of calcium into the extracellular fluid because of enhanced bone demineralization. The result would be hypercalciuria even in the setting of a low-calcium diet or an overnight fast.

If a primary defect in renal calcium reabsorption has led directly to hypercalciuria, there would be a fall in the serum calcium concentration that would stimulate production of PTH and 1,25(OH)\(_2\)D\(_3\). This production, in turn, would result in enhanced intestinal calcium absorption. With this mechanism, the renal loss of calcium would persist even with a low-calcium diet or overnight fast.

Hypercalciuria may also develop as a result of a renal defect in phosphorus reabsorption. The resultant hypophosphatemia would lead to enhanced 1,25(OH)\(_2\)D\(_3\) production, which would stimulate intestinal absorption of both phosphorus and calcium. The increased serum calcium would suppress PTH production. The increased filtered load of calcium in the setting of suppressed PTH would also lead to hypercalciuria. Enhanced bone demineralization would increase the serum calcium concentration, which in turn would suppress PTH production. The increase in the filtered load of calcium in this setting would result in hypercalciuria.

Thus, there are several potential mechanisms for hypercalciuria. Do human or animal data support one mechanism above all others? From a clinical therapeutic standpoint, is it worth differentiating among the various potential mechanisms in each patient with suspected IH?
Human Data

Larsen compiled the results of numerous calcium balance studies on patients with IH and normocalciuric control subjects and normalized the results for calcium intake. He found that intestinal calcium absorption was significantly higher in the subjects with IH.

Asplin and colleagues, also collecting data from published metabolic balance studies, compared net intestinal calcium absorption and urinary calcium excretion in hypercalciuric and normocalciuric adults. They also noted an increase in intestinal calcium absorption in subjects with IH but found that urinary excretion of calcium was increased to an even greater degree, placing many of these patients in net negative calcium balance.

Although these data confirm that enhanced intestinal absorption of calcium probably plays a role in the pathogenesis of IH, the investigators could not clarify whether this is the primary defect or is secondary to another lesion, such as a primary defect in renal handling of calcium. Others suggested that the increase in intestinal calcium absorption, in combination with excessive renal calcium excretion, indicated a more generalized defect in calcium homeostasis. Nonetheless, the finding of enhanced calcium absorption makes enhanced bone resorption an unlikely primary mechanism of IH, as the increase in serum calcium concentration resulting from bone resorption would suppress 1,25(OH)\(_2\)D\(_3\)-mediated intestinal calcium absorption.

In most published studies, patients with IH have higher serum levels of 1,25(OH)\(_2\)D\(_3\) than normocalciuric control subjects. Kaplan and colleagues determined that 1,25(OH)\(_2\)D\(_3\) levels were higher than control values in approximately one third of patients with IH and that intestinal calcium absorption was inappropriately high for the level of 1,25(OH)\(_2\)D\(_3\). These studies support either 1,25(OH)\(_2\)D\(_3\)-mediated intestinal calcium absorption or a primary defect in renal tubular calcium reabsorption as a primary mechanism for hypercalciuria in IH.

PTH levels in patients with IH have been reported as normal or slightly lower than those in controls. This finding argues against a reduction in renal tubular calcium reabsorption as the primary defect in IH because with this mechanism the hypercalciuria would lead to low serum calcium levels and stimulation of PTH secretion. This finding also does not support the hypothesis that elevated levels of PTH are the stimulus for the increased levels of serum 1,25(OH)\(_2\)D\(_3\) observed in many studies. It is, however, consistent with the other potential mechanisms for IH.

Bone mass in patients with IH has been assessed by a number of methods, including radiologic densitometry, quantitative computed tomography (CT), dual-energy x-ray absorptiometry, single-photon absorptiometry, and others. Studies of patients with IH have generally shown only a mild reduction of bone mineral density compared with values in controls. The studies were unable to reveal a unifying mechanism for the mild reduction in bone mineral density, but primary net bone resorption is unlikely because a much greater decrease in bone density would be expected in this setting. Altered 1,25(OH)\(_2\)D\(_3\) regulation would be consistent with this finding because the effects of 1,25(OH)\(_2\)D\(_3\) on bone resorption would be mitigated by the increased intestinal calcium absorption stimulated by the hormone.

Until recently, it was considered necessary to determine whether a patient with IH tended to have excessive GI calcium absorption (absorptive hypercalciuria) or excessive renal excretion (renal leak). Patients with excessive renal calcium excretion were prescribed thiazide diuretics, and those thought to have a predominantly absorptive defect were prescribed a low-calcium diet. Coe and colleagues undermined the validity of this approach in a study in which 24 patients with IH and 9 control subjects were given a low-calcium diet (2 mg/kg per day) for more than 1 week. Urine and blood tests revealed normal serum calcium levels, a mild decrease in PTH levels in the patients with IH, but no difference in 1,25(OH)\(_2\)D\(_3\) levels.

The striking finding was that whereas all the normocalciuric subjects excreted less calcium than they ingested, 16 of the 24 subjects with IH had urinary calcium excretion that exceeded their intake. Thus, most of the patients with IH receiving a low-calcium diet were in net negative calcium balance. No clear demarcation was noted between the patients who tended to excrete excessive amounts of calcium and those who did not. Instead, there was a smooth continuum of urinary calcium excretion among patients with IH that appeared not to be influenced by calcemic hormones. These findings have rendered insignificant not only the need to clinically distinguish IH mechanisms in humans but also the prescription of a low-calcium diet in any of these patients, as the result may be a dangerous reduction in bone mineral density, especially in women.
Genetic Hypercalciuric Stone-Forming Rats

To explain more fully the mechanism of IH in humans, we have developed an animal model of this disorder. Through more than 54 generations of successive inbreeding of the most hypercalciuric progeny of hypercalciuric Sprague-Dawley rats, we have established a strain of rats that excrete abnormally large amounts of urinary calcium (Fig. 28-1). As in humans, the principal mechanism for the excessive calcium excretion in these rats appears to be an increase in intestinal calcium absorption. The increased absorption appears to be mediated not by an increase in the serum level of 1,25(OH)_{2}D_{3} but by an increase in the number of intestinal vitamin D receptors.

When these hypercalciuric rats were fed a diet very low in calcium, their urinary calcium excretion remained elevated compared with that of similarly treated control rats, indicating a defect in renal calcium reabsorption or an increase in bone resorption, or both. Bone from these hypercalciuric rats released more calcium than the bone of control rats when exposed to increasing amounts of 1,25(OH)_{2}D_{3} and the administration of a bisphosphonate to rats fed a low-calcium diet significantly reduced urinary calcium excretion. In addition, a primary defect in renal calcium reabsorption was observed during clearance studies. We have shown that, besides the intestine, both the bone and kidney of the hypercalciuric rats have an increased number of vitamin D receptors.

Thus, hypercalciuric rats appear to have a systemic abnormality in calcium homeostasis; they absorb more intestinal calcium, they resorb more bone, and they do not adequately reabsorb filtered calcium. Because each one of the hypercalciuric rats forms renal stones, we have described them as genetic hypercalciuric stone-forming (GHS) rats. These studies suggest that an increased number of vitamin D receptors may be the underlying mechanism for hypercalciuria in these rats and perhaps, by analogy, in humans. Clinical studies are ongoing to determine whether patients with hypercalciuria have an increased number of vitamin D receptors.
CLINICAL PRESENTATION

Kidney stones vary in clinical presentation from asymptomatic to large, obstructing staghorn calculi that can significantly impair renal function and lead to end-stage renal disease. The severity of stone disease depends on the pathogenetic factors contributing to the rate of stone formation in addition to the stone type, size, and location. In its most classic form, nephrolithiasis is manifested as renal colic. This discomfort of abrupt onset intensifies over time into an excruciating, severe flank pain that resolves only with stone passage or removal. The pain may migrate anteriorly along the abdomen and inferiorly to the groin, testicles, or labia majora as the stone moves toward the ureterovesical junction. Gross hematuria, urinary urgency and frequency, nausea, and vomiting may be present. Stones smaller than 5 mm are likely to pass spontaneously with hydration, whereas larger stones often necessitate urologic intervention.

Certain disorders can lead to small, diffuse renal parenchymal calcifications termed nephrocalcinosis. The calcifications, usually calcium phosphate or calcium oxalate, may be present in the cortex or medulla. Among the most common causes of stone-related nephrocalcinosis are primary hyperoxaluria and medullary sponge kidney.
METABOLIC EVALUATION OF STONE FORMERS

Although it is uniformly accepted that patients with multiple stones merit a thorough investigation into the cause of nephrolithiasis, evaluation of the patient with a single stone is controversial. This is probably due to the difficulty in determining the cost-to-benefit ratio of stone evaluations and wide differences in reported rates of stone recurrence.

In 1988, the National Institutes of Health convened a consensus conference to resolve such issues related to the prevention and treatment of kidney stones. The panel determined that all patients, even those with a single stone, should undergo at least a basic evaluation in order to rule out a systemic etiologic mechanism. Patients with an increase in number or size of stones within a certain time frame (metabolically active stones), all children, all non-calcium oxalate stone formers, and those in demographic groups not typically susceptible to stone formation (nonwhites) warrant a more complete metabolic evaluation (Fig. 28-2).
THE BASIC EVALUATION (Table 28-2)

History
In addition to the medical history typically obtained from new patients, the evaluation of the stone former includes a stone history and a thorough review of diet, fluid intake, and lifestyle. Specific laboratory studies and radiographic tests are also required.

Stone History
The stone history begins with a chronology of stone events: age of incidence of first stone, size and number of stones formed, frequency of passage, stone type if known, and whether the stones occur equally in both kidneys or unilaterally. Also helpful is a report of the patient's symptoms with each episode as well as the need for and response to surgical intervention. This information imparts not only the severity of the stone disease but also clues to the origin of the patient's nephrolithiasis. For example, nephrolithiasis that begins at a young age may be attributable to an inherited metabolic disorder such as primary hyperoxaluria or cystinuria. Large staghorn calculi that are difficult to eradicate and that recur despite

### TABLE 28-2 -- The Basic Evaluation

<table>
<thead>
<tr>
<th>History</th>
<th>Medical history</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stone history</td>
<td></td>
</tr>
<tr>
<td>Family history</td>
<td></td>
</tr>
<tr>
<td>Medications</td>
<td></td>
</tr>
<tr>
<td>Occupation and lifestyle</td>
<td></td>
</tr>
<tr>
<td>Diet and fluid intake</td>
<td></td>
</tr>
<tr>
<td>Physical examination</td>
<td></td>
</tr>
<tr>
<td>Laboratory tests</td>
<td></td>
</tr>
<tr>
<td>Urinalysis</td>
<td></td>
</tr>
<tr>
<td>Urate culture and sensitivity</td>
<td></td>
</tr>
<tr>
<td>Cystine screening</td>
<td></td>
</tr>
<tr>
<td>Blood tests</td>
<td>Sodium, potassium, chloride, bicarbonate</td>
</tr>
<tr>
<td></td>
<td>Calcium, phosphorus, uric acid, creatinine</td>
</tr>
<tr>
<td></td>
<td>Intact parathyroid hormone if calcium elevated</td>
</tr>
<tr>
<td>Stone analysis</td>
<td></td>
</tr>
<tr>
<td>Radiology</td>
<td>Kidneys, ureter, and bladder examination</td>
</tr>
<tr>
<td></td>
<td>Intravenous pyelography</td>
</tr>
<tr>
<td></td>
<td>Ultrasonography</td>
</tr>
<tr>
<td></td>
<td>Spiral computed tomography</td>
</tr>
</tbody>
</table>


### TABLE 28-3 -- Causes of Calcium Stone Formation

<table>
<thead>
<tr>
<th>Hypercalciuria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Idiopathic hypercalciuria</td>
</tr>
<tr>
<td>Hypercalcemic disorders</td>
</tr>
<tr>
<td>Primary hyperparathyroidism</td>
</tr>
<tr>
<td>Malignancy</td>
</tr>
<tr>
<td>Granulomatous diseases Sarcoid</td>
</tr>
<tr>
<td>Immobilization</td>
</tr>
<tr>
<td>Thyrotoxicosis</td>
</tr>
<tr>
<td>Milk-alkali syndrome</td>
</tr>
<tr>
<td>Medications (see Table 28-5)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hyperoxaluria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dietary hyperoxaluria (urine oxalate 4060 mg/day)</td>
</tr>
<tr>
<td>Enteric oxaluria (urine oxalate 60100 mg/day)</td>
</tr>
<tr>
<td>Malabsorptive disorders</td>
</tr>
<tr>
<td>Crohn's disease</td>
</tr>
<tr>
<td>Sprue (celiac disease)</td>
</tr>
<tr>
<td>Jejunostial bypass</td>
</tr>
<tr>
<td>Chronic pancreatitis</td>
</tr>
</tbody>
</table>
Biliary obstruction

Primary hyperoxaluria types 1 and 2 (oxalate 80-300 mg/day)

Hyperuricosuria (see Table 28-4)

Hypocitraturia

Metabolic acidosis

Hypokalemia

Exercise

Infection

Starvation

Hypomagnesemia

Renal tubular acidosis (distal, type 1)

Anatomic genitourinary tract abnormalities

Medullary sponge kidney

Tubular ectasia

Congenital megacalix


TABLE 28-4 – Factors Associated with Noncalcium Stone Formation

<table>
<thead>
<tr>
<th>Stone Type</th>
<th>Associated Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uric acid stones</td>
<td>Low urine volume, Inadequate fluid intake, Diarrhea, Malabsorptive disorders, Excessive insensible losses, Low urine pH (5.5), Diarrhea, Diet high in animal protein, Hyperuricosuria, Hyperuricemia, Gout, Intracellular to extracellular uric acid shift, Myeloproliferative disorders, Hemolytic anemia, Inborn errors of metabolism, Lesch-Nyhan syndrome, Glucose-6-phosphatase deficiency, Medications (see Table 28-5)</td>
</tr>
<tr>
<td>Struvite stones</td>
<td>Urease-producing bacteria, Proteus, Pseudomonas, Haemophilus, Yersinia, Ureaplasma, Klebsiella, Corynebacterium, Serratia, Citrobacter, Staphylococcus, Escherichia coli (not a urease producer), High urine pH (&gt;6.5), Indwelling urinary catheter, Neurogenic bladder, Cystine stones, Autosomal recessive trait, Excessive excretion of cystine, ornithine, lysine, and arginine, Low solubility of cystine (&lt;250 mg/L)</td>
</tr>
</tbody>
</table>


Frequent surgical intervention are more likely to be composed of struvite instead of calcium oxalate. Cystine stones are not crushed thoroughly with the use of lithotripsy, and alternative surgical modalities are generally required for stone removal. In patients who tend to form stones in only one kidney, the possibility of congenital abnormalities of that kidney, such as megacalix or medullary sponge kidney, should be explored.

Medical History

Systemic disorders that can contribute to nephrolithiasis are sought in the medical history. For example, any disorder that can result in hypercalcemia, such as sarcoidosis or certain malignancies, may also lead to hypercalcemia. A variety of GI disorders associated with malabsorption (e.g., sprue, Crohn's disease) can cause calcium oxalate nephrolithiasis on the basis of enteric hyperoxaluria. Patients with gout are more likely to have uric acid stones (Table 28-3 and Table 28-4).

Family History
A number of stone disorders are inherited, making the family history an important component of the basic evaluation. IH appears to be a familial disorder. Although the exact chromosomes and genes have not yet been identified, the pattern of inheritance may prove to be polygenic.

Stones arising in childhood or young adulthood can be related to autosomal recessive disorders such as cystinuria and primary oxaluria, although the latter tends to affect younger children. Cystinuria is thought to result from a genetic defect on chromosome 2, whereas the exact genetic abnormality that causes primary oxaluria remains unknown. Several inherited metabolic disorders can also cause hyperuricosuria.

Medications

Medications can contribute to stone formation in several ways. Calcium-containing antacids, for example, can increase the amount of calcium absorbed and subsequently excreted. Loop diuretics, on the other hand, can directly promote renal tubular excretion of calcium. Acetazolamide, a weak diuretic, induces a mild metabolic acidosis and alkaline urine, favorable conditions for the development of calcium phosphate stones. Other medications, such as salicylates and probenecid, are implicated in uric acid lithiasis.

Certain crystals or stones may consist completely of precipitated medication. Such medications include intravenously administered acyclovir, triamterene, and indinavir. Although oxalate is a metabolic end product of vitamin C, there has been no obvious correlation between vitamin C ingestion and calcium oxalate nephrolithiasis. Nonetheless, patients are counseled to avoid large doses of the vitamin.

Lifestyle and Diet

Occupation and lifestyle are aspects of the social history that may contribute to stone formation. Surgeons and traveling salespeople, for example, tend to minimize fluid intake in order to avoid frequent micturition throughout the day. Loss of insensible fluid can also exacerbate nephrolithiasis and may be related to employment (e.g., construction work) or hobbies (running, gardening).

The evaluation proceeds with a thorough review of the patient's diet and fluid intake. Patients are asked to review what they eat at all meals and snacks. Particular attention is paid to ingestion of foods high in sodium (fast foods, canned foods, added salt or soy sauce) and the quantity of animal protein consumed (see later). Patients are also asked to list four or five favorite foods or snacks to assess whether they may be consuming foods high in oxalate or purine as well. Many patients are erroneously counseled by physicians to avoid calcium-containing foods. Doing so not only results in bone demineralization, a grave concern in women with stones, but also appears to be associated with increased stone formation.
Physical Examination

For most patients with nephrolithiasis, physical findings are normal; in some patients, however, the findings may reveal a systemic disorder related to the stone disease. An enterocutaneous fistula, for example, may be associated with Crohn’s disease, a common cause of enteric oxaluria. A paraplegic patient with an indwelling catheter may be susceptible to frequent urinary tract infections with urease-producing organisms and consequent struvite stone formation. Hyperuricosuria and uric acid stone formation may be seen in patients with tophi related to gout.
Laboratory Tests

Although valuable information is gleaned from the history and physical examination, it is often difficult to determine the metabolic cause of a patient's nephrolithiasis without laboratory data. The urinalysis is an easy and inexpensive test that provides a great deal of information. Uric acid and calcium oxalate stones, for example, grow more favorably at an acidic pH, and a consistently high urinary pH may suggest calcium phosphate or struvite nephrolithiasis. The specific gravity, if high, may confirm suspicions of inadequate fluid intake.

Hematuria is often present in active stone disease. Microscopic examination of the urine in this case may reveal characteristic crystals. Bacteria and pyuria noted in conjunction with a high urinary pH (>6.5) are characteristic of struvite stone disease. Urine specimens for culture should be obtained in this setting. Because enough urease may be produced to form struvite stones even when colony counts are low (<50,000 colony-forming units), the microbiology laboratory should be instructed specifically to identify the organism and to check for urease despite low colony counts.

If available, qualitative cystine screening should be performed on a urine specimen. Urine turns purple-red when sodium nitroprusside is added to a specimen containing cystine at a concentration greater than 75 mg/L.

Recommended blood tests in the basic evaluation include electrolytes (sodium, potassium, chloride, bicarbonate), serum creatinine to determine the overall renal function, uric acid, calcium, and phosphorus. If the calcium level is elevated or at the upper limit of normal or if the serum phosphorus level is reduced or at the lower limit of normal, a serum intact PTH level is also determined to rule out primary hyperparathyroidism. Low serum bicarbonate levels suggest a hypocitraturic disorder such as renal tubular acidosis (RTA) or acetazolamide ingestion.
Stone Analysis

Stone analysis should be performed, whenever possible, in patients with a new history of nephrolithiasis or in patients with long-standing stone disease who note a difference in clinical presentation or in the color, shape, or texture of any stone passed. Knowing the constituents of a stone can help the physician target certain elements of the medical history and specific urine studies. In most cases, the stone must be sent to an outside laboratory for examination. X-ray diffraction crystallography and infrared spectroscopy are currently the most accurate methods available for stone analysis.\(^5\)
Radiologic Evaluation

A number of useful radiologic tests can help determine the location and extent of the stone burden and may elucidate genitourinary abnormalities contributing to stone formation. In patients with known radiopaque stones (calcium, cystine, and struvite), a plain film of the abdomen that includes a view of the kidneys, ureters, and bladder (KUB) demonstrates most stones. Radiolucent stones composed of uric acid and xanthine, however, are not visible. The intravenous pyelogram (IVP) may be superior for screening, in that it can reveal all stones as filling defects regardless of composition. Abnormalities of the genitourinary tract that can predispose to nephrolithiasis (e.g., medullary sponge kidney and caliceal abnormalities) are also evident with IVP but not with KUB films. In an acute episode of renal colic, the osmotic diuresis generated by the contrast agent administered may flush out the offending stone.

Disadvantages of IVP include additional radiation exposure, since numerous radiographs must be taken and the patient is exposed to radiographic contrast material. Administration of contrast material should be avoided in patients with contraindications to this measure, such as renal insufficiency, diabetes mellitus, and proteinuria.

Despite the advantages of IVP, non-contrast-enhanced helical (or spiral) CT is increasingly considered the diagnostic test of choice in many institutions. Helical CT has proved to be at least as sensitive and specific as IVP in detecting stones of all types in both the kidneys and ureters. In addition, it can more accurately reveal causes of flank pain and hematuria not related to stones and requires no exposure to intravenous contrast material. Radiation exposure is a disadvantage of both CT and IVP, although the exposure to patients undergoing helical CT may be triple that of IVP. Helical CT takes less time to perform, a potential advantage in an emergency department setting, but tends to be more costly.

Renal ultrasound examinations are useful when radiation exposure must be limited. In general, ultrasonography is excellent in demonstrating stones within the kidneys but may miss ureteral stones.

Once a patient is known to have a certain type of stone, specific tests may be used in follow-up. For example, a patient known to have asymptomatic calcium stones can have a KUB test 6 to 12 months later to assess for any increase in stone size or number.
THE COMPLETE EVALUATION (see Table 28-3)

The complete evaluation comprises the entire basic examination as well as a 24-hour urine collection to determine volume and levels of calcium, oxalate, citrate, sodium, urate, phosphorus, and creatinine (Table 28-6). Creatinine is used to assess the adequacy of the collection; men should excrete at least 15 to 20 mg/kg of creatinine per day, and women should excrete 10 to 15 mg/kg of creatinine per day. In patients known to have cystine stones or in whom prior urine studies have been unrevealing, cystine should also be measured.

Patients should be instructed to collect the urine on a day when they perform usual activities and have their typical fluid and dietary intake. The first morning's urine specimen is discarded, and all urine for the next 24 hours (including the next morning's specimen) is collected in the jug. The collection jug should contain preservatives for some or all of the constituents being measured. In our laboratory, all the measurements may be done with hydrochloric acid as a preservative. Certain laboratories require various preservatives for the different factors measured. Physicians should ask their laboratory how many 24-hour urine collections and which preservatives are required for the complete evaluation.

The patients who require the complete evaluation are as follows: (1) all children, (2) nonwhites (demographic groups not typically prone to nephrolithiasis), (3) noncalcium stone formers, and (4) patients with metabolically active stone disease (metabolically active stones are those that grow in size or number within 1 year).

| Table 28-6 – Optimal 24-Hour Urine Values in Patients with Nephrolithiasis |
|------------------|-----------------|
| Volume           | >22.5 L         |
| pH               | >5.5, <7.0 (24-hr specimen not required) |
| Calcium          | <300 mg or <3.54.0 mg/kg in men |
|                  | <250 mg or <3.54.0 mg/kg in women |
| Oxalate          | <40 mg          |
| Sodium           | <3000 mg or <130 mEq |
| Uric acid        | <800 mg in men  |
|                  | <750 mg in women |
| Phosphorus       | <1100 mg        |
| Citrate          | >320 mg         |
| Creatinine       | >15 mg/kg in men |
|                  | >10 mg/kg in women in order to ensure adequacy of collection |

Table 28-6 – Optimal 24-Hour Urine Values in Patients with Nephrolithiasis
NONSPECIFIC THERAPY

Most patients are given general advice about fluid and dietary modification to prevent further stone formation. These nonpharmacologic interventions, which include an increase in fluid intake as well as restriction of dietary sodium and animal protein, can reduce the incidence of stone formation, a result termed the stone clinic effect. In one study, such interventions resulted in a 60% decrease in stone recurrence over 5 years. Many patients with calcium stones are erroneously advised to restrict their calcium intake, a recommendation that is harmful not only with respect to preservation of bone mineral but also with regard to stone formation.

Augmented fluid intake to ensure a urine volume greater than 2 to 2.5 L/day has proved efficacious in reducing stone formation and continues to be a mainstay of therapy for all stone types. With an increased urine volume, the supersaturation of calcium oxalate and other crystals is significantly reduced.

Renal calcium excretion is augmented by increased sodium excretion, and hypercalciuric patients tend to have a greater calciuric response to a sodium load than control subjects. Dietary sodium restriction with the consequent decrease in urinary sodium thus reduces calcium excretion. Patients are therefore counseled to limit their daily sodium intake to 3000 mg (130 mEq) in an attempt to reduce hypercalciuria.

A moderate reduction in animal protein (1.0 mg/kg per day) is beneficial in patients with nephrolithiasis because of the multiple mechanisms by which animal protein can contribute to stone formation. A mild metabolic acidosis develops when animal proteins are metabolized. In order to buffer the excess hydrogen ions, calcium is resorbed from bone, which increases the filtered load of calcium. Metabolic acidosis also diminishes renal tubular calcium reabsorption, which further enhances the hypercalciuria. In addition, metabolism of amino acids contained in animal protein generates sulfate ions, which couple with calcium ions to form insoluble complexes.

Citrate, a base, forms soluble complexes with calcium and is beneficial in lowering calcium oxalate and calcium phosphate oversaturation and in reducing stone formation. During metabolic acidosis, citrate is reabsorbed from bone, which increases the filtered load of calcium. Metabolic acidosis also diminishes renal tubular calcium reabsorption, which further enhances the hypercalciuria. In addition, metabolism of amino acids contained in animal protein generates sulfate ions, which couple with calcium ions to form insoluble complexes.

As indicated previously, hypercalciuric patients have often been categorized as having either excessive renal calcium excretion (renal leak) or excessive absorption of calcium from the GI tract (absorptive hypercalciuria). For several reasons, these distinctions and the numerous possible subdivisions are no longer favored.

First, both human and animal studies have shown that these are probably not two distinct entities but, rather, a linked abnormality that leads to a spectrum of both excessive calcium absorption and excretion. Second, patients with enhanced intestinal calcium absorption have generally been prescribed a low-calcium diet. Although some patients may absorb an excessive amount of calcium, many patients with IH given a low-calcium diet excrete more calcium than they absorb, rendering them vulnerable to excessive bone demineralization.

Studies have demonstrated the apparently paradoxical benefits of a high-calcium diet in patients with kidney stones. Ingested calcium binds oxalate in the intestine, reducing its absorption and consequent renal excretion. It has been suggested that a diet high in calcium decreases oxalate absorption and, consequently, urinary calcium oxalate supersaturation, leading to reduced stone formation.

---

**TABLE 28-7 -- Foods High in Oxalate**

<table>
<thead>
<tr>
<th>Beans (green and dried)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beer</td>
</tr>
<tr>
<td>Draft, stout, lager, pilsner</td>
</tr>
<tr>
<td>Beets</td>
</tr>
<tr>
<td>Berries (blackberries, blueberries, raspberries, strawberries, juice containing berries)</td>
</tr>
<tr>
<td>Black tea</td>
</tr>
<tr>
<td>Black pepper</td>
</tr>
<tr>
<td>Celery</td>
</tr>
<tr>
<td>Chocolate, cocoa</td>
</tr>
<tr>
<td>Eggplant</td>
</tr>
<tr>
<td>Figs, dried</td>
</tr>
<tr>
<td>Greens (collard greens, dandelion greens, endive, escarole, kale, leeks, mustard greens, parsley, sorrel, spinach, Swiss chard, watercress)</td>
</tr>
<tr>
<td>Green peppers</td>
</tr>
<tr>
<td>Lemon, lime and orange peel</td>
</tr>
<tr>
<td>Nuts</td>
</tr>
<tr>
<td>Pecans, peanuts, peanut butter</td>
</tr>
<tr>
<td>Okra</td>
</tr>
<tr>
<td>Rhubarb</td>
</tr>
<tr>
<td>Sweet potato</td>
</tr>
<tr>
<td>Tofu</td>
</tr>
</tbody>
</table>


---
ETIOLOGY AND SPECIFIC THERAPY

The optimal therapy for patients with metabolically active stone disease is directed at the particular metabolic abnormality. Before medications for nephrolithiasis are prescribed, all patients should be treated with the nonspecific measures noted earlier.

Before therapy is initiated, it is also worth assessing the patient’s existing stone burden with a radiologic examination (KUB, IVP, spiral CT, or ultrasonography). If stones are seen, the subsequent passage of stones would not necessarily indicate therapeutic failure. The basic and complete evaluations help direct the clinician to the specific treatments discussed here.

**Calcium Stones**

Most kidney stones (70%) contain calcium (see Table 28-1). More than one third of these are composed of calcium oxalate alone, and another 7% are composed of calcium phosphate alone. The remainder are composed of a combination of calcium oxalate and either urate or calcium phosphate. The stones tend to be gray, brown, or tan and rarely grow larger than 1 to 2 cm.

The main causes of calcium stone formation are as follows:

- Hypercalciuria (excessive urinary calcium excretion)
- Hyperoxaluria (excessive oxalate excretion)
- Hyperuricosuria (excessive uric acid excretion)
- Hypocitraturia (insufficient citrate excretion)
- Renal tubular acidosis
- Congenital abnormalities of the genitourinary tract
- Certain medications

**Hypercalciuria**

Patients with persistent hypercalciuria despite increased volume intake and a low-sodium diet often benefit from a thiazide diuretic. This class of drugs is inexpensive and extremely effective at reducing urinary calcium excretion and stone formation.

In order to maximize the efficacy of thiazides, patients must consume a sodium-restricted diet. Whereas hydrochlorothiazide is more commonly used for hypertension, chlorthalidone is favored for the treatment of hypercalciuria because it has a longer half-life and requires only daily dosing. The starting dose is 25 mg and can be increased to 50 mg. In petite patients or those with low blood pressure, therapy can be initiated with 12.5 mg.

One side effect of thiazides is an increase in serum lipid levels. For patients in whom this is a concern, such as those with hypercholesterolemia or other cardiac risk factors, indapamide, 1.25 to 2.5 mg, is a good alternative because it has less effect, compared with thiazides, on serum lipids.

Hypokalemia is another common side effect of thiazide therapy. Patients should be advised to increase their dietary intake of potassium-rich foods, and the potassium level should be checked 7 to 10 days after the start of the medication. Hypokalemia can result not only in cardiac and neuromuscular problems but also in hypocitraturia, another risk factor for stone formation. The supplement of choice, therefore, is potassium with a base, such as citrate, as the accompanying anion.

Potassium citrate is available as a liquid or as a wax matrix tablet. The latter form is preferable because patients generally find the liquid unpalatable. Patients with malabsorptive disorders, however, absorb potassium citrate better in the liquid form. Potassium citrate, in the wax matrix formulation, is available as 5- and 10-mEq tablets; 20 to 40 mEq/day in single or divided doses is usually adequate supplementation. Determination of follow-up potassium and bicarbonate levels may be required for further dose adjustment. Because citrate is a base, metabolic alkalosis can result with this medication and an alternative potassium supplement (e.g., potassium chloride) may be required. If hypokalemia persists or if large doses of supplemental medication are required, the patient may benefit from the addition of a potassium-sparing diuretic. Triamterene is generally avoided because it can precipitate into stones. Amiloride, however, may be initiated at a starting dose of 5 mg or in a combination tablet with thiazide.

After at least 4 weeks of the new medication, the 24-hour urine test should be repeated to assess the efficacy of therapy in reducing calcium levels; 24-hour urinary sodium and citrate levels should also be measured. The thiazide dose may need to be increased to decrease calcium excretion to less than 3 to 4 mg/kg per day. If sodium excretion remains high in conjunction with elevated urinary calcium excretion, further dietary counseling aimed at reducing dietary sodium may be required. Additional potassium citrate may be required if urinary citrate or serum potassium levels remain low.

**Hyperoxaluria**

Oxalate is produced predominantly by endogenous metabolism of glyoxylate and, to a lesser extent, by ascorbic acid. Some urinary oxalate is derived from dietary sources, such as rhubarb, cocoa, nuts, tea, and certain leafy green vegetables. Absorbed oxalate is excreted unchanged in the urine and raises urinary supersaturation with respect to calcium oxalate. Hyperoxaluria accounts for the formation of approximately 5% of all calcium stones.

The three main causes of hyperoxaluria are:

1. Excessive oxalate ingestion (dietary oxaluria).
2. Malabsorptive GI disorders (enteric oxaluria).
3. Excessive endogenous metabolism of oxalate related to a hepatic enzyme deficiency (primary hyperoxaluria).

Since ethylene glycol (used as an antifreeze in automobiles) is metabolized to oxalate, nephrolithiasis is frequently observed in patients after ingestion of ethylene glycol.

**Dietary Oxaluria**

Dietary oxaluria results in urinary oxalate levels that are mildly elevated (40 to 60 mg/day). Patients with hyperoxaluria should be given detailed lists of high-oxalate foods to avoid (Table 28-7). In addition, calcium carbonate (500 to 650 mg per tablet, two or three tablets with each meal and snacks) may further help reduce the amount of oxalate absorbed. Alternatively, patients can be advised to drink a glass of milk with meals, especially meals that might be high in oxalate. The calcium in milk binds the dietary oxalate and helps prevent its absorption.

**Enteric Oxaluria**

Enteric oxaluria results in higher urinary oxalate levels (60 to 100 mg/day). GI malabsorptive conditions associated with normal colonic function, such as Crohn's
disease, celiac sprue, jejunoileal bypass, chronic pancreatitis, and biliary obstruction, may lead to enteric oxaluria. In these disorders, malabsorbed fatty acids bind calcium in the gut, making more free calcium available for absorption in the colon. In addition, the colonic mucosa becomes more permeable to oxalate as a result of exposure to malabsorbed bile salts.

The mainstay of treatment, whenever possible, is therapy for the underlying disorder. A gluten-free diet, for example, can significantly reduce hyperoxaluria associated with sprue; for other conditions (e.g., surgical short-bowel syndrome), no specific therapy is feasible. In such cases, reduction of malabsorption and oxalate absorption may be achieved by instituting general therapy for steatorrhea, such as a low-fat diet, cholestyramine, and medium-chain triglycerides.

As in patients with dietary oxaluria, an oxalate-restricted diet and calcium carbonate with meals should be prescribed. Because of chronic diarrhea, these patients are also at risk for low urine volumes, hypocitruria, hypokalemia, and hypomagnesemia. The acidic, concentrated urine also predisposes to development of uric acid stones. Additional fluid intake must be stressed, and potassium citrate (the liquid form is generally better absorbed in these patients) and magnesium supplementation are often prescribed. Magnesium appears to be an inhibitor of stone formation and is supplied as magnesium oxide at 400 mg by mouth twice a day or magnesium gluconate at 0.5 to 1 g by mouth three times a day.

Primary Hyperoxaluria

Primary hyperoxaluria (PH) leads to nephrolithiasis, since hepatic enzyme deficiencies in these patients lead to massive endogenous oxalate production. PH results not only in severe hyperoxaluria (80 to 300 mg/day) but also in widespread deposition of oxalate in numerous organs and tissues such as the heart, bone marrow, muscle, and renal parenchyma at a young age. Cardiomyopathy, bone marrow suppression, and renal failure may ensue. In type 1 PH, the deficient hepatic enzyme is alanine:glyoxylate aminotransferase (AGT). In type 2 PH, which is an even more uncommon disorder, patients lack α-glycerate reductase and glyoxylate reductase. Patients with type 1 PH benefit from pyridoxine (vitamin B₆) therapy, which may reduce oxalate production.

All patients with PH should be treated with measures that reduce calcium oxalate precipitation, such as large fluid supplementation, potassium citrate, magnesium, and orthophosphate. Orthophosphate is an effective inhibitor of calcium oxalate crystalization but should be avoided in patients with a glomerular filtration rate below 50 mL/minute. Patients with renal failure may benefit from renal transplantation because dialysis is not as effective as a functioning kidney in oxalate removal.

The preceding measures should be continued after transplantation to avoid rapid loss of the allograft caused by calcium oxalate deposition. Ultimately, for patients with type 1 PH, liver transplantation can supply the missing AGT and may be curative, especially if it is performed before the development of end-stage renal failure. Some patients require combined liver-kidney transplantation.

Hyperuricosuria

Calcium oxalate crystals in the urine preferentially nucleate around other types of crystals or sloughed cells (heterogeneous nucleation). Uric acid crystals frequently form the nidus or internal core of calcium oxalate stones. In fact, up to 15% of calcium stones are found in patients with hyperuricosuria. In contrast to patients with pure calcium oxalate stones, these patients typically have elevated urinary uric acid levels but normal urinary calcium and oxalate levels. They also differ from patients with pure uric acid stones in that they tend to have a higher urinary pH (>5.5).

Therapy consists of dietary purine restriction and increased fluid intake. If urinary uric acid levels remain uncontrolled with these measures, allopurinol, 100 to 300 mg/day, may be added.

Hypocitruria

Citrate combines with calcium to form a soluble complex that reduces both calcium oxalate and calcium phosphate precipitation. In some patients, hypocitruria is the principal metabolic abnormality found in the 24-hour urine collection. Risk factors for hypocitruria include high protein intake, hypokalemia, metabolic acidosis, exercise, infection, starvation, androgens, and acetazolamide. Men tend to have lower urinary citrate concentrations than women, which may be responsible for the higher incidence of stone formation in men.

Furthermore, women with nephropthiasis have lower urinary citrate concentrations than nonstone-forming women.

Along with therapy for the underlying condition, such as moderating dietary protein intake, potassium citrate is prescribed. This salt is preferable to sodium citrate because sodium excretion promotes calcium excretion. Again, potassium citrate in the wax matrix formulation is preferred to the liquid preparation because of increased palatability. Large amounts may be required (30 to 75 mEq/day) in divided doses in order to raise the urinary citrate concentration to more than 320 mg/day. Potassium and bicarbonate levels should be closely monitored, especially in patients with renal insufficiency. If metabolic alkalosis ensues, partial supplementation with potassium chloride may be necessary.

Renal Tubular Acidosis

Distal (type 1) RTA is a disorder in which distal tubular hydrogen ion excretion is impaired, resulting in a nonanion gap metabolic acidosis and a persistently alkaline urine. The acidosis leads to calcium and phosphate release from bone as well as enhanced proximal tubular reabsorption of citrate. The net result is an increased filtered load of calcium and phosphate, severe hypocitraturia, and an elevated urinary pH, all of which promote calcium phosphate precipitation. Nephrocalcinosis, or renal parenchymal calcification, is frequently seen in this setting.

Twenty-four-hour urinary citrate levels are commonly below 100 mg in patients with distal RTA. Therapy consists of potassium citrate or potassium bicarbonate supplementation in order to treat both the metabolic acidosis and hypocitraturia. Large doses of these medications are often required (1 to 2 mEq/kg per day in two or three divided doses).
Nephrocalcinosis

Nephrocalcinosis is a process in which calcium is deposited in the renal parenchyma. There are two forms:

1. **Dystrophic calcification.** Calcium deposition arises from tissue necrosis secondary to neoplasm, infarction, or infection. It may be seen in the setting of renal transplant rejection, renal cortical necrosis, chronic glomerulonephritis, ethylene glycol toxicity, acquired immunodeficiency syndrome (AIDS)–related infections, and Alport’s syndrome. In general, serum calcium and phosphorus levels are normal and calcium-phosphate deposition occurs predominantly in the renal cortex.

2. **Metastatic calcification.** Patients may have elevated serum calcium and phosphate levels or an elevated urinary pH. Calcification in this setting occurs more commonly in the renal medulla. Common causes include RTA, primary hyperparathyroidism (or any disorder resulting in elevated serum calcium levels), medullary sponge kidney, papillary necrosis, acetazolamide, amphotericin B, triamterene, and primary hyperoxaluria. Primary hyperoxaluria can result in both medullary and cortical calcifications.

Both medullary and cortical parenchymal calcifications are easily noted with ultrasonography and CT scanning, even before they can be detected on plain radiographs. Therapy consists of treating the underlying disorder whenever possible. Otherwise, measures aimed at reducing hypercalcemia, oxalosis, and hyperphosphatemia should be attempted.
Uric Acid Stones (see Table 28-4)

Although uric acid stones make up only about 5% to 10% of all calculi formed in the United States, the prevalence of uric acid lithiasis is much greater in Mediterranean countries. These stones tend to be round, smooth, and yellow-orange. Because they are radiolucent, they are not visible on plain films but can be detected by ultrasonography or CT as filling defects on IVP. Uric acid is a purine metabolite and is also found in large quantities within cells. The three main causes of uric acid stone formation are (1) low urine volume, (2) low urinary pH, and (3) elevated urinary uric acid levels.

Urine Volume and pH

Any disorder that results in low urine volume (e.g., diarrheal disorders, diaphoresis, reduced fluid intake) can contribute to uric acid lithiasis. Diarrhea and diets high in animal protein can also contribute to an acidic urinary pH. Uric acid is increasingly soluble at an alkaline urinary pH such that urine with a pH of 6.5 can contain more than five times more uric acid than urine at pH 5.3 without inducing precipitation and may actually dissolve existing stones.  

Hyperuricosuria

Hyperuricosuria may be evident in patients who ingest large quantities of dietary purine or animal protein. Foods high in purine include organ meats, shellfish, certain fish, meat extracts, yeast, gravies, and stocks (Table 28-8). Hyperuricemic disorders such as gout, myeloproliferative disorders, tumor lysis syndrome, and certain inborn errors of metabolism (e.g., glucose-6-phosphatase deficiency, Lesch-Nyhan syndrome) may also contribute to an increased urinary filtered load of uric acid. Certain medications such as salicylates and probenecid can be hyperuricosuric as well.  

Therapy for patients with uric acid stones begins with nonspecific measures such as increasing fluid intake to maintain urine volume at about 3 L/day. Patients are prescribed a low-purine diet to decrease uric acid production. Despite dietary intervention, hyperuricemia often persists, especially in patients with disorders of cellular catabolism. In this setting, allopurinol should be prescribed at a starting dose of 100 mg/day, increasing to 300 mg/day as needed. 

A diet low in animal protein is also beneficial because the

<table>
<thead>
<tr>
<th>TABLE 28-8 -- Foods High in Purine</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Organ meats</strong></td>
</tr>
<tr>
<td>Brain, heart, kidney, liver, sweetbreads</td>
</tr>
<tr>
<td><strong>Meat extracts</strong></td>
</tr>
<tr>
<td>Bouillon, consomme, stock, gravies</td>
</tr>
<tr>
<td><strong>Meat</strong></td>
</tr>
<tr>
<td>Beef, chicken, goose, lamb, pork</td>
</tr>
<tr>
<td><strong>Shellfish</strong></td>
</tr>
<tr>
<td>Clams, mussels, scallops, shrimp, oysters</td>
</tr>
<tr>
<td><strong>Fish</strong></td>
</tr>
<tr>
<td>Anchovies, fish roe, herring, mackerel, sardines, others</td>
</tr>
<tr>
<td><strong>Certain vegetables</strong></td>
</tr>
<tr>
<td>Asparagus, cauliflower, kidney beans, lentils, lima beans, mushrooms, peas, spinach</td>
</tr>
</tbody>
</table>


Decreased endogenous acid production raises urinary pH. Ideally, the urinary pH should be elevated to approximately 6.5 to 7.0, a level that can dissolve existing crystals and stones. A urinary pH higher than 7.0 should be avoided, however, as calcium phosphate deposition may result. Potassium citrate at doses of 30 mEq by mouth twice a day or greater may be required to raise the urinary pH sufficiently. (See "Hypercalciuria" and "Hypocitruria" on available potassium citrate preparations.) Prescription of Nitrazine paper allows patients to monitor the urinary pH at various times of day and adjust their potassium citrate intake accordingly.

Although sodium bicarbonate may effectively alkalize the urine, it should be avoided because the additional sodium excretion encourages sodium urate formation. Sodium urate in the setting of alkaline urine can serve as a nidus for calcium oxalate precipitation. If the urinary pH cannot be raised adequately despite high doses of potassium citrate or if the dose prescribed results in hyperkalemia, the carbonic anhydrase inhibitor acetazolamide may be initiated. This medication results in an alkaline urine and mild systemic metabolic acidosis, a pattern similar to that in type 1 RTA. Again, the urinary pH should be maintained at less than 7.0 in order to avoid calcium phosphate precipitation.
Struvite Stones (see Table 28-4)

Struvite stones have also been termed triple phosphate stones, magnesium ammonium phosphate stones, and infection stones. Although they make up only about 10% to 15% of all stones formed, most staghorn calculi (i.e., large stones that extend beyond a single renal calix) are composed of struvite. The propensity of these stones to grow rapidly to a large size, to recur despite therapy, and to result in significant morbidity (and potential mortality) has also led to the appellation stone cancer. Infection with urease-producing bacteria must be present for these stones to form, and therefore severe renal infections as well as sepsis and loss of renal function may develop.

In contrast to other stone types, struvite stones occur with a higher incidence in women than in men, largely because of women’s increased susceptibility to urinary tract infections. Other groups at risk for development of struvite stones because of urinary stasis or infection include elderly people and patients with neurogenic bladders, indwelling urinary catheters, spinal cord lesions, or genitourinary abnormalities. Even in the absence of stone analysis, struvite stones should be suspected in patients with large stones, an alkaline urinary pH (>7), and the presence of urease-producing bacteria in the urine. Early detection and therapy are essential to avoid great potential morbidity.

Urease-Producing Bacteria

The formation of struvite stones is dependent on the presence of both ammonium ions and an alkaline urinary pH, conditions met clinically only through the actions of urease-producing bacteria. Ammonium, magnesium, and carbonate apatite (\( \text{Ca}_10(\text{PO}_4)_6\text{CO}_3 \)) in the urine combine with phosphate, which is present in this setting in its trivalent form.

Numerous bacteria, both gram-negative and gram-positive, as well as Mycoplasma and yeast species have been implicated in urease production. Bacteria in which urease is frequently isolated include Proteus species, Haemophilus, Corynebacterium, and Ureaplasma. Escherichia coli, despite its frequent role as a urinary tract pathogen, has not been shown to produce urease. Urease production adequate to stimulate stone formation may be present despite low bacterial colony counts. For this reason, the microbiology laboratory should be asked specifically to perform bacterial identification and to determine sensitivities even with colony counts lower than 100,000 colony-forming units. If no bacteria are isolated but a urease producer is suspected, special cultures for Ureaplasma urealyticum, a mycobacterium, should be ordered.

Therapy

In order to eradicate struvite stones, early and aggressive medical and urologic management is required. Appropriate antibiotic therapy is essential but must be combined with long-term bacterial suppression and complete surgical or medical stone removal. Extracorporeal shock wave lithotripsy is often adequate for fragmentation of stones less than 2 cm in size, but percutaneous nephroscopy or a combination of the two procedures is usually required for larger stones. Antibiotics should be continued on the basis of cultures of any stone fragments retrieved. After approximately 2 weeks of antibiotic therapy, when the urine culture is sterile, the dose of antibiotic should be halved. Suppressive antibiotics should continue at this dose as long as monthly surveillance cultures remain sterile for three consecutive months. At this point, antibiotics may be discontinued as long as surveillance urine cultures are obtained monthly for 1 year.

In addition to antimicrobial therapy, medical treatment may involve urease inhibition and chemolysis. In chemolysis, the kidney is irrigated with an acidic solution through a nephrostomy tube or ureteral catheter. Although rarely used today with the advent of less invasive surgical techniques, it can still play a role in dissolution of residual stone fragments. Ten percent hemiacidrin, the solution most commonly utilized, is composed of carbonic acid, citric acid, \( \alpha \)-gluconic acid, and magnesium at a pH of 3.9. The use of chemolysis was controversial in the past because high mortality rates were reported, but with close monitoring of serum magnesium levels, intrapelvic pressures, infection, and obstruction to flow, it is now thought to be relatively safe.

Urease inhibition has been shown to retard stone growth and to prevent new stone formation. It does not decrease bacterial counts and cannot eradicate existing stones. Combined with antimicrobial therapy, it serves primarily as palliative care for patients who cannot undergo definitive surgical management. The agent most commonly used is acetohydroxamic acid (AHA). These medications require adequate renal clearance for therapeutic efficacy and are contraindicated in patients with a serum creatinine level higher than 2mg/dL (or 176 µmol/L). Renal insufficiency can increase the incidence of side effects of the medications, which are numerous and limit their use. Side effects that result in discontinuation of the drug include neurologic symptoms, GI upset, hair loss, hemolytic anemia, and rash. Fortunately, the side effects all resolve with discontinuation of the drug. AHA is also teratogenic. The starting dose of AHA is 250 mg by mouth twice a day. If it is well tolerated for about 1 month, the dose is increased to 250 mg by mouth three times a day.
Cystine Stones (see Table 28-4)

Cystinuria is an autosomal recessive disorder that results in excessive renal tubular excretion of the dibasic amino acids cystine, ornithine, lysine, and arginine. The frequency of the abnormal gene, which is found on chromosome 2, is fairly high at 0.01%. The genetic defect would probably go unnoticed were it not for the low solubility of cystine, 250 mg/L. Cystinuria should not be confused with cystinosis, a more serious and debilitating disorder that results in extensive intracellular cystine accumulation. Whereas people with no tubular defect in cystine transport excrete approximately 30 to 50 mg of cystine per day, heterozygotes excrete about 400 mg/day and homozygotes often excrete larger amounts (>600 mg/day).

Stones usually develop in patients within the second or third decade. The stones may grow to large size and can appear as staghorn calculi or multiple stones. They are radiopaque because of the sulfur content of cystine molecules. The disease should be suspected in any patient with stone onset in childhood, frequent recurrence of nephrolithiasis, and a strong family history of the disease. The presence of the classic, hexagonal cystine crystals in the urine can verify the diagnosis. As these crystals may not be evident in a dilute or alkaline urine, qualitative screening with the sodium nitroprusside test better confirms the presence of cystinuria at a concentration greater than 75 mg/L. Quantitative cystine measures with a 24-hour urine sample should follow to determine the risk of stone formation and to guide therapy.

Therapy

The aim of treatment is to lower the urinary cystine concentration below the limits of solubility (<250 mg/L). As with the treatment of other stones, nonspecific therapy plays a large role. Patients are advised to drink large quantities of fluids. A patient with a cystine excretion of 750 mg/day, for example, should ideally drink enough fluid to increase urine output to more than 3 L/day. Large quantities of milk should be avoided because dairy products and foods high in protein contain large quantities of methionine, an essential amino acid that is a precursor of cystine. Since cystine is more soluble at a higher pH, juices are encouraged, as they tend to alkalize the urine. Potassium citrate (see "Hypercalciuria" and "Hypocitraturia" for details) is also prescribed to maintain the urinary pH between 6.5 and 7.0.

Approximately 50% of cystine stones are mixed stones. Patients with cystinuria often have other metabolic defects such as hypercalciuria, hypocitraturia, and hyperuricosuria. Therefore, a complete 24-hour urine collection for all stone-forming elements is necessary to treat nephrolithiasis fully in this setting.

If the preceding measures are inadequate in controlling stone formation or if the urinary cystine concentration is too high to make adequate fluid intake practical, chelating agents may be added. D-Penicillamine is a chelating agent that reduces the cystine concentration by forming a more soluble compound with cystine. The medication is associated with numerous serious side effects that limit its use. Second-generation and third-generation chelating agents such as -mercaptopropionylglycine and bucillamine are now available that reduce the cystine concentration with fewer side effects.

1423
Acknowledgments

This work was supported, in part, by grants DK 56788, DK 57716, and AR 46289 (all to David A. Bushinsky) from the National Institutes of Health.
References


Larsen: Williams Textbook of Endocrinology, 10th ed., Copyright © 2003 Elsevier


Section 8 - Disorders of Carbohydrate and Lipid Metabolism

Chapter 29 - Type 2 Diabetes Mellitus

John B. Buse
Kenneth S. Polonsky
Charles F. Burant

EPIDEMIOLOGY

Type 2 diabetes is the predominant form of diabetes worldwide, accounting for 90% of cases globally. An epidemic of type 2 diabetes is under way in both developed and developing countries, although the brunt of the disorder is felt disproportionately in non-European populations as evidenced by studies in Hispanic populations, Native American and Canadian communities, Pacific and Indian Ocean island populations, and in India and Australian Aboriginal communities. In the Pacific island of Nauru, diabetes was virtually unknown 50 years ago and is now present in approximately 40% of adults. Globally, the number of people with diabetes is expected to rise from the current estimate of 150 million to 220 million in 2010 and 300 million in 2025. Alarming increases in the prevalence of diabetes have occurred in various Chinese populations. Type 2 diabetes has become one of the world's most important public health problems.

Considerable information is available on the factors that are responsible for the development of type 2 diabetes, and these are summarized in Table 29-1. Type 2 diabetes is currently thought to occur in genetically predisposed individuals who are exposed to a series of environmental influences that precipitate the onset of clinical disease. The genetic basis of type 2 diabetes is discussed in detail later in this chapter, but the syndrome consists of monogenic and polygenic forms that can be differentiated both on clinical grounds and in terms of the genes that are involved in the pathogenesis of these disorders.

Sex, age, and ethnic background are important factors in determining risk for the development of type 2 diabetes. The disorder is more common in females, and the increased prevalence in certain racial and ethnic minority groups has already been alluded to. Age is also a critical factor. Type 2 diabetes has been viewed in the past as a disorder of aging with an increasing prevalence with age. This remains true today. However, a disturbing trend has become apparent in which the prevalence of obesity and type 2 diabetes in children is rising dramatically. In the past, it was believed that the overwhelming majority of children with diabetes had type 1 diabetes, with only 1% to 2% of children considered to have type 2 or other rare forms of diabetes. Later reports suggest that as many as 8% to 45% of children with newly diagnosed diabetes have nonimmune-mediated diabetes. The majority of these children have type 2 diabetes, but other types are being increasingly identified. An idiopathic nonimmune-mediated form of diabetes has been reported particularly in the black population.

<table>
<thead>
<tr>
<th>TABLE 29-1 – Epidemiologic Determinants and Risk Factors of Type 2 Diabetes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genetic factors</td>
</tr>
<tr>
<td>Genetic markers, family history, “thrifty gene(s)”</td>
</tr>
<tr>
<td>Demographic characteristics</td>
</tr>
<tr>
<td>Sex, age, ethnicity</td>
</tr>
<tr>
<td>Behavioral and lifestyle-related risk factors</td>
</tr>
<tr>
<td>Obesity (including distribution of obesity and duration)</td>
</tr>
<tr>
<td>Physical inactivity</td>
</tr>
<tr>
<td>Diet</td>
</tr>
<tr>
<td>Stress</td>
</tr>
<tr>
<td>Westernization, urbanization, modernization</td>
</tr>
<tr>
<td>Metabolic determinants and intermediate risk categories of type 2 diabetes</td>
</tr>
<tr>
<td>Impaired glucose tolerance</td>
</tr>
<tr>
<td>Insulin resistance</td>
</tr>
<tr>
<td>Pregnancy-related determinants (parity, gestational diabetes, diabetes in offspring of women with diabetes during pregnancy, intrauterine malnutrition or overnutrition)</td>
</tr>
</tbody>
</table>

DIAGNOSTIC CRITERIA FOR DIABETES MELLITUS

The diagnosis of diabetes rests on the measurement of plasma glucose levels. The diagnostic criteria for diabetes were changed in 1997 with the most significant changes being the level of fasting plasma glucose (FPG) that is recognized as diagnostic for diabetes, which was decreased from 140 to 126 mg/dL, and the introduction of a category of impaired fasting glucose (IFG). Current criteria for the diagnosis of diabetes, impaired fasting glucose and impaired glucose tolerance (IGT), are shown in Table 29-2.

Because plasma glucose concentrations range as a continuum, the criteria are based on estimates of the threshold for the complications of diabetes. The primary end point used to evaluate the relationship between glucose levels and complications is retinopathy. The prevalence of retinopathy in comparison with FPG and 2-hour plasma glucose has been evaluated.

TABLE 29-2 — Criteria for the Diagnosis of Diabetes

<table>
<thead>
<tr>
<th>Normoglycemia</th>
<th>Impaired Fasting Glucose or Impaired Glucose Tolerance</th>
<th>Diabetes</th>
</tr>
</thead>
<tbody>
<tr>
<td>FPG &lt;110 mg/dL</td>
<td>FPG 110 and &lt;126 mg/dL (IFG)</td>
<td>FPG 126 mg/dL</td>
</tr>
<tr>
<td>2-hr PG &lt;140 mg/dL</td>
<td>2-hr PG 140 and &lt;200 mg/dL (IGT)</td>
<td>2-hr PG 200 mg/dL</td>
</tr>
<tr>
<td></td>
<td>Symptoms of diabetes and casual plasma glucose concentration 200 mg/dL</td>
<td></td>
</tr>
</tbody>
</table>

FPG, fasting plasma glucose; IFG, impaired fasting glucose; IGT, impaired glucose tolerance; PG, plasma glucose.

* A diagnosis of diabetes must be confirmed, on a subsequent day, by measurement of FPG, 2-hour PG, or random plasma glucose (if symptoms are present). The FPG test is greatly preferred because of ease of administration, convenience, acceptability to patients, and lower cost. Fasting is defined as no caloric intake for at least 8 hours.

Both diagnostic criteria were able to predict the presence of retinopathy and, by inference, glucose levels that are diagnostic of diabetes [Fig. 29-1]. There is also an association between FPG and 2-hour plasma glucose and risk of macrovascular and cardiovascular disease. For instance, the Paris Prospective Study showed that the incidence of fatal coronary heart disease was related to both FPG and 2-hour plasma glucose that were determined at a baseline examination. Rates of disease were markedly increased at FPG greater than 125 mg/dL (6.9 mmol/L) or 2-hour plasma glucose greater than 140 mg/dL (7.8 mmol/L).

Reproducibility of the plasma glucose concentration is an important issue in the interpretation of the results of diagnostic tests for diabetes. There is significant variation in the results of repeated tests in adults after a 2- to 6-week interval. The intraindividual coefficient of variation in one study was 6.4% for the FPG and 16.7% for the 2-hour plasma glucose value. Thus, it is essential that abnormal results be confirmed by a repeated test. Although the oral glucose tolerance test (OGTT) is an invaluable tool in research, it is not recommended for routine use in the diagnosis of diabetes. It is inconvenient for patients, and in the vast majority of cases the diagnosis can be made on the basis of either an elevated fasting glucose concentration or an elevated random glucose determination in the presence of hyperglycemic symptoms.

In general, levels of hemoglobin A1c (HbA1c) are not currently recommended for the diagnosis of diabetes. The major reasons are the lack of standardization of the assays for HbA1c, and the imperfect correlation between HbA1c and FPG and 2-hour plasma glucose. However, HbA1c remains the most effective method for monitoring the effectiveness of diabetes treatment.
SCREENING FOR TYPE 2 DIABETES

Undiagnosed type 2 diabetes is common, with an estimated lag of 5 to 7 years between the onset of diabetes and diagnosis. It is estimated that in up to 50% of affected people the disease is undiagnosed. Subjects with IGT and undiagnosed type 2 diabetes are at significantly increased risk for coronary heart disease, stroke, and peripheral vascular disease. Thus, this delay in the diagnosis of type 2 diabetes causes an increase in microvascular and macrovascular disease. In addition, affected individuals have a greater likelihood of having dyslipidemia, hypertension, and obesity. Therefore, it is important for the clinician to screen for diabetes in a cost-effective manner in subjects who demonstrate major risk factors for diabetes as summarized in Table 29-3. Recommendations for screening are summarized in Table 29-4.
PATHOGENESIS OF TYPE 2 DIABETES

The pathogenesis of type 2 diabetes is complex and involves the interaction of genetic and environmental factors. A number of environmental factors have been shown to play a critical role in the development of the disease, particularly excessive caloric intake leading to obesity and a sedentary lifestyle. The clinical presentation is also heterogeneous with a wide range in age of onset, severity of associated hyperglycemia, and degree of obesity. From a pathophysiologic standpoint, persons with type 2 diabetes consistently demonstrate three cardinal abnormalities: (1) resistance to the action of insulin in peripheral tissues, particularly muscle and fat but also liver; (2) defective insulin secretion, particularly in response to a glucose stimulus; and (3) increased glucose production by the liver.

<table>
<thead>
<tr>
<th>TABLE 29-3 -- Major Risk Factors for Type 2 Diabetes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family history of diabetes (i.e., parents or siblings with diabetes)</td>
</tr>
<tr>
<td>Overweight (BMI 25 kg/m²)</td>
</tr>
<tr>
<td>Habitual physical inactivity</td>
</tr>
<tr>
<td>Race/ethnicity (e.g., African Americans, Hispanic Americans, Native Americans, Asian Americans and Pacific Islanders)</td>
</tr>
<tr>
<td>Previously identified IFG or IGT</td>
</tr>
<tr>
<td>Hypertension (140/90 mm Hg in adults)</td>
</tr>
<tr>
<td>HDL cholesterol 35 mg/dL (0.90 mmol/L) and/or a triglyceride level 250 mg/dL (2.82 mmol/L)</td>
</tr>
<tr>
<td>History of GDM or delivery of a baby weighing &gt;9 lb</td>
</tr>
<tr>
<td>Polycystic ovary syndrome</td>
</tr>
<tr>
<td>BMI, body mass index; GDM, gestational diabetes mellitus; HDL, high-density lipoprotein; IFG, impaired fasting glucose; IGT, impaired glucose tolerance.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE 29-4 -- Summary of Major Recommendations for Screening</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recommendations</td>
</tr>
<tr>
<td>Evaluation for type 2 diabetes should be performed within the health care setting. Patients should be screened at 3-year intervals beginning at age 45; testing should be considered at an earlier age or be carried out more frequently if diabetes risk factors are present.</td>
</tr>
<tr>
<td>Diabetes risk factors include a family history of diabetes; overweight, defined as BMI 25 kg/m²; habitual physical inactivity; belonging to a high-risk ethnic or racial group; previously identified IFG or IGT; hypertension; dyslipidemia; history of GDM or delivery of a baby weighing &gt;9 lb; and polycystic ovary syndrome.</td>
</tr>
<tr>
<td>The FPG is the recommended screening test. The OGTT may be necessary for the diagnosis of diabetes when the FPG is normal. The FPG is preferred for screenings because it is faster and easier to perform, more convenient, acceptable to patients, and less expensive.</td>
</tr>
<tr>
<td>Diagnostic testing should be performed in any clinical situation in which such testing is warranted; health care providers should not consider whether a person meets screening criteria in such cases.</td>
</tr>
<tr>
<td>Screening outside of health care settings, or community screening, has not been shown to be beneficial and may result in some harm; this type of screening is not recommended.</td>
</tr>
<tr>
<td>BMI, body mass index; FPG, fasting plasma glucose; GDM, gestational diabetes mellitus; IFG, impaired fasting glucose; IGT, impaired glucose tolerance; OGTT, oral glucose tolerance test.</td>
</tr>
</tbody>
</table>

Although the precise manner in which these genetic, environmental, and pathophysiologic factors interact to lead to the clinical onset of type 2 diabetes is not known, our understanding of these processes has increased substantially. With the exception of specific monogenic forms of the disease that may result from defects largely confined to the pathways that regulate insulin action in muscle, liver, and fat or defects in insulin secretory function in the pancreatic beta cell, there is an emerging consensus that the common forms of type 2 diabetes are polygenic in nature and are due to a combination of abnormal insulin secretion and insulin resistance. From a pathophysiologic standpoint, it is the inability of the pancreatic beta cell to adapt to the reductions in insulin sensitivity that occur over the lifetime of human subjects in response to puberty or pregnancy, a sedentary lifestyle, or overeating leading to weight gain that precipitates the onset of type 2 diabetes. An underlying genetic predisposition appears to be a critical factor in determining the frequency with which this occurs.

Genetic Factors in the Development of Type 2 Diabetes

Genetically, type 2 diabetes consists of monogenic and polygenic forms. The monogenic forms, although relatively uncommon, are nevertheless important and a number of the genes involved have been identified and characterized. The genes involved in the common polygenic form or forms of the disorder have been far more difficult to identify and characterize.

<table>
<thead>
<tr>
<th>TABLE 29-5 -- Monogenic Forms of Diabetes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Associated with insulin resistance</td>
</tr>
<tr>
<td>Mutations in the insulin receptor gene</td>
</tr>
<tr>
<td>Type A insulin resistance</td>
</tr>
<tr>
<td>Leprechaunism</td>
</tr>
<tr>
<td>Rabson-Mendenhall syndrome</td>
</tr>
<tr>
<td>Lipodystrophic diabetes</td>
</tr>
<tr>
<td>Mutations in the PPAR- gene</td>
</tr>
</tbody>
</table>

In the monogenic forms of diabetes, the gene involved is both necessary and sufficient to cause disease. In other words, environmental factors play little or no role in determining whether or not a genetically predisposed individual develops clinical diabetes. The monogenic forms of diabetes generally occur in young individuals, often in the first two to three decades of life, although if only mild asymptomatic elevations in blood glucose occur the diagnosis may be missed until later in life.
Mitochondrial gene mutations

- Maturity-onset diabetes of the young (MODY)
  - HNF-4 (MODY 1)
  - Glucokinase (MODY 2)
  - HNF-1 (MODY 3)
  - IPF-1 (MODY 4)
  - HNF-1 (MODY 5)
  - NeuroD1/Beta2 (MODY 6)

HNF, hepatocyte nuclear factor; IPF, insulin promoter factor; MODY, maturity-onset diabetes of the young; NeuroD1/Beta2, neurogenic differentiation 1/beta cell E-box trans-activator 2; PPAR, peroxisome proliferator-activated receptor.

The monogenic forms of diabetes are summarized in Table 29-5 and can be divided into those in which the mechanism is a defect in insulin secretion and those that involve defective responses to insulin or insulin resistance.

### Monogenic Forms of Diabetes Associated with Insulin Resistance

#### Mutations in the Insulin Receptor

More than 70 mutations have been identified in the insulin receptor gene in various insulin-resistant patients. There are at least three clinical syndromes caused by mutations in the insulin receptor gene. Type A insulin resistance is defined by the presence of insulin resistance, acanthosis nigricans, and hyperandrogenism. Patients with leprechaunism have multiple abnormalities, including intrauterine growth retardation, fasting hypoglycemia, and death within the first 2 to 3 years of life. The Rabson-Mendenhall syndrome is associated with short stature, protuberant abdomen, and abnormalities of teeth and nails; pineal hyperplasia was a characteristic in the original description of this syndrome.

These mutations may impair receptor function by a number of different mechanisms, including decreasing the number of receptors expressed on the cell surface (e.g., by decreasing the rate of receptor biosynthesis [class 1], accelerating the rate of receptor degradation [class 5], or inhibiting the transport of receptors to the plasma membrane [class 2]). The intrinsic function of the receptor may be abnormal if the affinity of insulin binding is reduced (class 3) or receptor tyrosine kinase is inactivated (class 4). The insulin resistance that is associated with insulin receptor mutations may be severe and present in the neonatal period, as with leprechaunism and the Rabson-Mendenhall syndrome, or occur in a milder form in adulthood leading to insulin-resistant diabetes with marked hyperinsulinemia, acanthosis nigricans, and hyperandrogenism.

#### Lipoatrophic Diabetes

In another monogenic form of diabetes, so-called lipoatrophic diabetes, severe insulin resistance is associated with lipoatrophy and lipodystrophy. This form of diabetes is characterized by a paucity of fat, insulin resistance, and hypertriglyceridemia. The disease has several genetic forms, including face-sparing partial lipoatrophy (the Dunnigan syndrome or the Koberling-Dunnigan syndrome), an autosomal dominant form caused by mutations in the lamin A/C gene, and congenital generalized lipodystrophy (the Seip-Berardinelli syndrome), an autosomal recessive form.

#### Mutations in Peroxisome ProliferatorActivated Receptor

It has been demonstrated that mutations in the transcription factor peroxisome proliferatoractivated receptor (PPAR) can cause type 2 diabetes of early onset. Two different heterozygous mutations were identified in the ligand-binding domain of PPAR in three subjects with severe insulin resistance. In the PPAR crystal structure, the mutations destabilize helix 12, which mediates trans-activation. Both receptor mutants showed markedly decreased transcriptional activation and inhibited the action of coexpressed wild-type PPAR in a dominant negative manner. A common amino acid polymorphism (Pro12Ala) in PPAR has been associated with type 2 diabetes. People homozygous for the Pro12 allele are more insulin resistant than those having one Ala12 allele and have a 1.25-fold increased risk of diabetes. There is also evidence for interaction between this polymorphism and fatty acids, linking this locus with diet.

### Monogenic Forms of Diabetes Associated with Defects in Insulin Secretion

#### Mutant-Insulin Syndromes

The first syndrome associated with diabetes to be characterized in terms of the clinical picture, genetic mechanisms, and clinical pathophysiology was that associated with mutant insulin or proinsulin. Persons with this disorder present clinically with a mild non-insulin-dependent form of diabetes. Affected individuals characterize themselves as having hyperinsulinemia on routine insulin assays. Increases in the concentration of insulin in association with diabetes are usually indicative of insulin resistance, but in this syndrome insulin resistance can be easily excluded because the patients respond normally to administration of exogenous insulin. Characterization of the insulin by high-performance liquid chromatography reveals that the hyperinsulinemia is due to the presence of the abnormal insulin or proinsulin and related breakdown products. The increased concentrations of insulin appear to be related to the presence of mutations in regions of the insulin molecule that are important for receptor binding, particularly the COOH terminus of the insulin B chain.

Because the liver is the major site of insulin clearance and the first-step hepatic insulin uptake and degradation are mediated by the insulin receptor, mutant forms of insulin with diminished insulin receptor binding ability are cleared more slowly from the circulation, and this reduction in insulin clearance leads to hyperinsulinemia. Alternatively, mutations in proinsulin may reduce the conversion of proinsulin to insulin, leading to accumulation of proinsulin.

Because proinsulin is cleared more slowly from the circulation than insulin, proinsulin levels increase. Proinsulin cross-reacts in most commercially available assays, and this insulin-like immunoreactivity can be characterized as related to the presence of proinsulin rather than insulin only by high-performance liquid chromatography or by the use of assays that are specific for insulin and proinsulin.

A patient with a mutation in prohormone convertase 1, one of the enzymes responsible for the conversion of proinsulin to insulin, has been described.

### Mitochondrial Diabetes

An A-to-G transition in the mitochondrial tRNALeu(UUR) gene at base pair 3243 has been shown to be associated with maternally transmitted diabetes and sensorineural hearing loss. In other subjects, this mutation is associated with diabetes and the syndrome of mitochondrial myopathy, encephalopathy, lactic acidosis, and stroke-like episodes (MELAS syndrome). The mitochondrial plays a key role in the regulation of insulin secretion, particularly in response to glucose. We have documented abnormal insulin secretion on at least one of a battery of tests in subjects with this mitochondrial mutation, even in subjects with normal or impaired glucose tolerance who have not developed overt diabetes.
Maturity-onset diabetes of the young (MODY) is a genetically and clinically heterogeneous group of disorders characterized by nonketotic diabetes mellitus, an autosomal dominant mode of inheritance, onset usually before 25 years of age and frequently in childhood or adolescence, and a primary defect in pancreatic beta cell function. A detailed review of MODY has been published,10 and the information contained in that review is summarized.

MODY can result from mutations in any one of at least six different genes. One of these genes encodes the glycolytic enzyme glucokinase (MODY5),11 and the other five encode transcription factors, hepatocyte nuclear factor (HNF)-4 (MODY7),12 HNF-1 (MODY3),13 insulin promoter factor-1 (IPF-1) (MODY4),14 HNF-1 (MODY5),15 and neurogenic differentiation 1/beta cell E-box trans-activator 2 (NeuroD1/BETA2) (MODY6).16 All of these genes are expressed in the insulin-producing pancreatic beta cells, and heterozygous mutations cause diabetes related to beta cell dysfunction. Abnormalities in liver and kidney function may occur in some forms of MODY, reflecting expression of the transcription factors in these tissues. Nongenetic factors that affect insulin sensitivity (infection, puberty, pregnancy, and rarely obesity) may trigger diabetes onset and affect the severity of hyperglycemia in MODY but do not play a significant role in the development of MODY.

The most common clinical presentation of MODY is a mild asymptomatic increase in blood glucose in a child, adolescent, or young adult with a prominent family history of diabetes often in successive generations, suggesting an autosomal dominant mode of inheritance. Some patients may have mild hyperglycemia for many years, whereas others have varying degrees of glucose intolerance for several years before the onset of persistent hyperglycemia.17 The diagnosis may be made only in adulthood even though the elevation in plasma glucose has been present for many years. Prospective testing indicates that in most patients the disease onset occurs in childhood or adolescence. In some patients, there may be a rapid progression to overt asymptomatic or symptomatic hyperglycemia necessitating therapy with an oral hypoglycemic drug or insulin. The presence of persistently normal plasma glucose levels in subjects with mutations in any of the known MODY genes is unusual, and the majority eventually experience diabetes (with the exception of many patients with glucokinase mutations; see later).

Although the exact prevalence of MODY is not known, current estimates suggest that MODY may account for 1% to 5% of all cases of diabetes in the United States and other industrialized countries.18 Several clinical characteristics distinguish patients with MODY from those with type 2 diabetes, including a prominent family history of diabetes in three or more generations, young age at presentation, and absence of obesity.

**Functional Effects of MODY Genes**

The identification of several genes associated with diabetes has provided a unique opportunity to characterize the pathophysiological mechanisms by which genetic mutations can lead to an increase in the plasma glucose concentration. All the susceptibility genes identified to date cause impaired insulin secretory responses to glucose, although the mechanisms differ.

**Glucokinase**

Glucokinase is expressed at its highest levels in the pancreatic beta cell and the liver. It catalyzes the transfer of phosphate from adenosine triphosphate (ATP) to glucose to generate glucose-6-phosphate (Fig. 29-2) (Figure Not Available). This reaction is the first rate-limiting step in glucose metabolism. Glucokinase functions as the glucose sensor in the beta cell by controlling the rate of entry of glucose into the glycolytic pathway (glucose phosphorylation) and its subsequent metabolism. In the liver glucokinase plays a key role in the ability to store glucose as glycogen, particularly in the postprandial state. Heterozygous mutations leading to partial deficiency of glucokinase are associated with MODY, and homozygous mutations resulting in complete deficiency of this enzyme lead to permanent neonatal diabetes mellitus. As predicted by the physiologic functions of glucokinase, the increase in plasma glucose concentrations seen in patients with this form of diabetes is due to a combination of reduced glucose-induced insulin secretion from the pancreatic beta cell and reduced glycogen storage in the liver after glucose ingestion.

**Liver-Enriched Transcription Factors: Hepatocyte Nuclear Factor-1, Hepatocyte Nuclear Factor-1, and Hepatocyte Nuclear Factor-4**

The transcription factors HNF-1, HNF-1, and HNF-4 play a key role in the tissue-specific regulation of gene expression in the liver.12 and are also expressed in other tissues including pancreatic islets, kidney, and genital tissues. HNF-1 and HNF-1 are members of the homeodomain-containing family of transcription factors, and HNF-4 is an orphan nuclear receptor.13,14 HNF-1, HNF-1, and HNF-4 make up part of an interacting network of transcription factors that function together to control gene expression during embryonic development and in adult tissues in which they are coexpressed. In the pancreatic beta cell, these transcription factors regulate the expression of the insulin gene as well as proteins involved in glucose transport and metabolism and mitochondrial metabolism, all linked to insulin secretion, and lipoprotein metabolism.15 The expression of HNF-1 is regulated at least in part by HNF-4. Persons with diabetes related to mutations in these genes have defects in insulin secretory responses to a variety of secretagogues, particularly glucose, that are present before the onset of hyperglycemia, suggesting that they represent the primary functional defect in the syndrome. Reduced glucagon responses to arginine have also been observed, suggesting that the pancreatic alpha cell is also involved in a broader pancreatic developmental abnormality.

**Insulin Promoter Factor-1**

IPF-1 is a homeodomain-containing transcription factor that was originally isolated as a transcriptional regulator of the insulin and somatostatin genes. It also plays a central role in the development of the pancreas as well as the regulation of expression of a variety of pancreatic islet genes including, besides insulin, glucokinase, islet amyloid polypeptide, and glucose transporter 2 genes. IPF-1 also appears to mediate glucose-induced stimulation of insulin gene transcription.16

**Neurogenic Differentiation 1 (NeuroD1/BETA2)**

The basic helix-loop-helix transcription factor NeuroD1/BETA2 was isolated on the basis of its ability to activate transcription of the insulin gene and is required for normal pancreatic islet development. Two patients with heterozygous mutations in NeuroD1 and diabetes have been described,16 and a third has been identified in an Icelandic population. Studies in other populations have failed to detect mutations in NeuroD1 even in subjects with a MODY phenotype. It therefore appears that mutations in NeuroD1 are a rare cause of MODY.

**Genetics of the Polygenic Forms of Type 2 Diabetes**

As alluded to earlier, the common polygenic form of type 2 diabetes has complex pathophysiology with genetic and environmental factors both playing a major role. The phenotypic manifestations of the disease are also complex, and subjects with established type 2 diabetes demonstrate abnormalities in a number of tissues that play a key role in the regulation of glucose metabolism. These include resistance to the action of insulin in muscle, fat, and liver; defects in insulin secretory responses from the pancreatic beta cell; and increases in hepatic glucose production. However, the primary defect or defects responsible for the development of the syndrome remain elusive and are likely not to be defined until more is known about the genes responsible for diabetes and the nature of the gene-environment interactions that are ultimately responsible for the development of the disorder in predisposed individuals.

Insulin resistance is present in individuals predisposed to type 2 diabetes before the onset of hyperglycemia, and this finding has been interpreted by some to indicate...
that insulin resistance is the primary abnormality that is responsible for the development of type 2 diabetes. However, defective beta cell function is also present before the onset of type 2 diabetes when IGT is present and in first-degree relatives of persons with type 2 diabetes who have completely normal plasma glucose concentrations. Thus, although there is still controversy about whether insulin resistance or abnormal insulin secretion represents the primary defect in type 2 diabetes, there is general consensus that both defects are present in essentially all subjects with the disorder, frequently from an early preclinical stage.

It is therefore evident that the identification and characterization of the genes responsible for type 2 diabetes will add an essential level to our understanding of the pathophysiology of the disorder. Unfortunately, progress in identifying these genes has been slow for reasons that are discussed in greater detail in the following.

The candidate gene approach has not been successful in identifying genes responsible for the common polygenic forms of type 2 diabetes. This approach involves identifying and then cloning genes known to be involved in pathways that regulate glucose metabolism, screening for variants in these genes, and assessing the frequency of the variants in cases and controls.

Because the candidate gene approach has not been productive in identifying the genes for the common forms of type 2 diabetes, linkage analysis has been applied. This approach involves defining regions of chromosomal DNA shared to excess by affected family members. Parents are genotyped at a particular marker, and the offspring are scored for sharing of zero, one, or two alleles inherited from their parents. Markers are genotyped in family members in the regions of polymorphic repeats called microsatellites or simple tandem repeats. Because in the case of type 2 diabetes (as well as other common polygenic human diseases), there is no prior knowledge of the gene defect, microsatellites at defined chromosomal locations are typed either in family members or in affected sibling pairs. In order to screen the whole genome, 300 to 400 microsatellites are genotyped in the study subjects at roughly 10-centimorgan intervals. This is generally adequate to define chromosomal regions sharing single gene defects. The evidence for linkage usually extends over a broad region, which may be 10 to 20 cM, and positional cloning strategies are then utilized to find the gene within this broad region.

This approach has demonstrated that genetic variation in the gene that encodes a ubiquitously expressed member of the calpain-like cysteine protease family, calpain-10 (CAPN10), is associated with increased risk for type 2 diabetes.

Until the observation was made on the grounds of genetic studies that calpains may be involved in the pathophysiology of diabetes, this possibility had not been considered. It is now critical that the role of calpains in regulating the metabolic pathways responsible for the maintenance of insulin secretion and insulin action be defined. Data available from the initial studies that have addressed these questions have begun to provide clues to the mechanisms that may be involved.

It appears that the effects of the at-risk polymorphisms are mediated by decreasing calpain-10 expression in relevant tissues. Thus, Baier and colleagues demonstrated that the polymorphism in calpain-10 that is responsible for the linkage to type 2 diabetes is associated with reduced muscle messenger ribonuclease acid (mRNA) and insulin resistance in Pima Indians, a population at extremely high risk for the development of type 2 diabetes. In order to simulate the physiologic effects of reduced calpain activity or expression, we have exposed rodent pancreatic islets or muscle strips to various calpain inhibitors for either 4 to 6 or 48 hours. The effects on pancreatic islets were dependent on the duration of the exposure. After 4 to 6 hours, inhibition of calpain activity was associated with enhanced glucose-induced insulin secretion that appeared to be mediated by effects on the secretion process itself. Exposure of mouse islets to calpain inhibitors of different structure and mechanism of action for 48 hours reversibly suppressed glucose-induced insulin secretion by 45% to 80%. Exposure of islets to other protease inhibitors (cathepsin B and proteasome) did not result in similar effects. The 48-hour incubation with calpain inhibitors also attenuated insulin secretory responses to the mitochondrial fuel-ketosocaproate and the Ca\(^{2+}\)-independent exocytosis trigger mastoparan. Glucose metabolism and intracellular calcium (Ca\(^{2+}\)), responses to glucose or ketosocaproate were also reduced, as was the exocytosis of insulin granules (measured by cell capacitance) in single beta cells. Thus, inhibition of calpain activity in islets attenuates insulin secretion, possibly by limiting the rate of glucose metabolism and exocytosis of insulin.

Exposure of rodent muscle strips to calpain inhibitors was shown to induce insulin resistance by impairing insulin-induced glucose transport and glycogen synthesis. Thus, these preliminary investigations suggest that calpain-sensitive pathways are present in pancreatic islets and muscle. Clearly, more detailed studies are needed to identify the specific calpain targets involved and to define their role in the normal control of insulin secretion and action and in the abnormalities in these processes that are responsible for the development of type 2 diabetes.

Why has it been so difficult to identify the genes responsible for the polygenic forms of type 2 diabetes, and what are the prospects for future progress? A perspective on the subject published by Cox emphasized that the experience to date gained from studies on CAPN16 is likely to provide a preview of the challenges that need to be overcome in the future.

The first important issue is that the CAPN16 variation implicated in susceptibility to type 2 diabetes was located in a noncoding rather than a coding sequence and leads to a significant reduction in skeletal muscle calpain-10 mRNA levels as well as insulin resistance. The latter functional result could not have been predicted from the genetic analyses and knowledge of the biology of calpains and could be obtained only from studies of the functional effects of CAPN16 variation in nondiabetic Pima Indians.

The second critical issue is that the mechanism for CAPN16 effects is complex and combinations of variants are more strongly associated with disease than individual polymorphisms. In the case of CAPN16, individuals heterozygous for the two different haplotypes that form the high-risk haplotype combination have approximately threefold increased risk compared with all other combinations of haplotypes and approximately eightfold increased risk compared with the lowest risk haplotype combination. Individuals homozygous for either of the haplotypes in the high-risk combination were not at increased risk of developing type 2 diabetes.

The third issue is that the physiologic mechanism or mechanisms by which genetic susceptibility loci cause disease may not be clear and may be difficult to determine. This is certainly the case with CAPN10 and with calpains in general, for which the primary physiologic function is unclear.

The fourth issue is that because multiple genetic mechanisms as well as environmental factors are involved in disease pathogenesis, replication may be difficult as the importance of specific genetic variations may differ in different populations.

In summary, therefore, as we move from studies of the monogenic forms of diabetes to studies that aim to understand the genetic, molecular, and physiologic basis of the polygenic forms of this condition, additional serious challenges must be met. However, in light of the complex nature of these disorders, understanding the genetic basis of these conditions is still the most likely pathway to solving the nature of these disorders.
INSULIN RESISTANCE AND THE RISK OF TYPE 2 DIABETES

Insulin Resistance

A substantial amount of data indicates that insulin resistance plays a major role in the development of glucose intolerance and diabetes. Insulin resistance is a consistent finding in patients with type 2 diabetes, and resistance is present years before the onset of diabetes. Prospective studies show that insulin resistance predicts the onset of diabetes.

The term insulin resistance indicates the presence of an impaired biologic response to either exogenously administered or endogenously secreted insulin. Insulin resistance is manifested by decreased insulin-stimulated glucose transport and metabolism in adipocytes and skeletal muscle and by impaired suppression of hepatic glucose output. Insulin sensitivity is influenced by a number of factors including age, weight, ethnicity, body fat (especially abdominal), physical activity, and medications. Insulin resistance is associated with the progression to IGT and type 2 diabetes, although diabetes is rarely seen in insulin-resistant persons without some degree of beta cell dysfunction. First-degree relatives of type 2 diabetics have insulin resistance even at a time when they are nonobese, implying a strong genetic component in the development of insulin resistance. There is also a strong influence of environmental factors on the genetic predisposition to insulin resistance and therefore to diabetes.

Obesity and Type 2 Diabetes

The association of obesity with type 2 diabetes has been recognized for decades. A close association between obesity and insulin resistance is seen in all ethnic groups and is found across the full range of body weights, across all ages, and in both sexes. A number of large epidemiologic studies showed that the risk for diabetes, and presumably insulin resistance, rises as body fat content increases from the very lean to the very obese, implying that the absolute amount of body fat has an effect on insulin sensitivity across a broad range. However, central (intra-abdominal) adiposity is more strongly linked to insulin resistance and a number of important metabolic variables, including plasma glucose, insulin, total plasma cholesterol and triglyceride concentrations, and decreased plasma high-density lipoprotein (HDL) cholesterol concentration, than is total adiposity. In addition, the effect of accumulation of abdominal fat on glucose tolerance is independent of total adiposity.

The reason for the relationship to intra-abdominal fat with abnormal metabolism is not clearly defined, but a number of hypotheses, which are not mutually exclusive, have been proposed. First, abdominal fat is more lipolytically active than subcutaneous fat, perhaps because of its greater complement of adrenergic receptors. In addition, the abdominal adipose store is resistant to the antilipolytic effects of insulin, including alterations in lipoprotein lipase activity, which leads to increased lipase activity and a greater flux of fatty acids into the circulation with the portal circulation receiving the greatest fatty acid load. The role of the liver in insulin resistance and hyperglycemia is discussed subsequently.

Skeletal Muscle Insulin Resistance

The primary site of glucose disposal after a meal is skeletal muscle, and the primary mechanism of glucose storage is through its conversion to glycogen. Studies using the hyperinsulinemic-euglycemic clamp technique have demonstrated that in insulin-resistant people with and without type 2 diabetes, there is a deficiency in the nonoxidative disposal of glucose related primarily to a defect in glycogen synthesis. Elevated free fatty acids (FFAs) predict the progression from IGT to diabetes. In the periphery, FFAs may not be markedly elevated because of efficient extraction by the liver and skeletal muscle. Thus, normal or minimally elevated FFA levels may not reflect the true exposure of fatty acids to peripheral tissues. Increases in fatty acid flux to skeletal muscle related to the increased visceral lipolysis have been implicated in the inhibition of muscle glucose uptake.

The Randle hypothesis, or the glucose-fatty acid cycle, was originally proposed to account for the ability of FFAs to inhibit muscle glucose utilization. Randle and colleagues demonstrated that fatty acids compete with glucose for substrate oxidation in isolated muscle. The increase in fatty acid metabolism leads to an increase...
in the intramitochondrial acetyl coenzyme A (CoA)/CoA and reduced nicotinamide adenine dinucleotide (NADH)/NAD⁺ ratios with subsequent inhibition of pyruvate dehydrogenase. The resulting increased intracellular mitochondrial (and cytosolic) citrate concentrations result in allosteric inhibition of phosphofructokinase, the key rate-limiting enzyme in glycolysis. Subsequent accumulation of glucose-6-phosphate would inhibit hexokinase II activity, resulting in an increase in intracellular glucose concentrations and decreased glucose uptake.

However, later studies have suggested that the primary effect of fatty acids, at least in the presence of raised insulin levels, is a decrease in glucose transport as measured by a reduction in the rate of accumulation of intracellular glucose and glycogen using 13C and 13N nuclear magnetic resonance spectroscopy. In normal subjects, elevated fatty acids, achieved by infusion of triglyceride emulsions and heparin (to activate lipoprotein lipase), resulted in a fall in intracellular glucose and glucose-6-phosphate concentrations that preceded the fall in glycogen accumulation. These results challenge the Randle hypothesis (which predicts a rise in intracellular glucose-6-phosphate concentrations) as the basis of the reduction in insulin sensitivity seen with elevated fatty acids. Similar decreases in glucose transport have been seen in patients with type 2 diabetes and lean, normoglycemic insulin-resistant offspring of type 2 diabetic individuals. These studies also found a decrease in the activity of phosphatidylinositol 3-phosphate kinase (PI 3-kinase) and increased protein kinase C-theta activity that may, in part, mediate the effect of high FFAs.

It has been found that insulin-stimulated glucose uptake is inversely related to the amount of intramuscular triglycerides. A strong correlation between intramuscular triglyceride concentration and insulin resistance has been demonstrated by evaluating intramuscular triglyceride with biopsy, computed tomographic scanning, and magnetic resonance imaging measurements. The latter method has been a valuable addition because the magnetic resonance signal can distinguish intramyocellular from extramyocellular fat and demonstrates the increased triglyceride accumulation within the myofiber itself. First-degree relatives of type 2 diabetics have an increase in intramyocellular fat, and in this group there is also a correlation with insulin resistance.

The mechanism for accumulation of triglyceride in the skeletal muscle of obese and insulin-resistant individuals is probably related to the mismatching of FFA uptake and oxidation. During resting postabsorptive conditions, about 30% of fatty acid flux in the plasma pool is accounted for by oxidation, with the remaining 70% of flux recycled into triglyceride, indicating a physiologic reserve that exceeds immediate tissue needs for oxidative substrates. The equilibrium between oxidation and reesterification within muscle is paramount in determining fatty acid storage within tissue. The uptake, transport, and metabolism of fatty acids are highly regulated processes (Fig. 29-6), and alteration of the balance between uptake and oxidation in skeletal muscle leads to increased intramyocellular triglyceride. The increased lipolysis associated with obesity provides an increased amount of FFA presented to muscle.

Increased muscle triglyceride content is not invariably linked to insulin resistance because exercise training is associated with increased muscle triglyceride content, and chronic exercise increases insulin sensitivity as well as the capacity for fatty acid oxidation.

There is a direct correlation between heart-type FABP content and insulin resistance in obese and insulin-resistant individuals. Some, but not all, studies have shown a decrease in heart-type FABP in insulin-resistant humans.

Fatty Acid Metabolism in Skeletal Muscle, Relationship to Insulin Sensitivity

The uptake of fatty acid from the serum, where it is mostly bound to albumen, is mediated by at least three families of proteins: fatty acid translocase, plasma membrane fatty acid binding protein (FABP-pm), and fatty acid transport protein. The levels of the putative transport proteins are regulated by exercise, correlated with body weight at least in men, and can be modulated by insulin infusion.

FABPs are capable of binding multiple hydrophobic ligands, including fatty acids, eicosanoids, and retinoids with high affinity. FABPs are thought to facilitate uptake of fatty acids and promote subsequent intracellular transport to subcellular organelles. There is a direct correlation between heart-type FABP content and oxidative capacity observed during development and among different muscle types. In mice with a disruption of the heart-type or adipocyte isoform of FABP, plasma fatty acid concentrations were significantly elevated while plasma glucose was decreased, suggesting a key role in normal regulation in fatty acid oxidation. Some, but not all, studies have shown a decrease in heart-type FABP in insulin-resistant humans.

Carotid palmitoyltransferase I (CPT-I) has been the subject of intense scrutiny for many years because of its central role in the balance between mitochondrial glycerol and fatty acid metabolism, primarily because of inhibition of mitochondrial fatty acid uptake by malonyl CoA. A specific isoform contributes 97% of the CPT-I in muscle and has 100-fold lower sensitivity to inhibition by malonyl CoA. This lower sensitivity to malonyl CoA inhibition suggests that the levels of CPT-I itself may be important in the balance of uptake and oxidation of fatty acids. Evidence for this in skeletal muscle stems from the finding that, as with other fatty acid-oxidizing enzymes, muscle CPT-I mRNA is regulated by PPAR activators, fat feeding, and exercise in rodents and is inversely correlated with obesity in humans.

Long-chain fatty acids, after passing through the inner mitochondrial membrane as acylcarnitines, are metabolized at the surface of the inner mitochondrial membrane by CPT-II and the long chain-specific oxidation system consisting of very-long-chain acyl CoA dehydrogenase (VLCAD) and the trifunctional protein (TFP) oxidation complex. Transfer of the acyl chain from carnitine to CoA catalyzed by CPT-II is followed by one cycle of oxidation catalyzed by VLCAD and TFP to yield a chain-shortened acyl CoA that can recycle through the same oxidation system. In actuality, four different acyl CoA dehydrogenase enzymes catalyze the initial dehydrogenation of straight-chain fatty acids in mitochondria. Three of themshort-chain acyl CoA dehydrogenase (SCAD), medium-chain acyl CoA dehydrogenase (MCAD), and long-chain acyl CoA dehydrogenase (LCAD) are soluble enzymes located in the mitochondrial matrix as homotrimers. A fourth, VLCAD, is attached to the inner membrane as a homodimer. Their names derive from the length of the fatty acids that they process. VLCAD and LCAD shorten the long-chain fatty acids into medium-chain fatty acids that can then be processed by MCAD and SCAD. The SCAD, MCAD, and LCAD monomers share a high degree of homology between them but not with VLCAD. At least some of these enzymes can be regulated in humans during exercise training.

Uncoupling protein-1 (UCP1) is clearly related to the uncoupling of oxidative phosphorylation in brown adipose tissue. UCP2 and UCP3 have structural similarities to UCP1, but it is not clear that they are actually uncouplers of oxidative phosphorylation. Newer members of the family, brain mitochondrial carrier protein 1 (BMCP1) and UCP4, have an even more distant sequence relationship. BMCP1 and UCP4 are predominantly expressed in neural tissues, namely the brain. UCP3 mRNA is found primarily in skeletal muscle and in brown adipose tissue. UCP2 has a ubiquitous tissue distribution. UCP2 and UCP3 mRNAs have been correlated with different physiologic states, and numerous studies indicate that expression of UCP2 and UCP3 is stimulated by thyroid hormones and in the presence of high levels of fatty acids. In humans, the levels of UCP2 and UCP3 mRNAs were up-regulated by a high-fat diet and the up-regulation was more
pronounced in humans with high proportions of type II A fibers. In a small study, exercise training in humans increased mitochondrial oxidative capacity but did not change UCP2 or UCP3 levels. Obesity itself was shown to be positively correlated with a splice isoform of UCP3. A unique polymorphism in the promoter region of UCP3 correlated with the expression of UCP3 in skeletal muscle.

Glucose Influence on Fatty Acid Metabolism

An emerging concept that could couple the increased fatty acid flux into skeletal muscle with impaired insulin action is the central role of malonyl CoA in regulating fatty acid and glucose oxidation (Fig. 29-7). Malonyl CoA is an allosteric inhibitor of CPT-I, the enzyme that controls the transfer of long-chain fatty acyl CoAs into the mitochondria. Even in insulin-resistant skeletal muscle, glucose uptake into the skeletal muscle is higher, especially at the elevated levels of glucose found in type 2 diabetes. The glucose is shunted toward the glycolytic pathway, generating acetyl CoA that can be converted to malonyl CoA in the cytoplasm by the action of the highly regulated enzyme acetyl CoA carboxylase (ACC).

In humans an infusion of insulin and glucose at a high rate leads to increases in the concentration of malonyl CoA in skeletal muscle and to decreases in whole-body and, presumably, muscle fatty acid oxidation. In the presence of elevated glucose and insulin levels, the tricarboxylic acid (TCA) cycle is activated, resulting in an increase in citrate in the cytoplasm through increased malate cycling in the mitochondria. The increased citrate is converted to acetyl CoA through citrate lyase and thus provides an indirect substrate for ACC. Citrate also allosterically activates ACC and makes ACC a better substrate for phosphatases that activate the enzyme. ACC is also regulated by a phosphorylation-dephosphorylation cycle, with adenosine monophosphate (AMP)-dependent protein kinase an important kinase, which inhibits both ACC basal activity and ACC activation by insulin. ACC then generates malonyl CoA that in turn allosterically inhibits CPT-I residing on the outer mitochondrial membrane, inhibiting uptake of acetyl CoA. The resulting buildup of long-chain acyl CoAs and diacylglycerols is proposed to activate one or more protein kinase C isoforms or other lipid-activated proteins, resulting in insulin resistance. Support for this hypothesis is the finding that exercise, which activates AMP-dependent kinase, inactivates ACC, lowers intracellular long-chain acyl CoA levels, and has an acute insulin-sensitizing effect.

Hyperinsulinemia and Insulin Resistance

Hyperinsulinemia per se has been proposed to cause insulin resistance. Elevated concentrations of insulin can cause insulin resistance by down-regulating insulin receptors and desensitizing postreceptor pathways. Del Prato and associates showed that 24 and 72 hours of sustained physiologic hyperinsulinemia in normal individuals specifically inhibited the ability of insulin to increase nonoxidative glucose disposal in association with an impaired ability of insulin to stimulate glycogen synthase activity. Suppression of insulin secretion in obese, insulin-resistant people results in increased insulin sensitivity.

Insulin Signaling

Insulin signaling is initiated through the binding to and activation of its cell-surface receptor and initiates a cascade of phosphorylation and dephosphorylation events, second messenger generation, and protein-protein interactions that result in the diverse metabolic events in nearly every tissue (Fig. 29-8) (Figure Not Available). The insulin receptor consists of two insulin-binding subunits and two catalytically active subunits that are disulfide linked into an 22 heterotetrameric complex. Insulin binds to the extracellular subunits, activating the intracellular tyrosine kinase domain of the subunit. One receptor subunit phosphorylates its partner on specific tyrosine residues that may have distinct functions such as stimulation of intercellular association of signaling molecules such as Shc and Grb, members of the insulin receptor substrate family (IRS1, 2, 3, 4), Shc adapter protein isoforms, and SIRP (signal regulatory protein) family members, Gab-1, Cbl, CAP, and APS; stimulation of mitogenesis; and receptor internalization.

Insulin receptor subunit has also been shown to undergo serine-threonine phosphorylation, which may decrease the ability of the receptor to autophosphorylate. The activities of a number of protein kinase C isoforms that catalyze the serine or threonine phosphorylation of the insulin receptor are elevated in animal models of insulin resistance and in insulin-resistant humans. Interventions that decrease serine phosphorylation of the insulin receptor result in increased insulin signaling. Termination of the insulin signaling event occurs by internalization and dephosphorylation of the receptor by protein tyrosine phosphatases. Increased activity of protein tyrosine phosphatase can attenuate insulin signaling. Two protein tyrosine phosphatases that have been shown to negatively regulate insulin signaling, PTP1B and LAR (leukocyte antigen related), have been reported to be elevated in insulin-resistant patients. Conversely, disruption of PTP1B in mice resulted in a marked increase in insulin sensitivity and resistance to diet-induced obesity.

Mutations in the insulin receptor are associated with rare forms of insulin resistance. These mutations affect insulin receptor number, splicing, trafficking, binding, and phosphorylation. The affected patients demonstrate severe insulin resistance, manifest as clinically diverse syndromes including the type A syndrome, leprechaunism, Rabson-Mendenhall syndrome, and lipodystrophic diabetes.

Downstream Events Following Insulin Receptor Phosphorylation

IRSs act as multifunctional docking proteins activated by tyrosine phosphorylation. The IRS proteins have multiple functional domains, including Pleckstrin homology (PH) and phosphotyrosine binding (PTB) and SH domains that interact with other proteins to mediate the insulin signaling events. Disruption of IRS1 in mice resulted in mild insulin resistance and growth retardation, whereas disruption of IRS2 resulted in beta cell failure and subsequent insulin resistance.

P3-kinase, which is regulated by interaction with IRS proteins, is necessary but not sufficient for the stimulation of the glucose transporter GLUT4-mediated increase in glucose transport in insulin-sensitive tissues. In addition, inhibition of PI 3-kinase activity with the fungal inhibitor wortmannin inhibited insulin-stimulated glucose uptake, glycogen synthesis, triglyceride accumulation, protein synthesis, and modulation of gene expression. P3-kinase generates 3,4,5-phosphoinositols, which activates several PIP3 (phosphatidylinositol-3,4,5 triphosphate)-dependent enzymes.

Figure 29-8 (Figure Not Available) Insulin signaling. The insulin receptor is autophosphorylated on multiple tyrosine residues, allowing the docking and activation of multiple signaling molecules, which mediate the increases in glucose uptake and metabolism as well as changes in protein and lipid metabolism. (From Satelli AR, Kahn CR. Insulin signaling and the regulation of glucose and lipid metabolism. Nature 2001, 414:798906.)

serine-threonine kinases, such as PI-dependent protein kinases 1 and 2, which in turn activates Akt, salt- and glucocorticoid-induced kinases. Protein kinase C, wortmannin-sensitive and insulin-stimulated serine kinase, and others.

Akt kinase (also known as protein kinase B) exists as three distinct isoforms that are activated by phosphorylation on specific threonine and serine residues. Activated Akt has the ability to phosphorylate proteins that regulate lipid synthesis, glycogen synthesis, protein synthesis, and apoptosis. Disruption of Akt2 resulted in insulin resistance and diabetes in mice. Several investigators have examined the role of PI 3-kinase and Akt in individuals with insulin resistance. Studies have shown a decrease in IRS-associated PI 3-kinase and Akt activity in insulin-resistant skeletal muscle; however, in some patients with reduced PI 3-kinase activity there was normal activation of Akt.

A primary effect of insulin is to stimulate translocation of the glucose transporter GLUT4 from an intracellular pool to the surface of cells, primarily in skeletal muscle and adipose tissue and heart. The mechanism by which the signaling pathways converge on the intracellular GLUT4-containing vesicles to cause GLUT4 translocation is not well understood. It appears that the number of glucose transporters in skeletal muscle of insulin-resistant persons is not changed but the ability of
insulin to effect this translocation is disrupted.

**Glucocorticoid-Induced Insulin Resistance**

Cushing's syndrome and exogenous glucocorticoid treatment have long been known to induce significant insulin resistance in humans. The exact mechanism is unknown, but it is associated with redistribution of fat from the periphery to the central compartment. Elevations in triglyceride and FFA levels also occur. At a molecular level, dexamethasone has differential effects on the proteins involved in the early steps in insulin action in liver and muscle. In both tissues, dexamethasone treatment results in a reduction in insulin-stimulated IRS1-associated PI 3-kinase, which may play a role in the pathogenesis of insulin resistance at the cellular level in these animals.

**Tumor Necrosis Factor**

Studies in humans and animal models of obesity have identified changes in the expression and activity of key molecules involved in the insulin signaling pathway. Decreases in the number and the kinase activity of insulin receptors[157] and impairment in the activation of IRS1, PI 3-kinase,[158][159] and protein kinase B[160] have been observed. Although the basis for this is, in general, unknown, a tumor necrosis factor (TNF)-mediated decrease in the activity of the intermediate steps in the insulin signaling cascade has been proposed. TNF-α made and secreted by adipocytes, is elevated in a variety of experimental models of obesity.[161] The kinase activity of the insulin receptor in rats[162] or in 3T3-L1 adipocytes[163] treated with TNF-α was reduced, possibly by increased serine phosphorylation.[164] Fat-fed mice with genetic ablation of TNF-α production had increased kinase activity of the insulin receptor compared with control mice and demonstrated increased insulin sensitivity.[165] In addition, rats treated with neutralizing antisera or soluble TNF receptors demonstrated an amelioration of their insulin resistance. As described later, other interventions to decrease TNF-α action result in increased insulin sensitivity.

**Glucotolerance, Glucosamine**

Hyperglycemia is a primary factor in the development of the complications of the disease, and decreases in average blood glucose have a profound effect to prevent complications in both type 1 and type 2 diabetes.[166] Hyperglycemia itself can cause insulin resistance. In Pima Indians, the level of fasting glycaemia is the primary determinant of insulin sensitivity.[167] The defect is primarily in skeletal muscle[168] and is related to the degree of hyperglycemia.

Entry of glucose into the cell results in its phosphorylation to glucose-6-phosphate, which has multiple metabolic fates. The enzyme glucose:fructose-6-phosphate amidotransferase (GFAT) carries out the rate-limiting step of the hexosamine pathway.[169] Evidence suggests that the hexosamine pathway underlies the defect in glucose utilization associated with hyperglycemia. Hexosamines, such as glucosamine, when incubated with adipose tissue, induce insulin resistance in fat cells[170] and in skeletal muscle.[171] Infusion of glucosamine into rats resulted in a dose-dependent increase in insulin resistance of skeletal muscle.[172] Finally, transgenic mice that overexpress GFAT specifically in skeletal muscle acquired severe insulin resistance.[173] By a pathway that is unclear, glucosamine overproduction resulted in a disruption of the ability of GLUT4 to translocate to the cell surface. Through its anti-insulin action, the hexosamine pathway has been hypothesized to be a glucose sensor that allows the cell to sense and adapt to the prevailing level of glucose.[174]

**Insulin Resistance and Lipodystrophy Associated with Human Immunodeficiency Virus Infection**

A syndrome with many of the clinical and metabolic features of insulin resistance is increasingly being recognized in patients with human immunodeficiency virus (HIV) infection. A novel form of lipodystrophy is observed in certain of these patients. This form of lipodystrophy is associated with significant insulin resistance and type 2 diabetes, dyslipidemia, elevated total and low-density lipoprotein (LDL) cholesterol and suppressed HDL cholesterol concentrations, and a susceptibility to lactic acidemia.[175][176][177][178]

Epidemiologic studies have associated this syndrome with previous or current treatment with antiretroviral protease inhibitors or nucleoside reverse transcriptase inhibitors. There is also an association with increased age.[179] Other potential contributing factors are male sex, diagnosis of the acquired immunodeficiency syndrome, responsiveness to antiretroviral treatment, and increases in CD4 T-cell counts. An increased emphasis has been placed on the role of protease inhibitors in the pathogenesis of the syndrome. Administration of these drugs or ritonavir to normal subjects caused increases in plasma triglyceride and very-low-density lipoprotein (VLDL) cholesterol and decreased plasma HDL cholesterol levels.[180][181] Indinavir administration for 4 weeks resulted in small increases in serum glucose and insulin levels and decreased insulin-mediated glucose disposal as assessed with a hyperinsulinemic, euglycemic clamp. In this study there were no changes in lipoprotein, triglycerides, or FFA levels.[182][183]

The metabolic basis of the molecular deficit is not clear. A number of protease inhibitors can inhibit glucose transport in vitro and in vivo, and there is evidence for a direct interaction with the GLUT4 glucose transporter[184] that could inhibit glucose uptake specifically in insulin-responsive tissue. Mitochondrial abnormalities have been described in subcutaneous adipose tissue biopsy specimens of HIV-infected patients with lipodystrophy compared with those without the syndrome.[185] A direct effect of protease inhibitors on differentiation of adipocytes has also been described.[186][187] At present, the precise mechanism for the lipodystrophy associated with HIV infection is not known.

**Mechanisms of Reducing Insulin Resistance**

The most effective measures to improve insulin sensitivity are weight loss and exercise. Both modalities are effective and can be additive in their ability to improve insulin action. Later in this chapter, the role of these interventions in the treatment of patients with type 2 diabetes is discussed. The scientific basis and molecular mechanisms responsible for the improvements in insulin sensitivity seen with these interventions are now summarized.

**Mechanisms for Improved Insulin Sensitivity with Weight Loss**

Weight loss can be a highly effective treatment for overweight patients with type 2 diabetes and other cardiovascular risk factors, and indeed it is advocated as the first line of therapy. Weight loss may also play a role in the prevention of type 2 diabetes.[188][189] In overweight patients with type 2 diabetes, weight loss can reduce hepatic glucose production, insulin resistance, and fasting hyperinsulinemia and can improve glycemic control. Weight loss in type 2 diabetes is also associated with a reduction in blood pressure and an improvement in the lipid profile. These benefits can occur with as little as 5% to 10% weight loss.[190][191][192] Indinavir administration for 4 weeks resulted in small increases in serum glucose and insulin levels and decreased insulin-mediated glucose disposal as assessed with a hyperinsulinemic, euglycemic clamp. In this study there were no changes in lipoprotein, triglycerides, or FFA levels.[193][194]

One possible mechanism for improvements in insulin sensitivity through weight loss may be effects on the pattern of muscle fatty acid metabolism and the accumulation of lipid within muscle. In this context, it would be important to know whether alterations in the pathways of fatty acid utilization in skeletal muscle represent primary defects in obese individuals or arise secondarily, after an individual has become obese. This question cannot be answered by cross-sectional comparisons of lean and obese subjects. One prospective clinical study indicated that lower rates of lipid oxidation were a predisposing factor for greater weight gain,[195] and collateral studies implicated altered skeletal muscle enzymes activities in impaired lipid oxidation.[196][197] A reduced reliance on lipid oxidation has also been identified as a risk factor for weight regain after weight loss.[198] These data raise the possibility that a potential impairment in the capacity for lipid oxidation may be a primary determinant of obesity. Weight loss can markedly improve insulin-resistant glucose metabolism in skeletal muscle. When the patient's response indicates a substantial acquired or secondary component of obesity-related insulin-resistant glucose metabolism, it is important to determine whether weight loss can modulate patterns of skeletal muscle metabolism of fatty acids, including the content of fat within muscle.

**Mechanisms for Improved Insulin Sensitivity with Exercise**

Exercise is clearly effective in increasing insulin sensitivity in animals and in humans. There appear to be two separable but related effects of exercise on insulin action. A single bout of exercise can result in an acute increase in insulin-independent glucose transport measurable during and for a relatively short period after exercise.[199][200][201][202][203] Like insulin, exercise and muscle contractions increase glucose transport by translocation of intracellular GLUT4 glucose transporters to
Acute Exercise

The signaling pathway leading to the exercise-induced increase in glucose transporter translocation and glucose transport is unknown, although there is ample evidence that the pathway is independent of the insulin-stimulated, receptor-mediated pathway. The effect of exercise and contractions on translocation and transport is additive to the maximal effect of insulin. Insulin-stimulated glucose transport in muscle is inhibited by specific inhibitors of PI 3-kinase, such as wortmannin, whereas transport or translocation stimulated by muscle contractions is insensitive to these inhibitors. Stimulation of muscle contractions in situ and exercise do not increase insulin receptor phosphorylation or tyrosine kinase activity, IRS phosphorylation, or PI 3-kinase activity. In addition, in many insulin-resistant states the acute exercise-stimulated (but not insulin-stimulated) glucose transport and GLUT4 translocation is normal. This has been demonstrated in the obese, insulin-resistant Zucker rat and in type 2 diabetic patients. Finally, hypoxia, a stimulus for glucose transport that is also independent of the insulin receptor-mediated pathway, is effective in increasing glucose transport in muscle strips from obese, insulin-resistant individuals and in patients with type 2 diabetes.

The acute effect of exercise and hypoxia may be mediated by AMP-dependent protein kinase (AMPK). AMPK is thought to be a sensor of intracellular energy stores and is activated by increases in intracellular AMP. A stable AMP analogue, 5-amino-4-imidazole carboxamide ribotide (ZMP), can be generated intracellularly from 5-aminoimidazole-4-carboxamide ribonucleoside (AICAR) and can activate AMPK in cells, leading to increased phosphorylation of known substrates for AMPK including 3-hydroxy-3-methylglutaryl CoA reductase, acetyl-CoA carboxylase, and creatine kinase. Treatment of incubated skeletal muscle with AICAR resulted in increased glucose uptake and glucose transporter translocation. Similarly, the inclusion of 2 mM AICAR in the perfusate of the rat hindlimb resulted in inactivation of ACC, decreases in malonyl CoA levels and a twofold increase in glucose uptake. The euglycemic clamp technique has been used in conscious rats to demonstrate that infusion of AICAR resulted in a more than twofold increase in glucose utilization. Uptake of the glucose analogue 2-deoxyglucose was also increased twofold in vivo in soleus and gastrocnemius muscles. As with previous studies, this uptake was not associated with PI 3-kinase activation, again indicating a separate pathway from that of insulin.

A second effect of exercise, which becomes evident as the acute effect on glucose transport reverses, is a large increase in the sensitivity of glucose transport to stimulation by insulin. This effect is due to translocation of more GLUT4 glucose transporters to the cell surface for any given dose of insulin. As with the acute stimulation of transport by exercise, the cellular mechanisms leading to enhanced translocation in response to submaximally effective stimuli are unknown. However, several studies have shown that steps in the insulin signaling cascade leading to activation of PI 3-kinase are not enhanced after a bout of exercise. There is no change in insulin binding to its receptor, insulin stimulation of receptor tyrosine kinase activity, increase in insulin-stimulated tyrosine phosphorylation of IRS1 or PI 3-kinase activity associated with IRS1.

Exercise Training

Exercise training also results in increases in insulin sensitivity and can delay or prevent the onset of type 2 diabetes in those at high risk. Using the hyperinsulinemic-euglycemic clamp, Perseghin and co-workers compared exercise training for 45 minutes on a stair-climbing machine 4 days per week for 6 weeks in normal insulin-sensitive subjects and a group of high-risk, insulin-resistant relatives of type 2 diabetics. A 100% increase in insulin sensitivity was seen in both groups without a significant change in body weight. Interestingly, the higher basal and glucose-stimulated insulin release seen in the insulin-resistant subjects was not altered after exercise training. The effect of exercise training on insulin sensitivity has been proposed to be due to up-regulation of glucose transporter number, changes in capillary density, and increases in the number of red, glycolytic (type Ila) fibers.
Mechanisms That Link Cardiovascular Disease and Insulin Resistance: Syndrome X

Myocardial infarction, stroke, and nonischemic cardiovascular disease are the cause of death in up to 80% of individuals with type 2 diabetes. Independent of other risk factors, type 2 diabetes increases the risk for cardiovascular morbidity and mortality but also provides a synergistic interaction with other risk factors such as smoking, hypertension, and dyslipidemia. In a Finnish population, diabetes increased the risk for myocardial infarction fivefold. and insulin resistance as measured by elevated fasting insulin levels increased the risk of death from heart disease. Women are particularly vulnerable to the cardiovascular effects of type 2 diabetes as they appear to lose the protective effects of estrogen in the premenopausal period.

A constellation of metabolic derangements that are frequently seen in patients with insulin resistance and type 2 diabetes are individually associated with an increased risk of cardiovascular disease. These patients have been variously designated as having syndrome X, the dysmetabolic syndrome, hypertension, obesity, noninsulin-dependent diabetes mellitus (NIDDM), dyslipidemia, and atherosclerotic cardiovascular disease (HONDRA); or the “deadly quartet.” The syndrome has also been associated with easily oxidized, small LDL particles; heightened blood clotting activity (e.g., increased plasminogen activator inhibitor 1); and elevated serum uric acid concentration. The central abnormality associated with syndrome X has been proposed to be insulin resistance. Some of the abnormalities have also been proposed to contribute to insulin resistance.

Perhaps the overriding risk factor for coronary artery disease in insulin resistance and type 2 diabetes is the associated dyslipidemia. The profile includes hypertriglyceridemia, low plasma HDL, and small, dense LDL particle concentrations. The percentage of men with type 2 diabetes with abnormal cholesterol levels is not different from that of nondiabetic men. However, diabetic women have nearly double the rate of hypercholesterolemia and greater changes in other lipid parameters that increase the risk for cardiovascular disease (Fig. 29-9). The physiologic basis for this abnormal lipid profile appears to be overproduction of apolipoprotein B-containing VLDL particles. The apolipoprotein B production by the liver is primarily post-translational and augmented by insulin and by the increased availability of FFAs in the portal circulation, probably as a result of increased lipolysis in the visceral adipose tissue. Part of the post-translational regulation may be due to insulin and fatty acid-mediated increases in microsomal triglyceride transfer protein levels that catalyze the transfer of lipids to apolipoprotein B and decrease the ubiquination-dependent degradation of apolipoprotein B.

The overproduction of VLDL triglyceride results in increased transfer of VLDL triglyceride to HDL particles in exchange for HDL cholesterol esters mediated by the cholesterol ester transfer protein. The triglyceride-rich HDL is hydrolyzed by hepatic lipase, which results in the generation of small HDL, which is degraded more readily by the kidney, resulting in low HDL levels in serum. Cholesterol ester transfer proteinmediated exchange of VLDL triglyceride for LDL cholesterol esters and subsequent triglyceride hydrolysis by hepatic lipase probably result in the generation of the small, dense LDL particles found in insulin-resistant subjects.

The increased risk of heart disease in patients with diabetes has prompted the recommendation that individuals with diabetes be treated for their dyslipidemia as aggressively as individuals who have had a previous myocardial infarction. In addition, individuals with the metabolic syndrome of insulin resistance and obesity are considered to be in a higher risk category and should also be aggressively treated in order to lower lipids.

Hypertension and overt diabetes double the risk of cardiovascular disease. Defects in vasodilatation and alterations in blood flow may provide a link to hypertension in insulin-resistant subjects. The normal vasodilatory response of insulin is disrupted in obese, insulin-resistant, and diabetic individuals, perhaps through inability of insulin to increase the production of the potent vasodilator nitric oxide by endothelial cells. The defect may be magnified by increases in plasma FFAs. Other proposed mechanisms for insulin resistance leading to hypertension are the activation of the sympathetic nervous system by insulin and the intrinsic ability of insulin to cause salt and water reabsorption in the kidney resulting in expanded plasma volume.

Hypertension itself, independent of other risk factors, has been associated with the propensity to become diabetic. A prospective cohort study found that type 2 diabetes was nearly 2.5 times more likely to develop in subjects with hypertension than in subjects with normal blood pressure. A possible mechanism is that an intrinsic defect in vasodilatation may contribute to insulin resistance by decreasing the surface area of the vasculature perfusing skeletal muscle, decreasing the efficiency of glucose uptake. Conversely, vasodilatory agents may improve glucose uptake and may even prevent the onset of diabetes, as has been observed with angiotensin-converting enzyme inhibitor therapy.

Several factors involved in clotting and fibrinolysis, including fibrinogen, factor VII, and plasminogen activator inhibitor 1 (PAI-1), have been shown to be increased in individuals with insulin resistance. PAI-1 expression in hepatocytes, endothelial cells, and adipose tissue, and insulin-sensitizing thiazolidinediones decreased PAI-1 activity.

Upper body rather than lower body obesity (the apple rather than the pear shape) is highly correlated with insulin resistance and risk for type 2 diabetes. Thus, the anatomic distribution of fat, rather than the overall degree of obesity, appears to determine risk for the metabolic syndrome. The reported association between increased abdominal (upper body) fat and an increased risk of coronary heart disease is related to visceral fat, for which the waist-to-hip ratio is a convenient index. A waist-to-hip ratio greater than 1.0 in men and 0.8 in women indicates abdominal obesity. Recently the National Cholesterol Education Program (NCEP) has suggested that a waist circumference >40 inches in men and >35 inches in women is a marker for the metabolic syndrome.
The Role of Increased Hepatic Glucose Production in the Hyperglycemia of Type 2 Diabetes and Impaired Glucose Tolerance

The disposal of glucose after meals depends on the ability of insulin to increase peripheral glucose uptake and simultaneously decrease endogenous glucose production. Although studies have suggested that the kidney can contribute up to 25% of endogenous glucose production, the defect in type 2 diabetes is primarily in defective regulation of glucose production from the liver [hepatic glucose output (HGO)]. Two routes of glucose production by the liver are glycolysis from stored glycogen and gluconeogenesis from two- and three-carbon substrates derived primarily from skeletal muscle. Under different conditions and at different times postprandially, the contribution of each of these to maintenance of glucose levels may vary. Using $^{13}$C nuclear magnetic resonance spectroscopy combined with measurement of wholebody glucose production in normal human subjects at different intervals after fasting, it was found that gluconeogenesis accounted for 50% to 96% of glucose production with the proportion increasing with increasing duration of fasting.

The production of glucose by the liver is regulated primarily by the relative actions of glucagon and insulin to activate or suppress glucose production, respectively, although the nervous system and glucose autoregulation of hepatic glucose production probably play less important roles. The ability of insulin to reduce HGO is an important mechanism for maintaining normal glucose tolerance. Under normal circumstances, insulin suppresses up to 85% of glucose production in normal individuals by directly inhibiting gluconeogenesis, especially at lower insulin concentrations. Under circumstances in which gluconeogenesis is enhanced by glucagon, the effects of insulin to suppress hepatic glucose production may be even greater. Glucagon increases gluconeogenesis by activation of the classical protein kinase cascade involving cyclic AMP (cAMP)-dependent protein kinase and phosphorylase and also increases gluconeogenesis in part by increasing the transcription of phosphoenolpyruvate carboxykinase through the cAMP response element binding protein.

Insulin decreases in endogenous glucose production by both direct and indirect mechanisms (Fig. 29-10). In its direct action, portal insulin suppresses glucose production by inhibiting gluconeogenesis through an increase in phosphodiesterase activity or changes in the assembly of protein phosphatase complexes. Insulin can also directly suppress gluconeogenesis by inhibiting the activation of phosphoenolpyruvate carboxykinase transcription through the insulin-dependent phosphorylation of the forkhead transcription factor, sequestering it in the cytoplasm.

The indirect or peripheral effect of insulin in controlling glucose production by the liver is twofold. First, insulin has a profound effect on decreasing glucagon secretion by the alpha cell of the pancreas through systemic and paracrine effects. The decrease in glucagon secretion decreases the activation of glycogenolysis and gluconeogenesis. The second important peripheral action of insulin is decreasing FFA levels by suppressing lipolysis. FFAs increase hepatic glucose production by stimulating gluconeogenesis. When the reduction in plasma FFAs during a hyperinsulinemic clamp was prevented by infusion of triglyceride emulsions with heparin (which results in increased FFA levels by activation of lipoprotein lipase), insulin-mediated suppression of HGO was reduced. The suppression of glucagon secretion and decrease in FFA delivery to the liver are additive in reducing liver glucose production.

Hepatic insulin resistance plays an important role in the hyperglycemia of type 2 diabetes, and the impairment in suppression of HGO appears to be quantitatively similar to or even larger than the defect in the stimulation of peripheral glucose disposal. There is a direct relationship between increases in HGO and fasting hyperglycemia (Fig. 29-11). Insulin-mediated suppression of HGO is impaired at both low and high plasma insulin levels in type 2 diabetic patients. Treatment of patients with metformin, which suppress hepatic glucose production, results in improvements in glucose tolerance.

Alterations in both the direct and indirect effects of insulin in type 2 diabetic patients appear to play a role in the elevation in hepatic glucose production. Defects in the direct effect of insulin to suppress hepatic glucose production that have been demonstrated in humans appear to be due to a large rightward shift in the steep dose-response curve for insulin's inhibition of gluconeogenesis. However, peripheral insulin resistance may play the bigger role in elevated hepatic glucose production in type 2 diabetes. The resistance of adipose tissue, especially visceral fat, to suppression of lipolysis by insulin is responsible for part of the inability of insulin to suppress hepatic glucose production by the indirect route, resulting in enhanced gluconeogenesis. In addition, the suppression of glucagon levels in humans with insulin resistance may be impaired, again leading to an increase in endogenous glucose production.
INSULIN SECRETION AND TYPE 2 DIABETES

Normal insulin secretory function is essential for the maintenance of normal glucose tolerance, and abnormal insulin secretion is invariably present in patients with type 2 diabetes. In this section, the physiology of normal insulin and the alterations that are present in persons with type 2 diabetes are reviewed.

Quantitation of Beta Cell Function

The measurement of peripheral insulin concentrations by radioimmunoassay is still the most widely used method for quantifying beta cell functions in vivo. Although this approach provides valuable information, it is limited by the fact that 50% to 60% of the insulin produced by the pancreas is extracted by the liver without ever reaching the systemic circulation. The standard radioimmunoassay for the measurement of insulin concentrations is also unable to distinguish between endogenous and exogenous insulin, making it ineffective as a measure of endogenous beta cell reserve in the insulin-treated diabetic patient. Anti-insulin antibodies that may be present in patients treated with insulin interfere with the insulin radioimmunoassay, making insulin measurements in insulin-treated patients inaccurate. Conventional insulin radioimmunoassays are also unable to distinguish between levels of circulating proinsulin and true levels of circulating insulin.

Insulin is derived from a single-chain precursor, proinsulin. Within the Golgi apparatus of the pancreatic beta cell, proinsulin is cleaved by convertases to form insulin, C peptide, and two pairs of basic amino acids. Insulin is subsequently released into the circulation at concentrations equimolar with those of C peptide. In addition, small amounts of intact proinsulin and proinsulin conversion intermediates are released. Proinsulin and its related conversion intermediates can be detected in the circulation, where they constitute 20% of the total circulating insulin-like immunoreactivity. In vivo, proinsulin has a biologic potency that is only about 10% of that of insulin and the potency of split proinsulin intermediates is between those of proinsulin and insulin. C peptide has no known conclusive effects on carbohydrate metabolism although certain physiologic effects of C peptide have been proposed. Unlike insulin, C peptide is not extracted by the liver and is excreted almost exclusively by the kidneys. Its plasma half-life of approximately 30 minutes contrasts sharply with that of insulin, which is approximately 4 minutes.

Because C peptide is secreted in equimolar concentrations with insulin and is not extracted by the liver, many investigators have used levels of C peptide as a marker of beta cell function. The use of plasma C-peptide levels as an index of beta cell function is dependent on the critical assumption that the mean clearance rates of C peptide are constant over the range of C-peptide levels observed under normal physiologic conditions. This assumption has been shown to be valid for both dogs and humans and this approach can be used to derive rates of insulin secretion from plasma concentrations of C peptide under steady-state conditions. However, because of the long plasma half-life of C peptide, under nonsteady-state conditions (e.g., after a glucose infusion) peripheral plasma levels of C peptide do not change in proportion to the changing insulin secretory rate. Thus, under these conditions, insulin secretion rates are best calculated with use of the two-compartment model initially proposed by Eaton and co-workers. Modifications to the C-peptide model of insulin secretion have been introduced. This approach combines the minimal model of insulin action with the two-compartment model of C-peptide kinetics and allows insulin secretion and insulin sensitivity to be derived after either intravenous or oral administration of glucose.
Signaling Pathways in the Beta Cell and Insulin Secretion

Figure 29-12 depicts the signaling pathways in the pancreatic beta cell that are involved in the stimulus-secretion coupling of insulin release. These pathways provide the mechanism whereby insulin secretion rates respond to changes in blood glucose concentrations. Glucose enters the pancreatic beta cell by a process of facilitated diffusion mediated by the glucose transporter GLUT2. Although levels of GLUT2 on the beta cell membrane are reduced in diabetic states for various reasons, it is not currently believed that this is a rate-limiting step in the regulation of insulin secretion.

The first rate-limiting step in this process is the phosphorylation of glucose to glucose-6-phosphate. This reaction is mediated by the enzyme glucokinase. There is considerable evidence that glucokinase, by determining the rate of glycolysis, functions as the glucose sensor of the beta cell and that this is the primary mechanism whereby the rate of insulin secretion adapts to changes in blood glucose. According to this view, as blood glucose levels increase more glucose enters the beta cell, the rate of glycolysis increases, and the rate of insulin secretion increases. A fall in blood glucose levels results in a fall in the rate of glycolysis and a reduction in the rate of insulin secretion.

Glucose metabolism produces an increase in cytosolic ATP, the key signal that initiates insulin secretion by causing blockade of the ATP-dependent K⁺ channel (KATP) on the beta cell membrane. Blockade of this channel induces membrane depolarization, which leads to an increase in cytosolic Ca²⁺ and insulin secretion. The biochemical events that link the increase in glycolysis to an increase in ATP are complex. Dukes and co-workers proposed the glycolytic production of NADH during the oxidation of glyceraldehyde-3-phosphate as the key process because NADH is subsequently processed into ATP by mitochondria through the operation of specific shuttle systems.

The rate of pyruvate generation has also been proposed as an explanation for the link between glucose metabolism and the increase in insulin secretion. According to this view, pyruvate generated by the glycolytic pathway enters the mitochondria and is metabolized further in the TCA cycle. Electron transfer from the TCA cycle to the respiratory chain by NADH and reduced flavin adenine dinucleotide (FADH₂) promotes the generation of ATP, which is exported into the cytosol. The increase in ATP closes ATP-sensitive K⁺ channels, which depolarizes the beta cell membrane and opens the voltage-dependent Ca²⁺ channels, leading to an increase in intracellular Ca²⁺. The increase in cytosolic Ca²⁺ is the main trigger for exocytosis, the process by which insulin-containing secretory granules fuse with the plasma membrane, leading to the release of insulin into the circulation. The increase in ATP not only closes KATP channels but also serves as a major permissive factor for movement of insulin granules and for priming of exocytosis.

Cyclic AMP also plays an important role in beta cell signal transduction pathways. This second messenger is generated at the plasma membrane from ATP and potentiates glucose-stimulated insulin secretion, particularly in response to glucagon, glucagon-like peptide 1 (GLP-1), and gastric inhibitory polypeptide. The cAMP-dependent pathways appear to be particularly important in the exocytic machinery.

KATP channels play an essential role in beta cell stimulus-secretion coupling. The reader is directed to an excellent review for more complete information. KATP channels comprise sulfonylurea receptors (SURs) and potassium inward rectifiers, Kir6.1 and Kir6.2, that assemble to form a large octameric channel with a (SUR/Kir6.x) stoichiometry. In the pancreatic beta cell the SUR1/Kir6.2 pairs constitute the KATP channel. KATP channels control the flux of potassium ions driven by an electrochemical potential. Opening these channels can set the resting membrane potential of beta cells below the threshold for activation of voltage-gated Ca²⁺ channels when plasma glucose levels are low, thus reducing insulin secretion. Changes in the cytosolic concentrations of ATP and adenosine diphosphate (ADP) as summarized earlier lead to closure of the channels and depolarization of the beta cell membrane. Mutations in both components of the beta cell KATP, SUR1, and Kir6.2 have been shown to lead to hypersecretion of insulin resulting clinically in either a recessive form of familial hyperinsulinemia or persistent hyperinsulinemic hypoglycemia of infancy.
Physiologic Factors Regulating Insulin Secretion

Carbohydrate Nutrients

The most important physiologic substance involved in the regulation of insulin release is glucose.\[439\] \[440\] \[441\] The effect of glucose on the beta cell is dose-related. Dose-dependent increases in concentrations of insulin and C peptide and in rates of insulin secretion have been observed after oral and intravenous glucose loads, with 1.4 units of insulin, on average, being secreted in response to an oral glucose load as small as 12 g.\[404\] \[405\] \[406\] \[407\] The insulin secretory response is greater after oral than after intravenous glucose administration.\[408\] \[409\] \[410\] \[411\] Known as the incretin effect,\[412\] \[413\] this enhanced response to oral glucose has been interpreted as an indication that absorption of glucose by way of the gastrointestinal tract stimulates the release of hormones and other mechanisms that ultimately enhance the sensitivity of the beta cell to glucose (see following discussion of hormonal factors). In a study involving nine normal volunteers in whom glucose was infused at a rate designed to achieve levels previously attained after an oral glucose load, the amount of insulin secreted in response to the intravenous load was 26% less than that secreted in response to the oral load.\[414\]

Insulin secretion does not respond as a linear function of glucose concentration. The relationship of glucose concentration to the rate of insulin release follows a sigmoidal curve, with a threshold corresponding to the glucose levels normally seen under fasting conditions and with the steep portion of the dose-response curve corresponding to the range of glucose levels normally achieved postprandially.\[415\] \[416\] \[417\] The sigmoidal nature of the dose-response curve has been attributed to a gaussian distribution of thresholds for stimulation among the individual beta cells.\[418\] \[419\] \[420\] \[421\]

When glucose is infused intravenously at a constant rate, an initial biphasic secretory response is observed that consists of a rapid, early insulin peak followed by a second, more slowly rising peak.\[422\] \[423\] \[424\] \[425\] The significance of the first-phase insulin release is unclear, but it may reflect the existence of a compartment of readily releasable insulin within the beta cell or a transient rise and fall of a metabolic signal for insulin secretion.\[426\] Despite early suggestions to the contrary,\[427\] \[428\] \[429\] \[430\] a subsequent study demonstrated that the first-phase response to intravenous glucose is highly reproducible within subjects.\[431\] \[432\] \[433\] \[434\] After the acute response, a second phase of insulin release occurs that is directly related to the level of glucose elevation. In vitro studies of isolated islet cells and the perfused pancreas have identified a third phase of insulin secretion commencing 1.5 to 3.0 hours after exposure to glucose and characterized by a spontaneous decline in secretion to 15% to 25% of the amount released during peak secretory level subsequently maintained for more than 48 hours.\[435\] \[436\] \[437\] \[438\]

In addition to its acute secretagogue effects on insulin secretion, glucose has intermediate and longer term effects that are physiologically and clinically relevant. In the intermediate term, exposure of the pancreatic beta cell to a high concentration of glucose primes its response to a subsequent glucose stimulus leading to a shift to the left in the dose-response curve relating glucose and insulin secretion.\[439\] \[440\] \[441\] \[442\] However, when pancreatic islets are exposed to high concentrations for prolonged periods, a reduction of insulin secretion is seen. Although all the precise mechanisms responsible for these adverse effects that have been termed glucotoxicity are not known, there is evidence that long-term exposure to high glucose reduces expression of a number of genes that are critical to normal beta cell function, including the insulin gene.\[443\] \[444\] \[445\] \[446\]

Noncarbohydrate Nutrients

Amino acids have been shown to stimulate insulin release in the absence of glucose, the most potent secretagogues being the essential amino acids leucine, arginine, and lysine.\[447\] \[448\] \[449\] The effects of arginine and lysine on the beta cell appear to be more potent than that of leucine. The effects of amino acids on insulin secretion are potentiated by glucose.\[450\] \[451\] \[452\] \[453\]

In contrast to amino acids, various lipids and their metabolites appear to have only minor effects on insulin release in vivo. Although carbohydrate-rich fat meals stimulate insulin secretion, carbohydrate-free fat meals have minimal effects on beta cell function.\[454\] Ketone bodies and short- and long-chain fatty acids have been shown to stimulate insulin secretion acutely both in isolated islet cells and in humans.\[455\] \[456\] \[457\] \[458\] The effects of elevated FFAs in the insulin secretory responses to glucose are related to the duration of the exposure. Zhou and Grill\[459\] first suggested that long-term exposure of pancreatic islets to FFAs inhibited glucose-induced insulin secretion and biosynthesis. This observation has been confirmed in rats.\[460\] In humans, it was demonstrated that the insulin resistance induced by an acute (90-minute) elevation in FFAs was compensated for by an appropriate increase in insulin secretion.\[461\] After chronic elevation of FFAs (48 hours), the beta cell compensatory response for insulin resistance was not adequate. Additional studies have demonstrated that the adverse effects of prolonged FFAs on glucose-induced insulin secretion are not seen in individuals with type 2 diabetes. From these results, it appears that elevated FFAs may contribute to the failure of beta cell compensation for insulin resistance.

Hormonal Factors

The release of insulin from the beta cell after a meal is facilitated by a number of gastrointestinal peptide hormones, including glucose-dependent insulinotropic peptide (GIP), cholecystokinin, and GLP-1.\[462\] \[463\] \[464\] \[465\] These hormones are released from small intestinal endocrine cells postprandially and travel in the blood stream to reach the beta cells, where they act through second messengers to increase the sensitivity of the islet cells to glucose. In general, these hormones are not of themselves secretagogues, and their effects are evident only in the presence of hyperglycemia.\[466\] \[467\] \[468\] \[469\] The release of these peptides may explain why the modest postprandial glucose levels achieved in normal subjects in vivo have such a dramatic effect on insulin production, whereas similar glucose concentrations in vitro elicit a much smaller response.\[470\] Similarly, this incretin effect could account for the greater beta cell response observed after oral as opposed to intravenous glucose administration.

Whether impaired postprandial secretion of incretin hormones plays a role in the inadequate insulin secretory response to oral glucose and to meals in IGT or diabetes is controversial.\[471\] \[472\] \[473\] \[474\] but pharmacologic doses of these peptides may have future therapeutic benefit. Subcutaneous administration of GIP-1, the most potent of the incretins, lowers glucose in type 2 diabetic patients by stimulating endogenous insulin secretion and perhaps by inhibiting glucagon secretion and gastric emptying.\[475\] \[476\] Because of its short half-life, however, its longer acting analogue, exendin-4, has greater therapeutic promise.\[477\]

Treatment with suryphysiologic doses of GIP during hyperglycemia has been shown to augment insulin secretion in normal\[478\] but not in diabetic humans.\[479\] \[480\] \[481\] \[482\] Although cholecystokinin has the ability to augment insulin secretion in humans, whether it is an incretin at physiologic levels is not firmly established.\[483\] \[484\] \[485\] \[486\] Its effects are also seen largely at pharmacologic doses.\[487\]

The postprandial insulin secretory response may also be influenced by other intestinal peptide hormones, including vasoactive intestinal polypeptide, somatostatin, gastrin, and gastrin,\[488\] \[489\] but the precise role of these hormones remains to be elucidated.

The hormones produced by pancreatic alpha and beta cells also modulate insulin release. Whereas glucagon has a stimulatory effect on the beta cell,\[490\] \[491\] somatostatin suppresses insulin release.\[492\] It is currently unclear whether these hormones reach the beta cell by traveling through the islet cell interstitium (thus exerting a paracrine effect) or through islet cell capillaries. Indeed, the importance of these two hormones in regulating basal and postprandial insulin levels under normal physiologic circumstances is in doubt. Paradoxically, the low insulin levels observed during prolonged periods of starvation have been attributed to the elevated glucagon concentrations seen in this setting.\[493\] \[494\] \[495\] \[496\] Other hormones that exert a stimulatory effect on insulin secretion include growth hormone,\[497\] glucocorticoids,\[498\] prolactin,\[499\] \[500\] placental lactogen,\[501\] \[502\] and the sex steroids.\[503\]

Whereas all of the preceding hormones may stimulate insulin secretion indirectly by inducing a state of insulin resistance, some may also act directly on the beta cell, possibly to augment its sensitivity to glucose. Thus, hyperinsulinemia is associated with conditions in which these hormones are present in excess, such as acromegaly, Cushing's syndrome, and the second half of pregnancy. Furthermore, treatments with placent lactogen,\[504\] hydrocortisone,\[505\] and growth hormone\[506\]
are all effective in reversing the reduction in insulin response to glucose that is observed in vitro after hypophysectomy. Although hyperinsulinemia after an oral glucose load has been observed in patients with hyperthyroidism, the increased concentration of immunoreactive insulin in this setting may reflect elevations in serum proinsulin rather than a true increase in serum insulin.

**Neural Factors**

The islets are innervated by both the cholinergic and adrenergic limbs of the autonomic nervous system. Although both sympathetic stimulation and parasympathetic stimulation enhance secretion of glucagon, the secretion of insulin is stimulated by vagal nerve fibers and inhibited by sympathetic nerve fibers. Adrenergic inhibition of the beta cell appears to be mediated by the $\beta$-adrenoceptor because its effect is attenuated by the $\beta$-antagonist phentolamine and reproduced by the $\alpha_2$-agonist clonidine. There is also considerable evidence that many indirect effects of sympathetic nerve stimulation play a role in regulation of beta cell function through stimulation or inhibition of somatostatin, $\gamma$-adrenoceptors, and neuropeptides galanin and neuropeptide Y.

Parasympathetic stimulation of islets results in stimulation of insulin, glucagon, and pancreatic polypeptide directly and through the neuropeptides vasoactive intestinal polypeptide, gastrin-releasing polypeptide, and pituitary adenylate cyclase-activating polypeptide. In addition, sensory innervation of islets may play a role in tonic inhibition of insulin secretion through the neuropeptides calcitonin generelated peptide and, less clearly, substance P.

The importance of the autonomic nervous system in regulating insulin secretion in vivo is unclear. The neural effects on beta cell function cannot be entirely dissociated from the hormonal effects because some of the neurotransmitters of the autonomic nervous system are, in fact, hormones. Furthermore, the secretion of insulinotropic hormones such as GIP and GLP-1 postprandially has been shown to be under vagal and adrenergic control.
Temporal Pattern of Insulin Secretion

It has been estimated that, in any 24-hour period, 50% of the total insulin secreted by the pancreas is secreted under basal conditions and the remainder is secreted in response to meals. The estimated basal insulin secretion rates range from 18 to 32 units per 24 hours (0.7 to 1.3 mg). After meal ingestion, the insulin secretory response is rapid and insulin secretion increases approximately fivefold over baseline to reach a peak within 60 minutes (Fig. 29-13; see Fig. 28-12). In these studies subjects consumed 20% of calories with breakfast and 40% with lunch and dinner, respectively. However, the amount of insulin secreted after each meal did not differ significantly. The rapidity of the insulin secretory response to breakfast is underscored by the fact that 71.6% ± 1.6% of the insulin secreted in the 4 hours after the meal was produced in the first 2 hours and the remainder in the next 2 hours. Insulin secretion did not decrease as rapidly after lunch and dinner and 62.8% ± 1.6% and 59.6% ± 1.4% of the total meal response were secreted in the first 2 hours after these meals.

The normal insulin secretory profile is characterized by a series of insulin secretory pulses. After breakfast, 1.8 ± 0.2 secretory pulses were identified in normal volunteers and the peaks of these pulses occurred 42.8 ± 3.4 minutes after the meal. Multiple insulin secretory pulses were also identified after lunch and dinner. After these meals, up to four pulses of insulin secretion were identified in both groups of subjects. Thus, in the 5-hour time interval between lunch and dinner, an average of 2.5 ± 0.3 secretory pulses were identified, and 2.6 ± 0.2 were identified in the same period after dinner.

Pulses of insulin secretion that did not appear to be meal-related were also identified. Between 11:00 AM and 6:00 AM and in the 3 hours before breakfast, on average 3.9 ± 0.3 secretory pulses were present in normal subjects. Thus, over the 24-hour period of observation, a total of 11.1 ± 0.5 pulses were identified in normal subjects. Close to 90% (87% ± 3%) of postmeal pulses in insulin secretion but only 47% ± 8% of nonmeal-related pulses were concomitant with a pulse in glucose.

Oscillatory Insulin Secretion

In vivo studies of beta cell secretory function have demonstrated that insulin is released in a pulsatile manner. This behavior is characterized by rapid oscillations occurring every 8 to 15 minutes that are superimposed on slower (ultradian) oscillations occurring at a periodicity of 80 to 150 minutes. The rapid oscillations persist in vitro and are therefore likely to be the result of metabolic pathways in the pancreatic beta cell that involved negative feedback loops with time lags.

The rapid oscillations of insulin are of small amplitude in the systemic circulation, averaging between 0.4 and 3.2 μU/mL in several published human studies. Because these values are close to the limits of sensitivity of most standard insulin radioimmunoassays, the characterization of these oscillations is subject to considerable pitfalls. Not the least of which is the need to differentiate between true oscillations of small amplitude and random assay noise. The latter problem has been overcome by the development of extremely sensitive enzyme-linked immunosorbent assays that allow the detection of extremely small changes in peripheral insulin concentrations. The application of these assays in studies involving frequent sampling from the peripheral circulation has led to a series of studies of the role of these oscillations in the overall regulation of insulin secretion.

These studies have suggested that increases in overall insulin secretion seen in response to a variety of secretagogues in various physiologic and pathophysiologic states are due to an increase in the amplitude of the bursts of insulin secretion. The studies have proposed that 75% of insulin secretion is accounted for by secretory bursts and the responses to GLP-1, sulfonylureas, and oral glucose are all mediated by an increase in the amplitude of insulin secretory pulses. Furthermore, consistent with observations made by O'Rahilly and colleagues a number of years ago, relatives of patients with type 2 diabetes demonstrate a disorderly profile of the insulin secretory oscillations. A number of mathematical programs have been developed that allow these insulin secretory oscillations to be evaluated and studied. The latest addition to the list is the development of ApEn and cross ApEn, which are statistics that measure temporal regularity of the oscillations in the insulin secretory profile.

The low amplitude of the rapid oscillations in the systemic circulation contrasts sharply with observations in the portal vein, where pulse amplitudes of 20 to 40 μU/mL have been recorded in dogs. Although the physiologic importance of these low-amplitude rapid pulses in the periphery is unclear, they are likely to be of physiologic importance in the portal vein. It is possible that the liver responds more readily to insulin delivered in a pulsatile fashion than to insulin delivered at a constant rate.
secretion without change in frequency, and there is a slight temporal advance of the glucose versus the insulin oscillation.

These findings suggest that the ultradian oscillations may be entirely accounted for by the major dynamic characteristics of the insulin-glucose feedback system, with no need to postulate the existence of an intrapancreatic pacemaker. In support of this hypothesis, Sturis and colleagues demonstrated that when glucose is administered in an oscillatory pattern, ultradian oscillations in plasma glucose and insulin secretion are generated that are 100% concordant with the oscillatory period of the exogenous glucose infusion. This close relationship between the ultradian oscillations in insulin secretion and similar oscillations in plasma glucose was further exemplified in a series of dose-response studies in which the largest amplitude oscillations in insulin secretion were observed in the subjects exhibiting the largest amplitude glucose oscillations, which in turn were directly related to the infusion dose of glucose. It has been shown that in normal humans, insulin is more effective in reducing plasma glucose levels when administered intravenously as a 120-minute oscillation than when delivered at a constant rate. These results indicate that the ultradian oscillations have functional significance.

Circadian variations in the secretion of insulin have also been reported. When insulin secretory responses were measured during a 24-hour period during which subjects received three standard meals, the maximal postprandial responses were observed after breakfast. These findings are mirrored by the results of studies in which subjects were tested for oral glucose tolerance at different times of the day and were found to exhibit maximal insulin secretory responses in the morning and lower responses in the afternoon and evening. These diurnal differences are also noted in tests for intravenous glucose tolerance. Furthermore, although ultradian glucose and insulin oscillations are closely correlated during a constant 24-hour glucose infusion, the nocturnal rise in mean glucose levels is not accompanied by a similar increase in the insulin secretory rate. It has been postulated that these diurnal differences may reflect diminished responsiveness of the beta cell to glucose in the afternoon and evening.
Insulin Secretion in Obesity and Insulin Resistance

Obesity and other insulin-resistant states are associated with a substantially greater risk for the development of type 2 diabetes. The ability of the pancreatic beta cell to compensate for insulin resistance determines whether blood glucose levels remain normal in insulin-resistant subjects or whether the subjects develop glucose intolerance or diabetes.

The nature of the beta cell compensation for insulin resistance involves hypersecretion of insulin even in the presence of normal glucose concentrations. This can occur only if beta cell sensitivity to glucose is increased. The increase in beta cell sensitivity to glucose seen in obesity appears to be mediated by two factors. First, increased beta cell mass is observed in obesity and other insulin-resistant states. Second, insulin resistance appears to be associated with increased expression of hexokinase in the beta cell relative to the expression of glucokinase. Because hexokinase has a significantly lower Michaelis constant (Km) for glucose than glucokinase, the functional effect of increased hexokinase expression is to shift the glucose-insulin secretion dose-response curve to the left, leading to increased insulin secretion across a wide range of glucose concentrations.

Assessment of the adequacy of the beta cell compensation for insulin resistance is important because this is the major determinant of the development of diabetes. In insulin-resistant states it is important to evaluate beta cell function in relation to the degree of insulin resistance. Kahn and co-workers studied the relationship between insulin sensitivity and beta cell function in 93 relatively young, apparently healthy human subjects of varying degrees of obesity. The sensitivity index SI was calculated using the minimal model of Bergman as a measure of insulin sensitivity and was then compared with various measures of insulin secretion. The relationship between the SI and the beta cell measures was curvilinear and reciprocal for fasting insulin (P < .0001), first-phase insulin response (AIR - acute insulin response) glucose; P < .0001), glucose potentiation slope (n = 56; P < .005), and beta cell secretory capacity (AIRmax; n = 43; P < .0001). The curvilinear relationship between SI and the beta cell measures could not be distinguished from a hyperbola, that is, SI × beta cell function = constant. The nature of this relationship is consistent with a regulated feedback loop control system such that for any difference in SI, a proportionate reciprocal difference occurs in insulin levels and responses in subjects with similar carbohydrate tolerance. Thus, in human subjects with normal glucose tolerance and varying degrees of obesity, beta cell function varies quantitatively with differences in insulin sensitivity. The increase in insulin secretion that is observed with a fall in SI should be viewed as the beta cell compensation that allows normal glucose tolerance to be maintained in the presence of insulin resistance.

The insulin resistance of obesity is characterized by hyperinsulinemia. Hyperinsulinemia in this setting reflects a combination of increased insulin production and decreased insulin clearance, but most evidence suggests that increased insulin secretion is the predominant factor. Both basal and 24-hour insulin secretory rates are three to four times higher in obese subjects and are strongly correlated with body mass index. Insulin secretory responses to intravenous glucose have been studied in otherwise healthy insulin-resistant subjects in comparison with insulin-sensitive subjects by means of a graded glucose infusion. Figure 29-15 depicts insulin concentrations and insulin secretion rates at each level of plasma glucose achieved, thereby constructing a glucose-insulin or glucose-insulin secretion rate dose-response relationship. Both insulin concentrations and insulin secretion rates are increased in insulin-resistant subjects, resulting from a combination of increased insulin production and decreased insulin clearance. For each level of glucose, insulin secretion rates are higher in the insulin-resistant subjects, reflecting an adaptive response of the beta cell to peripheral insulin resistance. Similar compensatory hyperinsulinemia has been demonstrated using other clinical techniques such as the frequency sampled intravenous glucose tolerance test in obesity and other insulin-resistant states such as late pregnancy.

The temporal pattern of insulin secretion is unaltered in obese subjects compared with normal subjects. Basal insulin secretion in obese subjects accounts for 50% of the total daily production of insulin, and secretory pulses of insulin occur every 1.5 to 2 hours. However, the amplitude of these pulses postprandially is greater in obese subjects. Nevertheless, when these postprandial secretory responses are expressed as a percentage of the basal secretory rate, the postprandial responses in obese and normal subjects are identical.
Insulin Secretion in Subjects with Impaired Glucose Tolerance

It has been suggested that insulin secretion may be normal in subjects with IGT. However, substantial defects in insulin secretion have been demonstrated in people with normal fasting glucose and glycated hemoglobin concentrations with glucose values greater than 140 mg/dL or 7.8 mmol/L 2 hours after ingestion of 75 g of glucose orally. Thus, defects in insulin secretion can be detected before the onset of overt hyperglycemia.

Detailed study of insulin secretion in patients with IGT has demonstrated that consistent quantitative and qualitative defects are seen in this group. During oral glucose tolerance testing, there is a delay in the peak insulin response. The glucose-insulin secretion dose-response relationship is flattened and shifted to the right (Fig. 29-16), and first-phase insulin responses to an intravenous glucose bolus are consistently decreased in relation to ambient insulin sensitivity. Further, abnormalities in first-phase insulin secretion were observed in first-degree relatives of patients with type 2 diabetes who exhibited only mild intolerance to glucose, and an attenuated insulin response to oral glucose was observed in normoglycemic co-twins of patients with type 2 diabetes. This pattern of insulin secretion during the so-called prediabetic phase was also seen in subjects with IGT who later developed type 2 diabetes and in normoglycemic obese subjects with a recent history of gestational diabetes. Another group at high risk for type 2 diabetes, Beta cell abnormalities may therefore precede the development of overt type 2 diabetes by many years.

The temporal pattern of insulin secretory responses is altered in IGT and is similar to but not as pronounced as that seen in diabetic subjects (see later). There is a loss of coordinated insulin secretory responses during oscillatory glucose infusion, indicating that the ability of the beta cell to sense and respond appropriately to parallel changes in the plasma glucose level is impaired. Abnormalities in rapid oscillations of insulin secretion have also been observed in first-degree relatives of patients with type 2 diabetes who have only mild glucose intolerance, further suggesting that abnormalities in the temporal pattern of beta cell function may be an early manifestation of beta cell dysfunction preceding the development of type 2 diabetes. Because an elevation in serum proinsulin is seen in subjects with diabetes, the contribution of proinsulin to the hyperinsulinemia of IGT has been questioned. Therefore, elevations in proinsulin, although elevations in fasting and stimulated proinsulin or proinsulin/insulin ratios have been found by many, although not all, investigators.

For by an increase in proinsulin, although elevations in fasting and stimulated proinsulin or proinsulin/insulin ratios have been found by many, although not all, investigators. Correlation of elevated proinsulin levels in IGT as a predictor of future conversion to diabetes has also been observed.
Insulin Secretion in Type 2 Diabetes Mellitus

Because of the presence of concomitant insulin resistance, patients with type 2 diabetes are often hyperinsulinemic, but the degree of hyperinsulinemia is inappropriately low for the prevailing glucose concentrations. Nevertheless, many of these patients have sufficient beta cell reserve to maintain a euglycemic state by dietary restriction with or without an oral agent. The beta cell defect in patients with type 2 diabetes mellitus is characterized by an absent first-phase insulin and C-peptide response to an intravenous glucose load and a reduced second-phase response. Although hyperglycemia may play a role in mediating these changes, the abnormal first-phase response to intravenous glucose persists in patients whose diabetic control has been greatly improved, consistent with the idea that patients with type 2 diabetes have an intrinsic defect in the beta cell. Furthermore, abnormalities in first-phase insulin secretion have also been observed in first-degree relatives of patients with type 2 diabetes who have only mild glucose intolerance, and an attenuated insulin response to oral glucose has been observed in normoglycemic co-twins of patients with type 2 diabetes group at high risk for type 2 diabetes and who can legitimately be classified as prediabetic. This pattern of insulin secretion during the so-called prediabetic phase is also seen in subjects with IGT who later develop type 2 diabetes and in normoglycemic obese subjects with a recent history of gestational diabetes, who are also at high risk for type 2 diabetes. Beta cell abnormalities may therefore precede the development of overt type 2 diabetes by many years.

Type 2 diabetes also affects proinsulin levels in serum. Increased levels of proinsulin are consistently seen in association with increases in the proinsulin/insulin molar ratio. The amount of proinsulin produced in this setting appears to be related to the degree of glycemic control rather than to the duration of the diabetic state, and in one series proinsulin levels contributed almost 50% of the total insulin immunoreactivity in type 2 diabetes patients who had marked hyperglycemia. In addition to intact proinsulin, the beta cell secretes one or more of the four major proinsulin conversion products (split 32,33-, split 65,66-, des-31,32-, and des-64,65-proinsulin) into the circulation. These conversion products are produced within the secretory granules of the islet as a result of the activity of specific conversion enzymes at the two cleavage sites in proinsulin linking the C peptide to the A and B chains.

The composition of the elevated proinsulin-like immunoreactivity (PLI) in patients with type 2 diabetes compared with control subjects has not been fully characterized. Hales and colleagues have developed immunoradiometric assays for this purpose. Using these assays, split 32,33-proinsulin was reported to be the predominant proinsulin conversion product in the circulation, although des-31,32-proinsulin levels may also be elevated. Insulin, proinsulin, and conversion product concentrations were also measured with these assays 30 minutes after oral glucose in patients with type 2 diabetes. Insulin was reduced in all patients, with no overlap between patients and controls, and concentrations of proinsulin and conversion products were elevated in the diabetic patients. These data highlight the importance of the potentially confounding effects of proinsulin and proinsulin conversion products in the interpretation of circulating immunoreactive insulin in patients with type 2 diabetes and emphasize the need to measure the concentrations of the individual peptides.

Abnormalities in the temporal pattern of insulin secretion have also been demonstrated in patients with type 2 diabetes. In contrast to normal subjects, in whom equal amounts of insulin are secreted basally and postprandially, in a given 24-hour period, patients with type 2 diabetes secrete a greater proportion of their daily insulin under basal conditions. This reduction in the proportion of insulin secreted postprandially appears to be related in part to a reduction in the amplitude of the secretory pulses of insulin occurring after meals rather than to a reduction in the number of pulses. In contrast to normal subjects, patients with type 2 diabetes have ultradian oscillations in insulin secretion that are less tightly coupled with oscillations in plasma glucose. Similar findings were observed in patients with IGT studied under the same experimental conditions and in a further group of type 2 diabetic patients studied under fasting conditions. The rapid insulin pulses are also abnormal in type 2 diabetes because the persistent regular rapid oscillations present in normal subjects are not observed. Instead, the cycles are of shorter duration and are irregular in nature. Similar findings were observed in a group of first-degree relatives of patients with type 2 diabetes who had only mild glucose intolerance, suggesting that abnormalities in oscillatory activity may be an early manifestation of beta cell dysfunction.

The effects of therapy on beta cell function in patients with type 2 diabetes have also been investigated. Although interpretation of the results in many instances is limited by the fact that beta cell function was not always studied at comparable levels of glucose before and during therapy, the majority of the studies indicated that improvements in diabetic control are associated with an enhancement of beta cell secretory activity. This increased endogenous production of insulin appears to be independent of the mode of treatment and is in particular associated with increases in the amount of insulin secreted postprandially. The enhanced beta cell secretory activity after meals reflects an increase in the amplitude of existing secretory pulses rather than an increased number of pulses. Despite improvements in glycemic control, beta cell function is not normalized after therapy, suggesting that the intrinsic defect in the beta cell persists.

Treatment with the sulfonylurea gliburide increases the amount of insulin secreted in response to meals but does not correct the underlying abnormalities in the pattern of insulin secretion. In particular, the abnormalities in the pulsatile pattern of ultradian insulin secretory oscillations persist on treatment with glyburide despite the increase in the amount of insulin secreted.

We have also investigated the effects on insulin secretion of improving insulin resistance in subjects with IGT by using the insulin-sensitizing agent troglitazone, a thiazolidinedione. Troglitazone therapy improved insulin sensitivity, and this was associated with enhanced ability of the pancreatic beta cell to respond to a glucose stimulus as judged by improvements in the dose-response relationships between glucose and insulin secretion as well as enhanced ability of the pancreatic beta cell to detect and respond to small oscillations in the plasma glucose concentration.
RODENT MODELS OF TYPE 2 DIABETES

A number of spontaneous and genetically selected animal models of type 2 diabetes have been identified. Most of the models combine the two main features of type 2 diabetes, obesity-associated insulin resistance and beta cell dysfunction with or without diminished beta cell mass. As with diabetes in humans, the different rodent models of type 2 diabetes have similarities but a number of overt and subtle differences make them useful surrogates for intensive study of the syndromes associated with type 2 diabetes.

An interesting observation is the striking sexual dimorphism in most rodent models of type 2 diabetes, with the male most often being affected exclusively, earlier, or more severely in most instances. In this regard, it is not like the human situation. The advent of transgenic and knockout technology in mice has produced a wide range of models of insulin resistance and beta cell dysfunction that results in hyperglycemia. It is beyond the scope of this chapter to review each of these, and the reader is referred to the primary literature for review of these animals. We limit our discussion to the well-documented spontaneous or derived models of the disease in rodents.

Mouse Models of Type 2 Diabetes

Leptin (Lep) and the Leptin Receptor (db)

The ob mutation, now designated Lep, was first described in 1960, but the gene mutation responsible for the syndrome was not described until the ob mutation was found to be in the gene for leptin. Mice homozygous for the ob mutation do not produce the satiety factor leptin and become markedly hyperphagic, obese, insulin-resistant, and hyperinsulinemic. They have a multitude of other pathohalamic functions that render them hypometabolic, contribute to the obesity, and also result in infertility. Leptin treatment of these mice results in decreased food intake and reverses many of their other metabolic defects. The ob mice develop obesity at weaning that becomes progressive because of hyperphagia. Insulin resistance is seen in muscle, adipose tissue, and liver, with a variety of signaling defects that are also reversible with insulin administration. The ob mouse becomes hyperglycemic and has a profound hyperinsulinemia associated with beta cell hyperplasia with up to a 10-fold increase in islet mass.

Parabiotic experiments between the ob and db mice suggested that the db mutation would be in the receptor for ob. This was confirmed with the identification of multiple mutations in the leptin receptor in db. Like ob mice, db mice are hyperphagic and begin to surpass their littermates in weight at weaning. They are progressively hyperinsulinemic, become hyperglycemic at 6 to 8 weeks, and because of a decline in beta cell function become markedly hyperinsulinemic at 4 to 6 months. The reason for the more severe diabetes in the db mouse is not clear, but it may be due to background strain differences as similar defects in insulin signaling are seen in this animal model as well. Treatment of both ob and db mice with insulin-sensitizing agents such as thiazolidinediones reversed the insulin resistance and ameliorated or prevented the onset of diabetes.

Agouti Mouse

Dominant "yellow" mutations in the agouti gene produce obesity and hyperglycemia. Depending on the background strain, the agouti mutation has a variable phenotype. In susceptible strains, the onset of hyperinsulinemia begins at 6 weeks of age and insulin levels continue to increase with age with beta cell hyperplasia and hypertrophy. The agouti mutation results in systemic production of a protein normally expressed in the skin, most frequently because of a retrotransposon insertion into the promoter region of the gene. Interestingly, a number of genes, including the fatty acid synthase gene, both have insulin and agouti response elements, which result in a marked increase in expression leading to increased hepatic fatty acid synthesis and enhanced fat deposition in adipocytes. The hyperglycemia is postprandial, and the fasting glucose levels are usually normal. The exact function of the agouti gene is unknown, but the animals are hyperphagic and show enhanced growth.

KK Mouse

These mice were originally bred for enhanced size but are not as obese as most other obese mice (usually less than 60 g). Breeding the KK into various background strains has produced variable insulin resistance, hyperinsulinemia, and hyperglycemia. The most studied strain is the KK, produced in Japan. This mouse has markedly increased insulin levels (>1000 µU/mL) when fed a high-fat diet. As the male mouse ages, glucose levels fall toward the normal range. The mutation responsible for the KK phenotype is unknown.

NZO Mouse

New Zealand obese (NZO) mice were derived by inbreeding of abdominally obese outbred mice. NZO neonates have high birth weights, and mice of both sexes are large and at weaning exhibit an elevated carcass fat content. Approximately 40% to 50% of group-caged NZO males, but not females, develop type II diabetes between 12 and 20 weeks of age when maintained with a chow diet containing 4.5% fat. Obesity in NZO mice is characterized by widespread accumulation of subcutaneous as well as visceral fat. The obesity in these mice is accompanied by glucose intolerance in males associated with increased hepatic and peripheral insulin resistance. In contrast to those in ob and db mice, genes encoding certain gluconeogenic and glycolytic enzymes in the liver retain normal responsiveness to insulin, although there is evidence for an inappropriately active fructose-1,6-biphosphatase. Defective beta cell insulin secretion from NZO islets in vitro and in vivo has been described. There appears to be a defect in the glycolytic pathway in beta cells leading to defective glucose-stimulated insulin release.

The genetics of NZO mice show a polygenic disorder, and none of the allelic variants have been discovered. Complicating the analysis of the model is the susceptibility of the mice to autoimmune disorders including a lupus-like syndrome and the insulin receptor. There is also a maternal influence in the periparturient period in the development of the disorder, which may reflect substances in the maternal milk.

Gold Thioglucose-Induced Diabetes

Gold thioglucose induces specific lesions in the ventromedial hypothalamus and induces an initial chronic hyperinsulinemia that leads to hypoglycemia, hyperphagia, obesity, and the development of insulin resistance and hyperglycemia. This model has been used as an example of pancreatic dysfunction preceding the induction of insulin resistance as opposed to pancreatic compensation for insulin resistance.

Diabetes Induced by Fat Ablation

Three models of insulin-resistant diabetes have been created in which adipose tissue has been genetically eliminated by overexpression of foreign genes using the fat-specific promoter aP2. Expression of an attenuated diphtheria toxin in adipose tissue resulted in an age-dependent loss of fat, progressive insulin resistance, and significant glucose intolerance, and diabetes. These mice represent a model of the human condition lipodystrophic diabetes and demonstrate the importance of fat in normal glucose homeostasis. It has been suggested that the lack of fat depots results in elevated
fatty acid delivery to liver and muscle and the development of insulin resistance. The diabetes in these animals can be variously treated by thiazolidinediones, \[587\] leptin administration, \[592\] and fat transplantation. \[591\] Interestingly, human lipodystrophy also responds to thiazolidinedione treatment, \[593\] suggesting that some of the effects of these compounds are not wholly dependent upon adipose tissue.

**C57BL/6J Mice Fed a High-Fat Diet**

Male C57BL/6J (also know as B6) mice fed a high-fat, high-carbohydrate diet (58% fat by kilocalories) or a "Western" diet developed hyperglycemia, hyperinsulinemia, hyperlipidemia, and increased adiposity. \[594\] Glucose-stimulated insulin secretion was blunted, and there was significant insulin resistance. \[595\] Despite obesity, plasma leptin levels in the Western diet-fed B6 mice were significantly lower than in control mice in the absence of hyperphagia. \[596\] The weight gain is due primarily to an increase in mesenteric adiposity, which makes this a good model for adult-onset type 2 diabetes.

**Nagoya-Shibata-Yasuda (NSY) Mice**

The NSY mouse shows male-specific, mild IGT with only a minority of the females becoming diabetic. \[597\] An impairment in beta cell function and obesity are present. These mice do not show the typical islet hyperplasia associated with insulin resistance.

**TallyHo Mice**

The TallyHo mouse also has a male-only development of diabetes associated with beta cell hyperplasia. Both male and

female TallyHo mice are obese, hyperinsulinemic, and hyperlipidemic, with the males having glucose levels greater than 500 mg/dL. \[600\]
**Rat Models of NIDDM**

**Zucker Diabetic Fatty (ZDF) Rat**

The orthologue of the db mouse, the obese Zucker rat (fa/fa), has a mutation in the leptin receptor that results in significant hyperphagia. This mutation is distinct from the mutations in ob in that it does not disrupt leptin receptor gene expression and does not affect ligand binding. This mutation results in a constitutive intracellular signaling domain, which may induce a desensitization of the leptin signaling pathways.

The selection of the inbred ZDF strain utilized Zucker (fa/fo) rats that had progressed to a diabetic phenotype. Brother-sister mating resulted in nearly 100% diabetes in the male rats receiving a 5% fat diet. Hyperglycemia begins to develop in males at 7 weeks of age, with serum glucose levels rising to 500 mg/dL by 12 weeks of age. The hyperinsulinemia precedes hyperglycemia with marked islet hyperplasia with dysmorphogenesis but by 19 weeks of age insulin levels drop concomitantly with islet atrophy, in part because of an imbalance of hyperplasia and apoptosis. The islets of prediabetic ZDF rats secrete significantly more insulin in response to glucose with elevated basal levels of insulin secretion and a leftward shift but a blunted glucose dose-response curve. Islets of prediabetic male ZDF rats also have defects in the normal oscillatory pattern of insulin secretion.

In contrast to the male ZDF rat, the female rat has significant insulin resistance but does not become diabetic unless given a proprietary high-fat diet (GMI 13004). The high-fat diet appears to have a direct effect on the beta cell as there is no change in peripheral insulin sensitivity (P. Hansen and C. F. Burant, unpublished). Interestingly, there is a decrease in peripheral triglyceride and FFA levels in the female rat after the institution of the high-fat diet.

The underlying genetic defect that results in beta cell failure in the ZDF rat is unknown. The beta cell number and insulin content are not different from those in homoyzogous normal animals, but insulin promoter activity is doubled in the ZDF rat. Insulin promoter mapping studies suggest that a critical region in the promoter of the insulin gene is affected. A number of other gene expression differences have been described in ZDF islets, including decreases in the expression of GLUT2, increases in glucokinase and hexokinase activity; decreases in mitochondrial metabolism; accumulation of intraslit lipid and long-chain fatty acyl CoA, which is associated with abnormal beta cell secretion; and increases in nitric oxide and ceramide accumulation, which is associated with apoptosis. Other gene expression changes are also found in the prediabetic rat islet. Which of these defects are important for the development of the diabetes is not clear. Despite the fixed genetic defect in the male animal that leads to diabetes, this defect interacts with the insulin resistance because treatment with insulin-sensitizing agents can prevent the onset of diabetes in the male and female. These agents are not effective in the male after establishment of diabetes; however, the female rat can respond to thiazolidinediones, even after significant hyperglycemia.

**Goto-Kakizaki (GK) Rats**

The GK inbred rat strain was derived from outbred Wistar rats by selection for IGT. Early in the development of diabetes, there are mild elevations of both glucose and insulin levels in the GK rat, but as the animals age, reduced beta cell mass is evident with markedly diminished insulin stores and abnormal secretory responses to glucose. A number of biochemical defects have been described in the islets of these animals, including decreased energy production, expression of proteins involved in insulin granule movement, and decreased adenylyl cyclase activity. Defects in peripheral signaling include decreased maximal and submaximal insulin-stimulated IRS1 tyrosine phosphorylation, IRS1-associated PI 3-kinase activity, and Akt activation in muscle and defective regulation of protein phosphatase-1 (PP-1), PP-2A, and mitogen-activated protein kinase activation by upstream insulin signaling components in adipocytes. Some of these defects may be due to hyperglycemia because they can be reversed by phlorizin-induced normalization of serum glucose.

**Bureau of Home Economics (BHE/Cdb) Rats**

The BHE/Cdb rat is a subline of the parent BHE obtained by selection for hyperglycemia and dyslipidemia without obesity. Glucose-stimulated insulin secretion is markedly diminished in these rats, a trait that is maternally inherited. A significant defect appears to be in the liver, where increased gluconeogenesis and lipogenesis precede the hyperglycemia, which may be due to defects in mitochondrial respiration associated with mitochondrial DNA mutations.

**Psammomys obesus (Sand Rat)**

This is a nutritionally induced obesity model of type 2 diabetes. Genetically, the Sand Rat is in reality a gerbil and the animal usually lives on a low-calorie vegetable diet. When given a high-carbohydrate diet, the Sand Rat rapidly becomes hyperglycemic secondary to weight gain associated with significant insulin resistance and enhanced hepatic glucose production. When a relatively hypocaloric diet is restored, the metabolic syndrome reverts to normoglycemia. A subpopulation of the Sand Rat develops frank beta cell failure and becomes ketogenic.

**Otsuka Long-Evans Tokushima Fatty (OLETF) Rats**

The OLETF rat strain was derived from the Long-Evans rat with polyuria, polydipsia, and mild obesity. About 90% of the male animals become diabetic by 1 year of age. Statistical tests have determined that the locus containing the cholecystokinin A receptor is responsible for about 50% of the NIDDM in the OLETF rats. The receptor is disrupted in the OLETF rat because of a 165-bp deletion in exon 1. Genetic segregation analysis has also shown interaction with a second locus, Odb2, which acts in a synergetic fashion to result in NIDDM, and both of these loci are required in homozygous OLETF rats to cause elevated plasma glucose.

The role of sex hormones is pronounced in this strain as orchietomy markedly reduces the incidence of diabetes whereas ovariectomy increases hyperglycemia to 30% in the female. Further, treatment of castrated males with testosterone restores the incidence to 89%. The islets undergo a progressive inflammatory reaction with progressive fibrosis. This reaction is associated with the impairment of beta cell function. Obesity and insulin resistance appear to precede the development of beta cell failure. Studies have also shown that obesity is necessary for the development of NIDDM in OLETF males and that insulin resistance may be closely related to fat deposition in the abdominal cavity. Troglitazone and metformin have been used successfully to treat the diabetes in the OLETF rat, with troglitazone completely preventing the beta cell morphologic and functional deterioration.

**Neonatal Streptozotocin**

Two models have been described in which a single dose of the beta cell toxic streptozotocin is given to 2-day-old female Wistar or male Sprague-Dawley rats. These animals have a transient hyperglycemia but develop IGT at 4 to 6 weeks of age. There is an initial reduction of beta cell mass with regeneration resulting in an approximately 50% reduction in adulthood.
MANAGEMENT OF TYPE 2 DIABETES

Over the last 10 years, a conceptual transformation in the principles of management of type 2 diabetes has occurred. Fundamentally, there has been a change in the level of concern about diabetes as a public health issue as well as in attitudes toward its treatment. Dramatic advances in the spectrum of pharmacologic agents and monitoring technology available for the treatment of diabetes have made it possible to lower glucose safely to the near-normal range in the majority of patients. Great strides have been made in establishing an evidence base for guidelines regarding glycemic control and efforts to reduce the risk of complications. Both corporate and government health insurance providers have greatly improved the extent to which diabetes equipment and supplies are covered.

A comprehensive review of all the subtleties of diabetes management in the 21st century is beyond the scope of this chapter. In the following pages, we deal with the salient features of the epidemiology of the complications of type 2 diabetes, diagnostic strategies, treatment guidelines, lifestyle interventions, and pharmacotherapy before turning briefly to a discussion of preventive measures for type 2 diabetes and its complications. An excellent source of information on these issues that is updated annually is the American Diabetes Association's Clinical Practice Recommendations. It is published as the first supplement to the journal Diabetes Care each January and is available online at www.diabetes.org by clicking "For Health Care Professionals"; near the end of that document is a listing of technical reviews, which are generally recent, fairly exhaustive treatments of most areas of interest in diabetes care.

Scope of the Problem

Type 2 diabetes is estimated to affect some 17 million to 20 million people in the United States. There is an epidemic of diabetes nationwide with a 6% annual growth rate in the prevalence of the disease. Worldwide, the prevalence of diabetes is increasing even faster. This increase is driven by population aging; population growth, particularly among ethnic groups with greater susceptibility to the disease; and dramatic increases in rates of obesity as a consequence of increasingly sedentary lifestyles and greater consumption of simple sugars and high-caloric-density foods. At least in the United States, opportunistic screening for diabetes in high-risk populations is recommended by professional societies and many insurers, resulting in an increase in the proportion of affected individuals diagnosed in this country from approximately one half a decade ago to about two thirds.

The morbidity, mortality, and expense associated with diabetes are staggering. In Western society, people with diabetes are three times more likely to be hospitalized than nondiabetic individuals. In the United States, diabetes is the leading cause of blindness and accounts for over 40% of the new cases of end-stage renal disease. The risk of heart disease and stroke is 2 to 4 times higher and the risk of lower extremity amputation is approximately 20 times higher for people with diabetes than for those without. Life expectancy is reduced by approximately 10 years in people with diabetes, and although diabetes is the seventh leading cause of death in the United States, this is clearly an underestimate. Despite the fact that some 70% of people with diabetes die of heart disease and stroke, only approximately 10% have diabetes listed as a contributing cause on death certificates.

Tragically, this enormous burden of death and disability has not been reduced by huge health care expenditures. In fact, the epidemic of diabetes is one of the drivers of increasing health care costs, with annual disbursements for people with diabetes approximately three to five times higher per capita than those for individuals without diabetes. In the United States, at least 15% of health care expenditures are related to the treatment of people with diabetes. Nevertheless, whereas rates of coronary artery disease are declining in the United States in general, this is not the case for people with diabetes. However, there is evidence that increased effort to control diabetes and its comorbidities can even reduce costs associated with diabetes and that a public health approach to diabetes can reduce the burden of complications of diabetes.
Screening and Diagnosis

The role of screening to make the diagnosis of diabetes in asymptomatic individuals is an area of substantial controversy. No prospective randomized trials have examined the benefit of such a screening program. On the other hand, it seems self-evident that early diagnosis and intervention have at least the potential to reduce complications in a disease in which 20% to 50% of patients have a complication at the time of diagnosis. The cost-effectiveness of universal approaches to diabetes screening has been called into question. The American Diabetes Association (ADA) recommendations for screening are based on a review that concludes, “Periodic, targeted, and opportunistic screening within the existing health care system seems to offer the greatest yield and likelihood of appropriate follow-up and treatment.” The ADA suggests that patients (i.e., screening should be performed only in the context of a routine health care setting) should be screened at 3-year intervals beginning at age 45 and that testing should be considered at an earlier age or be carried out more frequently if diabetes risk factors are present. Those risk factors are listed in Table 29-3.

Most groups recommend FPG as the most practical screening test, although it is recognized that its sensitivity is substantially lower than that of the OGTT. More recent data suggest that alternative screening strategies may have advantages. In a study employing glucose meters to measure random capillary blood glucose, values of 120 mg/dL or higher obtained at random without regard to meals were 75% to 84% sensitive and 86% to 90% specific for detecting diabetes as defined by either FPG or oral glucose tolerance testing. In the future, it is possible that well-validated models will allow us to predict diabetes risk from standard biologic measures such as body mass index, blood pressure, and lipids with greater precision than today.

Classically, diabetes has been diagnosed on the basis of prospective epidemiologic data associating circulating glucose levels with the future development of diabetic retinopathy. In 1997, recommendations were made by an expert committee to change the diagnostic criteria for diabetes to improve the sensitivity of FPG for the diagnosis of diabetes. They determined, from a review of several large data sets, that FPG greater than or equal to 126 mg/dL (7.0 mM) identified a population of people with a risk of retinopathy similar to that of those with a 2-hour value in an OGTT of 200 mg/dL or higher. In an effort to simplify the OGTT, only the 2-hour plasma glucose after a 75-g oral glucose load needs to be measured for diagnostic purposes. Furthermore, patients with classical symptoms of diabetes in association with a random glucose level of 200 mg/dL or higher also meet diagnostic criteria for diabetes. To avoid misclassification, it is further suggested that patients should meet one of the three diagnostic criteria on at least two separate days before making the diagnosis of diabetes.

Because macrovascular disease accounts for the majority of the morbidity and almost all the mortality associated with diabetes and the diagnosis of diabetes is associated with more stringent guidelines for the treatment of comorbidities such as dyslipidemia and hypertension, it seems likely that the diagnostic criteria for diabetes will be lowered again to make the fasting glucose cut points more sensitive. This seems appropriate, as it is clear that glucose levels above normal but below the current thresholds for diabetes are associated with increased cardiovascular risk. A debate unlikely to be answered in the next decade is whether it is acceptable to measure only fasting glucose as an index of glucose intolerance or whether it is cost-effective to use an oral challenge to ascertain fully glucose-related risks.
Glucose Treatment Guidelines

Prospective randomized clinical trials have documented improved rates of microvascular complications in patients with type 2 diabetes treated to lower glycemic targets. In the UK Prospective Diabetes Study (UKPDS), patients with newonset diabetes were treated with diet and exercise for 3 months with an average reduction in glycosylated hemoglobin or HbA\textsubscript{1c} from approximately 9% to 7% (upper limit of normal 6%). Those with FPG greater than 108 mg/dL (6 mM) were randomly assigned to two treatment policies. In the standard intervention, subjects continued the lifestyle intervention. Pharmacologic therapy was initiated only if the FPG reached 15 mM (270 mg/dL) or the patient became symptomatic. In the more intensive treatment program, all patients were randomly assigned and treated with either sulfonylurea, metformin, or insulin as initial therapy, with the dose increased to try to achieve an FPG less than 108 mg/dL. Combinations of agents were used only if the patients became symptomatic or FPG became greater than 270 mg/dL (15 mM). As a consequence of the design, although the HbA\textsubscript{1c} fell initially to about 6%, over the average 10 years of follow-up it rose to approximately 8%. The average HbA\textsubscript{1c} in the standard treatment group was approximately 1% higher. The risk of severe hypoglycemia was smaller the order of 1% to 5% per year in the insulin-treated patient group and weight gain was modest; both were higher in patients randomly assigned to insulin and lower in those receiving metformin. \textsuperscript{43} Associated with this improvement in glycemic control, there was a reduction in the risk of microvascular complications (retinopathy, nephropathy, and neuropathy) in the intensive group. Although there was a trend toward reduced rates of macrovascular events in the more intensively treated group, it did not reach statistical significance. \textsuperscript{43}

Similar reductions in microvascular events were observed in another trial of entirely different design and much smaller size. In the Kumamoto study, Japanese patients of normal weight with type 2 diabetes treated with insulin were randomly assigned to standard treatment or an intensive program of insulin therapy designed to achieve normal glycemia. The control group maintained HbA\textsubscript{1c} values at approximately 9%, whereas the HbA\textsubscript{1c} in the intensive group was reduced to approximately 7% and the separation maintained for 6 years. Again, there was a modest increased risk of hypoglycemia and weight gain, a reduction in microvascular complications, and a non-statistically significant trend toward reduced rates of vascular end points. \textsuperscript{45}

Although no interventional studies have documented a reduced risk of vascular end points associated with an improvement in glycemic control, multiple epidemiologic studies have suggested that there is an association between cardiovascular risk and HbA\textsubscript{1c}, FPG, and the 2-hour level in the OGTT. \textsuperscript{671} \textsuperscript{672} \textsuperscript{673} \textsuperscript{674} \textsuperscript{675} \textsuperscript{676} In the UKPDS epidemiologic analysis, there was a 16% reduction in cardiovascular disease rates per 1% reduction in HbA\textsubscript{1c} without evidence of a threshold or lower limit of benefit all the way into the normal range. \textsuperscript{672}

In Table 29-6, guidelines from the ADA and the American College of Endocrinology (ACE) are presented. The ADA suggests that the goal of treatment in the management of diabetes should be an HbA\textsubscript{1c} value less than 7%. \textsuperscript{672} Although initially developed on an ad hoc basis, this goal is supported by the clinical trial data as this level of glycemia was associated with improved outcomes in patients with type 2 diabetes in the preceding studies. The ACE, on the basis of the fact that in the same outcome studies normal glycemia levels were targeted and achieved in at least some subjects and that epidemiologic analyses suggest no threshold to the benefit of glucose lowering, has recommended an HbA\textsubscript{1c} goal of less than 6.5%. \textsuperscript{675} \textsuperscript{676} Because the average HbA\textsubscript{1c} in the United States is estimated to be in the 7.5% to 9.5% range, the argument about whether the HbA\textsubscript{1c} target should be 6.5% or 7% is of limited practical significance.

However, it should be recognized that there are potential adverse events related to pursuit of more aggressive targetsglycemia, long-term exposure to poorly studied combinations of medications, expense, life disruption caused by greater attention and effort to achieve lower glycemic targets, and the potential that great efforts expended in achieving extremely stringent glycemic goals will result in less attention to other health risks by patient or provider. No cohort of patients of substantial size has ever been reported in which an average HbA\textsubscript{1c} level less than 7% has been achieved over a time frame that exceeds more than a few months. Several adequately powered, randomized controlled clinical trials are under way or being planned to explore the effects of seeking more intensive glycemic targets (HbA\textsubscript{1c} < 6%). Although it is clear that many patients can achieve lower glucose levels with currently available drugs and lifestyle interventions, it remains theoretically possible that the risks would exceed the benefits of seeking glucose targets less than 7%.

With respect to fasting, premeal, or postprandial targets, there is little support for any particular level of glycemic control in the management of type 2 diabetes as no large-scale outcome study has targeted particular levels of glucose with home glucose monitoring. The ADA target of fasting and premeal plasma glucose levels of 90 to 130 mg/dL, is based on an estimate of the range of average glucose values that would be associated with a low risk of hypoglycemia and HbA\textsubscript{1c} less than 7%. \textsuperscript{673} The ACE target of less than 110 mg/dL is an effort to achieve normal levels of glycemia. \textsuperscript{672} However, it should be recognized that consistent fasting and premeal glucose levels less than 110 mg/dL would be expected to be associated with an HbA\textsubscript{1c} of approximately 5.5%. \textsuperscript{673}

The ADA has not set any treatment goals for postprandial glucose levels because there are no published studies in which even safety, much less outcome, is documented for targeting a particular level of postprandial glucose. \textsuperscript{673} However, the ADA statement on postprandial glucose recognizes that there are effective HbA\textsubscript{1c} -lowering agents that primarily target postprandial glucose levels and suggests that monitoring postprandial glucose levels may allow dose adjustment of these agents. \textsuperscript{673} Furthermore, they recognize that there are patients with diabetes who have average fasting glucose levels within targets but whose HbA\textsubscript{1c} is elevated and that monitoring and specifically treating postprandial elevations in these patients may provide improvements in HbA\textsubscript{1c}, perhaps with a lower risk of hypoglycemia and weight gain than further lowering fasting and premeal glucose levels. The ACE recommends targeting a 2-hour postprandial glucose less than 140 mg/dL (7.8 mM) in an effort to achieve near-normal glycemia. \textsuperscript{672} Consistent postprandial glucose values less than 140 mg/dL would be associated with average HbA\textsubscript{1c} levels of approximately 5.5%. \textsuperscript{673}

\medskip

\begin{table}[h]
\centering
\caption{Glycemic Targets}
\begin{tabular}{|l|c|c|c|}
\hline
Parameter & Normal & ADA & ACE \\
\hline
Premeal plasma glucose (mg/dL) & <110 (mean 90) & 90/130 & <110 \\
Postprandial plasma glucose (mg/dL) & <140 & <140 & <140 \\
HbA\textsubscript{1c} & 4%-6% & <7% & <6.5% \\
\hline
\end{tabular}
\end{table}


Larsen: Williams Textbook of Endocrinology, 10th ed., Copyright © 2003 Elsevier
Lifestyle Intervention

The components of lifestyle intervention include medical nutrition counseling, exercise recommendations, and comprehensive diabetes education with the purpose of changing the paradigm of care in diabetes from provider-focused to patient-focused. Arguably, over the last 5 years, nothing has changed more fundamentally than the emphasis on lifestyle intervention. For decades, physicians and patients have paid lip service to the notion that lifestyle intervention is important. Now we have significant clinical trial evidence that each component of lifestyle intervention, when appropriately administered, can contribute to improved outcomes. Furthermore, since the Balanced Budget Act of 1997 and the passage of complementary legislation in most state governments, lifestyle intervention has been a covered benefit for most people. Although full implementation of these regulations is still in progress, they have dramatically expanded the proportion of the population with diabetes with insurance coverage for these essential services.

Education of Patients

Diabetes is a lifelong disease, and health care providers have almost no control over the extent to which patients adhere to the day-to-day treatment regimen. The appropriate role of the health care provider is to serve as a coach to the patient, who has primary responsibility for the delivery of daily care.

As a result, health care providers must carefully engage patients as partners in the therapeutic process. It is critical for the health care professional to understand the context in which patients are taking care of their disease. Using a prescriptive approach in which patients are told what to do can work in some situations but fails more often than not because of unrecognized barriers to the execution of a particular plan. For long-term success, diabetes self-management education is critical.

As defined by the ADA,259 diabetes self-management education is the process of providing the person with diabetes with the knowledge and skills needed to perform self-care, manage crises, and make lifestyle changes. As a result of this process, the patient must become a knowledgeable and active participant in the management of his or her disease. To achieve this,

<table>
<thead>
<tr>
<th>TASK</th>
<th>ROLE OF THE PLAYERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Be a source of accurate information and to coordinate with other sources of information as necessary</td>
<td>Primary care provider</td>
</tr>
<tr>
<td>To provide guidance in developing goals of treatment</td>
<td>Other providers</td>
</tr>
<tr>
<td>To provide guidance in developing goals of treatment and to help the primary care provider develop strategies to achieve treatment goals and avoid complications</td>
<td>Patient</td>
</tr>
</tbody>
</table>


A team of providers is generally required to implement fully the process of diabetes self-management education as the amount of information that needs to be exchanged is large and the range of expertise broad. It is generally not possible to cover the recommended content fully in the context of several or even many brief encounters with a physician in an office setting. Potential providers in a team care approach could include nurses, dietitians, exercise specialists, behavioral therapists, pharmacists, and other medical specialists including diabetologists or endocrinologists, podiatrists, medical subspecialists, obstetrician-gynecologists, psychiatrists, and surgeons. The potential role of the community in which the patient lives and works in the diabetes self-care process is enormous, including family, friends, employers, and health care insurance providers. Each potential member of the team has a role to play in the process, which must be reviewed and assessed frequently (Table 29-8). The primary role of the provider in this process is to provide guidance in goal setting to manage the risk of complications, suggest strategies to achieve goals and techniques to overcome barriers, provide training in skills, and help screen for complications. For this process to be a success, the patient must commit to the principles of self-care, participate fully in the development of a treatment plan, make ongoing decisions regarding self-care from day to day, and communicate honestly and with sufficient frequency with the team.

Fortunately, barriers to providing team care are becoming less daunting. Diabetes education programs are being established at a rapid rate. The American Association of Diabetes Educators (800-TEAM-UP4) and the ADA (800-DIABETEs) can provide information regarding diabetes educators and education programs in your area.

For team care to be most effective, communication, trust, and mutual respect are critical. Unfortunately, in many communities, the full benefit of the consultation and ongoing care with diabetes educators, nurses, dietitians, pharmacists, or others is not achieved because of overly hierarchical approaches to care. Nonphysicians, including patients, ought to provide suggestions regarding medication and lifestyle adjustments and

TABLE 29-7 – Curricular Areas That Should Be Addressed in Diabetes Self-Management Education

| PATHOPHYSIOLOGY OF THE PATIENT’S DIABETES AND ITS RELATIONSHIP TO TREATMENT OPTIONS |
|------|---------------------|
| INCORPORATING APPROPRIATE NUTRITIONAL MANAGEMENT |
| UTILIZING PHYSICAL ACTIVITY INTO LIFESTYLE |
| MONITORING BLOOD GLUCOSE, URINE KETONES (WHEN APPLICABLE), AND USING THE RESULTS TO IMPROVE CONTROL |
| PREVENTING, DETECTING, AND TREATING ACUTE COMPLICATIONS INCLUDING “SICK DAY RULES” AND HYPOGLYCEMIA |
| PREVENTING, DETECTING, AND TREATING CHRONIC COMPLICATIONS |
| GOAL SETTING TO PROMOTE HEALTH, AND PROBLEM SOLVING FOR DAILY LIVING |
| INTEGRATING PSYCHOSOCIAL ADJUSTMENT TO DAILY LIFE |
| PROMOTING PRECONCEPTION CARE, MANAGEMENT DURING PREGNANCY, AND GESTATIONAL DIABETES MANAGEMENT (IF APPLICABLE) |

A significant clinical trial evidence that each component of lifestyle intervention, when appropriately administered, can contribute to improved outcomes.

Furthermore, since the Balanced Budget Act of 1997 and the passage of complementary legislation in most state governments, lifestyle intervention has been a covered benefit for most people. Although full implementation of these regulations is still in progress, they have dramatically expanded the proportion of the population with diabetes with insurance coverage for these essential services.

To commit to diabetes self-management (as defined above)
help in the process of identifying barriers to effective management such as lack of knowledge, lack of time, and lack of resources and strategies to overcome those barriers.

Perhaps some of the most overlooked contributors to ineffective care in the setting of type 2 diabetes are the relatively common barriers created by psychiatric, neurocognitive function, and adjustment disorders, which are largely responsive to psychosocial therapies.

With respect to self-management education, a technical review documents the effect of medical nutrition therapy and specific advice on diabetes-related outcomes such as HbA1c, weight, and proteinuria. These are summarized in Table 29-9. A comprehensive, individually negotiated nutrition program in which each patient's circumstances, preferences, and cultural background as well as the overall treatment program are considered is most likely to result in optimal outcomes. Because of the complexity of both the medical and nutritional issues for most patients, it is recommended that a registered dietitian, with specific skill and experience in implementing nutrition therapy in diabetes management, work collaboratively with the patient and other health care team members in providing medical nutrition therapy.

Analogously, physicians and other members of the health care team need to understand the major issues in diabetes and nutrition and support the nutritional plan developed collaboratively. Individualized dietary advice can be developed by a physician from a brief diet history obtained by asking: "What do you eat for breakfast? . . . lunch? . . . supper? Do you have snacks between breakfast and lunch? . . . lunch and supper? . . . supper and bedtime? What do you drink during the day?" Ideally, this information should be obtained at each visit, with specific suggestions for change that both patient and provider agree are important and achievable.

Easy issues to address include caloric beverages, which tend to elevate glucose levels dramatically and can generally be replaced quite painlessly by artificially sweetened alternatives. Juices generally are perceived as healthy but can significantly affect glycemic control and total calorie intake. Substituting low-fat products for higher fat alternatives is useful but needs to be done with the recognition that they are generally higher in carbohydrates. "Fat-free" and "sugar-free" foods need to be recognized as food that is not "free." Portion control and recipe modification are excellent dietary techniques, particularly for meals and fried foods.

Adequate spacing between meals is usually good advice for patients with type 2 diabetes because postprandial glucose levels generally peak 2 hours after a meal, when a snack would normally be taken. Eating approximately every 4 hours while awake is the most practical dietary plan for most overweight people. Frequent small meals have been shown to be of benefit when used in a controlled inpatient setting, but in general when overweight patients are encouraged to eat more frequently they often overeat more frequently. At a minimum, avoiding high-calorie snacks is reasonable advice for most people with diabetes. A repeated diet history and additional modest changes negotiated every few weeks to months by all health care providers (doctor, nurse, or dietitian) allow assessment of whether previously agreed to changes were enacted, reinforcement of the importance of dietary efforts, and slow eneishment of patients into more healthful dietary choices.

In general, the critical nutrient for glycemic consistency is carbohydrate. Essentially every molecule of carbohydrate consumed is converted to glucose in the gut and requires the action of insulin to be cleared from the circulation. A dietary term called carbohydrate counting can be used in patients with type 2 diabetes to facilitate consistent carbohydrate intake or to allow insulin dose adjustment in response to changes in carbohydrates consumed. Whereas the beta cell in type 2 diabetes has generally lost its responsiveness to glucose, the second phase of insulin secretion is largely spared in type 2 diabetes and is in part driven by amino acids and fatty acids. Therefore, including some protein and fat in each meal and snack is useful.

Dietary fat is the nutrient most closely associated in epidemiologic studies with the risk of developing type 2 diabetes. Although dietary fats clearly have an impact on glucose and triglycerides but are much less calorically dense and have a higher thermic effect, both of which tend to promote weight loss.

Dietary protein similarly has a minimal impact on glucose levels, although amino acids do promote insulin secretion, which may be advantageous in the setting of type 2 diabetes. Metabolism of protein results in the formation of acids and nitrogenous waste, which may result in bone demineralization and glomerular hyperfiltration. At least 0.8 g of high-quality protein per kilogram is generally recommended; otherwise, restriction of protein intake to 10% to 20% of total calories minimizes potential adverse long-term effects of high protein intake.

The role of vitamins, trace minerals, and nutritional supplements in the treatment of diabetes is poorly understood. There are some who are absolutely convinced of adverse long-term effects of high protein intake.

Table 29-9 -- Major Nutrition Recommendations for Diabetes

<table>
<thead>
<tr>
<th>TABLE 29-9 – Major Nutrition Recommendations for Diabetes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Carbohydrate</strong></td>
</tr>
<tr>
<td>Foods containing carbohydrate from whole grains, fruits, vegetables, and low-fat milk are important components, and should be included, in a healthy diet.</td>
</tr>
<tr>
<td>With regard to the glycemic effects of carbohydrates, the total amount of carbohydrate in meals or snacks is more important than the source or type.</td>
</tr>
<tr>
<td>Individuals receiving intensive insulin therapy should adjust their premeal insulin doses based on the carbohydrate content of meals.</td>
</tr>
<tr>
<td>As sucrose does not increase glycemia to a greater extent than isocaloric amounts of starch, sucrose and sucrose-containing foods do not need to be restricted by people with diabetes; however, they should be substituted for other carbohydrate sources or, if added, covered with insulin or other glucose-lowering medication.</td>
</tr>
<tr>
<td>Nonnutritive sweeteners are safe when consumed within the acceptable daily intake levels established by the Food and Drug Administration.</td>
</tr>
<tr>
<td>Individuals receiving fixed daily insulin doses should try to be consistent in day-to-day carbohydrate intake.</td>
</tr>
<tr>
<td>Although the use of low-glycemic-index foods may reduce postprandial hyperglycemia, there is not sufficient evidence of long-term benefit to recommend use of low-glycemic-index diets as a primary strategy in food or meal planning.</td>
</tr>
<tr>
<td>As with the general public, consumption of dietary fiber is to be encouraged; however, there is no reason to recommend that people with diabetes consume a greater amount of fiber than other Americans.</td>
</tr>
<tr>
<td>Carbohydrate and monounsaturated fat should together provide 60%-70% of energy intake. However, the metabolic profile and need for weight loss should be considered when determining the monounsaturated fat content of the diet.</td>
</tr>
<tr>
<td>Sucrose and sucrose-containing foods should be eaten in the context of a healthy diet.</td>
</tr>
<tr>
<td><strong>Protein</strong></td>
</tr>
<tr>
<td>In persons with controlled type 2 diabetes, ingested protein does not increase plasma glucose concentrations, although protein is just as potent a stimulant of insulin secretion as carbohydrate.</td>
</tr>
</tbody>
</table>
For persons with diabetes, especially those not in optimal glucose control, the protein requirement may be greater than the recommended dietary allowance (RDA) but not greater than usual intake.

The long-term effects of diets high in protein and low in carbohydrate are unknown. Although such diets may produce short-term weight loss and improved glycemia, it has not been established that weight loss is maintained long-term. The long-term effect of such diets on low-density lipoprotein (LDL) cholesterol is also a concern.

Fat

Less than 10% of energy intake should be derived from saturated fats. Some individuals (i.e., persons with LDL cholesterol 100 mg/dL) may benefit from lowering saturated fat intake to less than 7% of energy intake.

To lower LDL cholesterol, energy derived from saturated fat can be reduced if weight loss is desirable or replaced with either carbohydrate or monounsaturated fat if weight loss is not a goal.

Dietary cholesterol intake should be less than 300 mg/day. Some individuals (i.e., persons with LDL cholesterol 100 mg/dL) may benefit from lowering dietary cholesterol to less than 200 mg per day.

Intake of trans-unsaturated fatty acids should be minimized.

Polyunsaturated fat intake should be approximately 10% of energy intake.

Reduced-fat diets, when maintained long-term, contribute to modest loss of weight and improvement in dyslipidemia.

Energy balance and obesity

In insulin-resistant individuals, reduced energy intake and modest weight loss improve insulin resistance and glycemia in the short term.

Structured programs which emphasize lifestyle changes including education, reduced fat (<30% of daily energy) and energy intake, regular physical activity, and regular participant contact can produce long-term weight loss on the order of 5% to 7% of starting weight.

Exercise and behavior modification are most useful as adjuncts to other weight loss strategies. Exercise is helpful in maintenance of weight loss.

Standard weight reduction diets, when used alone, are unlikely to produce long-term weight loss. Structured, intensive lifestyle programs are necessary.

Micronutrients

There is no clear evidence of benefit from vitamin or mineral supplementation in people with diabetes who do not have underlying deficiencies.

Routine supplementation of the diet with antioxidants is not advised because of uncertainties related to long-term efficacy and safety.

Alcohol

If individuals choose to drink alcohol, daily intake should be limited to one drink for adult women and two drinks for adult men. One drink is defined as 12 oz of beer, 5 oz of wine, or 1.5 oz of 80 proof spirits.

To reduce risk of hypoglycemia, alcohol should be consumed with food.

Children and adolescents with diabetes

Individualized food or meal plans and intensive insulin regimens can provide flexibility for children and adolescents with diabetes to accommodate irregular mealtimes and schedules, varying appetite, and varying activity levels.

Nutrient requirements for children and adolescents with type 1 or type 2 diabetes appear to be similar to those of other same-age children and adolescents.

Older adults

Energy requirements for older adults are less than for younger adults.

Physical activity should be encouraged.

In the elderly, undernutrition is more likely than overnutrition and therefore caution should be exercised when prescribing weight loss diets.

Acute complications

Glucose is the preferred treatment for hypoglycemia, although any form of carbohydrate that contains glucose may be used.

Ingestion of 15 to 20 g of glucose is an effective treatment for hypoglycemia but blood glucose may be corrected only temporarily.

Blood glucose should be evaluated in approximately 60 minutes, as additional treatment may be necessary.

During acute illnesses, testing blood glucose and blood or urine for ketones, drinking adequate amounts of fluids, and ingesting carbohydrates are important.

Hypertension

In both normotensive and hypertensive individuals, a reduction in sodium intake lowers blood pressure. The goal should be to reduce sodium intake to 2400 mg (100 mmol) or 6000 mg sodium chloride (salt) per day.

A modest amount of weight loss beneficially affects blood pressure.

Dyslipidemia

For persons with elevated LDL cholesterol, saturated fatty acids and trans-saturated fatty acids should be limited to less than 10% and perhaps to less than 7% of energy.

Energy derived from saturated fat can be reduced if weight loss is desirable or replaced with either carbohydrates or monounsaturated fats if weight loss is not a goal.

For persons with elevated plasma triglycerides, reduced high-density lipoprotein (HDL) cholesterol, and small dense LDL cholesterol (the metabolic syndrome), improved glycemic control, modest weight loss, dietary saturated fat restriction, increased physical activity, and incorporation of monounsaturated fats may be beneficial.

Nephropathy

In individuals with microalbuminuria, reduction of protein to 0.8 to 1.0 g/kg body weight per day and in individuals with overt nephropathy, reduction to 0.8 g/kg body weight per day may slow the progression of nephropathy.

Catabolic illness

The energy needs of most hospitalized patients can be met by providing 2535 kcal/kg body weight.

Protein needs are between 1.0 to 1.5 g/kg body weight; the higher end of the range being for more stressed patients.

Prevention of diabetes

Structured programs which emphasize lifestyle changes including education, reduced fat and energy intake, regular physical activity and regular participant contact can produce long-term weight loss of 5%-7% of starting weight and reduce the risk for developing diabetes.

All individuals, especially family members of persons with type 2 diabetes, should be encouraged to engage in regular physical activity to decrease risk of developing type 2 diabetes.


often counterproductive to engage in scholarly discussion of the nature of the evidence base for their decision. At a minimum, discussion should include the
documented efficacy of more classical lifestyle and pharmacologic intervention and the idea that these efforts should not be left by the wayside when budgetary constraints affect potentially more effective interventions. A multivitamin containing at least 400 μg of folic acid is probably a reasonable nutritional supplement for most patients with type 2 diabetes, and supplementation with folic acid (1 mg), vitamin B12 (400 μg), and pyridoxine (10 mg) has been shown to reduce the rate of restenosis after coronary angioplasty, presumably by reducing homocysteine levels.

Although there are proponents of a wide range of dietary composition, there are few data to support these recommendations from long-term outcome studies of prescribed diets. Mixed meals containing 10% to 20% of calories from protein, no more than 10% of calories from saturated fat, and the remainder largely from monounsaturated fats (seeds, nuts, avocados, olives, olive oil, canola oil) and carbohydrates, particularly whole grains, fruit, vegetables, and low-fat milk, are probably more acceptable. The most readily available. There is evidence to suggest that a diet rich in complex carbohydrates and low in fat and animal protein is a beneficial component of comprehensive lifestyle management in the setting of cardiovascular disease.

Weight loss is a goal of many patients with and without diabetes and certainly is associated with improvements in glycemic control, insulin resistance, circulating lipids, and blood pressure. Numerous clinical studies demonstrate that intensive lifestyle programs involving frequent contact with patients, individualized counseling, and education aimed at reducing dietary fat and calorie intake coupled with regular physical activity and efforts to understand and control behaviors that result in overeating can produce modest weight loss that can be largely maintained with sustained effort.

Self-Monitoring of Blood Glucose

Self-monitoring of blood glucose (SMBG) has not been demonstrated in clinical trials to change outcomes in type 2 diabetes when evaluated in isolation. However, multiple diabetes management programs have been demonstrated to help reduce complications. In all of these, SMBG is an integral part of the process, suggesting that SMBG is at least a component of effective therapy. The frequency and type of monitoring in diabetes therapy should be determined in consultation with the patient, taking into account the nature of the diabetes, the overall treatment plan and goals, and the patient's abilities. SMBG is particularly recommended for all patients with type 2 diabetes taking insulin or sulfonylureas as it allows the identification of minimal or asymptomatic episodes of hypoglycemia. Although severe hypoglycemia is relatively rare in type 2 diabetes, it can have devastating consequences such as trauma or self-injury or change in the perceived ability of a patient to continue to live independently as a result of confusion or loss of consciousness. Also, it is essential to have patients critically assess the nature of any hypoglycemic symptoms that may occur. Many patients are fearful or overconcerned about hypoglycemia and routinely consume extra calories in response to a variety of life's circumstances, such as when they are hungry, nervous, or upset. Monitoring generally document that most symptoms in patients with type 2 diabetes are not related to hypoglycemia and should not be treated with excessive caloric consumption.

Timing of SMBG varies depending on the diabetes type. It is important to advise patients to vary the time of the day at which blood glucose levels are checked. For some patients, the highest blood glucose of the day is the morning glucose, whereas for others the highest is before bed. Particularly in early diabetes, gestational diabetes, and well-controlled diabetes, monitoring 1 to 2 hours after meals allows patients to assess the effect of their lifestyle and pharmacologic efforts in controlling postprandial glucose levels, which are the only glycomic abnormality present. Monitoring and thus targeting therapy at just one time of day can leave the patient with a less than ideal overall response to therapy.

When glucose control is poor, having patients concentrate on premeal glucose levels is adequate. Once the premeal glucose levels reach the middle to low 100s, many advocate that patients switch to checking 1- to 2-hour postprandial glucose levels because it amplifies the observed effect of diet on glycemic control and enables patients to see that moderate changes in meal plan, activity, and medications have a significant impact on glycemic control. Even after substantial inappropiate changes in food intake, activity, or timing or dose of medication, blood sugar values often return to near-normal levels overnight or by the time of the next meal.

The frequency of glucose monitoring needs to be matched to individual patients' needs and treatment. Many clinicians ask patients to monitor at least once a day, varying among before breakfast, lunch, supper, bedtime, and mid-sleep as well as with hypoglycemic symptoms. Others ask intensively insulin-treated patients to monitor with intensity similar to that described for patients with type 1 diabetes (four times per day before meals with weekly checks at least once after breakfast, lunch, supper, and at midsleep as well as with symptoms). Some ask for sets of glyemic readings more infrequently (e.g., fasting and 1 hour after the biggest meal). In the subset of patients who achieve stable blood glucose levels without significant hypoglycemia, it is generally appropriate to decrease the frequency of SMBG to a few times a week. It is critical that SMBG be performed enough that both patient and provider have a good understanding of both the adequacy of the treatment regimen and the stability of glycemic control.

It has been widely assumed that the benefits of SMBG stem from the effect of putting patients in a situation in which they can be in control of their own therapy. If patients are aware of the glycemic targets associated with the outcomes they seek to achieve, SMBG enables them to evaluate critically their response to therapy and assure themselves that they are reaching their goals. It is generally useful for patients to keep a daily diary of their SMBG results, not only so that they can assess their own progress but also so that their health care providers can be informed as well. Fortunately, many patients now have the technological ability to perform daily or more frequent SMBG, record the results as instructed, and discuss them with their health care team only at quarterly or semiannual visits despite the fact that their control is inadequate. Unless SMBG results are entirely within agreed to targets, they should be communicated and reviewed at least monthly with a member of the health care team by telephone, fax, mail, or e-mail or at an interim visit to trigger changes in therapy as the need arises. Unfortunately, such services are generally unremunerated and may
become an unsustainable burden on health care teams.

Finally, one of the most difficult areas in which to keep current is the area of available equipment and supplies, particularly for glucose monitoring. A useful resource in this regard is the annual Resource Guide, which comes out as the January issue of Diabetes Forecast, a magazine for lay people with diabetes and their families. It is available online at the ADA Web site (www.diabetes.org) by clicking on “Community and Resources” on the left and then “Diabetes Forecast” below and finally on “Back Issues” to find the most recent January issue.
Pharmacotherapy of Type 2 Diabetes

The revolution in the treatment of type 2 diabetes since 1995 in the United States has been driven by the release of multiple new classes of drugs that independently address different pathophysiological mechanisms that contribute to the development of diabetes. The available oral antidiabetic agents can be divided by mechanism of action into insulin sensitizers with primary action in the liver, insulin sensitizers with primary action in peripheral tissues, insulin secretagogues, and agents that slow the absorption of carbohydrates. Insulin therapy in the setting of type 2 diabetes effectively is a supplement to endogenous insulin secretion. The relative benefits of lifestyle intervention and the six classes of drugs available for the management of type 2 diabetes are found in Table 29-10 (Table Not Available). This area has been the subject of extensive reviews. Because of limitations of space, in the following discussion the principles outlined in these reviews are summarized and limited additional references provided.

Insulin Sensitizers with Predominant Action in the Liver: Biguanides

Metformin is the only biguanide available in the United States. Phenformin was removed from the United States market in the 1970s because of deaths associated with lactic acidosis. Phenformin and buformin remain available in some countries around the world. Although metformin has been available in Europe for almost 40 years, it has been marketed in the United States only since 1995. The precise mechanism of action of metformin is unknown. Its major activity is to reduce hepatic insulin resistance and thereby gluconeogenesis and glucose production. It has more inconsistently demonstrated effects to improve insulin sensitivity in peripheral tissues. Because of its limited duration of action, it is generally taken at least twice daily, although a sustained-release formulation is now available.

As biguanides do not increase insulin levels, they are not associated with a significant risk of hypoglycemia. The most common adverse events are gastrointestinal nausea, diarrhea, crampy abdominal pain, and dysgeusia. About one third of patients have some gastrointestinal distress, particularly early in their course of treatment. This distress can be minimized by starting with a low dose once daily with meals and titrating upward slowly (over weeks) to effective doses. Sustained-release metformin is associated with less frequent and severe upper gastrointestinal symptoms, the more common of the adverse effects of metformin, but can increase the frequency of diarrhea, a much less common adverse effect overall. The vast majority of patients note no adverse effects with metformin therapy, and at least 90% tolerate it adequately with long-term use. Perhaps as a result of clinical or subclinical gastrointestinal effects, metformin is associated with less weight gain than other antidiabetic agents and in some studies modest mean weight loss.

The other side effect of concern with metformin is lactic acidosis, which is quite rare and occurs almost exclusively in patients who are at high risk of developing lactic acidosis apart from metformin therapy. As a result, it is recommended that high-risk patients avoid use of metformin. The package insert suggests that metformin is absolutely contraindicated in patients with renal insufficiency as the drug is cleared renally; it states that the drug should not be used in males with a serum creatinine greater than or equal to 1.5 mg/dL and in females at 1.4 mg/dL.

Obviously, there is a complex relationship between serum creatinine and renal function. Thus, reasonable practice would generally involve avoiding the use of metformin entirely in patients with an estimated creatinine clearance from the Cockcroft-Gault equation of less than 50 mL/min and avoiding greater than half-maximal doses of metformin in patients with an estimated creatinine clearance between 50 and 70 mL/min. As a reminder, according to the Cockcroft-Gault equation, creatinine clearance equals [(140 age)(weight in kg)]/(72 × serum creatinine in mg/dL)

TABLE 29-10 -- Comparisons of Therapies for Type 2 Diabetes

(Not Available)

in males, and this is multiplied by 0.85 in females. Therefore, a 30-year-old, 250-pound construction worker with a creatinine of 1.6 has a normal creatinine clearance of 103 mL/min, whereas an 80-year-old, 110-pound woman with a creatinine of 0.8 has a low creatinine clearance of 44 mL/min. Metformin is also contraindicated in patients with congestive heart failure requiring treatment, in those with hepatic insufficiency, and in the setting of alcohol abuse. Caution is required in elderly people, patients with acute illness or poorly controlled chronic illness, and in the setting of simultaneous treatment with nephrotoxic drugs (e.g., contrast dye).

The glucose-lowering efficacy and the prevalence of adverse gastrointestinal effects increase proportionately in the dose range 500 to 2000 mg/day. The maximal dose of 2550 mg does not generally provide additional benefit beyond that seen at 2000 mg daily. A new formulation of metformin combined with glyburide has been developed to maximize glucose-lowering effectiveness through the synergy of using an insulin secretagogue and thereby minimize gastrointestinal effects and is available in tablets with metformin/glyburide ratios of 250/1.25, 500/2.5, and 500/5. The formulations containing proportionally lower glyburide doses (250/1.25 and 500/2.5) seem to perform similarly to the higher dose combinations and thus are preferred.

Arguably, metformin has the best record of accomplishment among oral antidiabetic agents in outcome studies. In the UKPDS, among overweight subjects, those randomly assigned to metformin not only had improvements in microvascular complications similar to those of subjects randomly assigned to insulin and sulfonylurea but also demonstrated a reduction in diabetes-related deaths and myocardial infarction. The validity of this observation has been challenged because of unusual responses in a subsequent subrandomization. The beneficial effect of metformin on macrovascular complications through mechanisms independent of glycemic control is certainly plausible and supported by such observations as metformin-associated modest reductions in LDL, triglycerides, blood pressure, and procoagulant factors.

Insulin Sensitizers with Predominant Action in Peripheral Insulin-Sensitive Tissues: Thiazolidinediones

The thiazolidinedione class of drugs, often termed TZDs or glitazones, has engendered great enthusiasm and controversy since the first agent, troglitazone, was approved in 1997. Troglitazone was withdrawn from the United States market in the year 2000, largely because the remaining agents (pioglitazone and rosiglitazone) were thought to be safer than troglitazone, with which rare fatal hepatotoxicity was associated. These agents are believed to work through binding and modulation of the activity of a family of nuclear transcription factors termed peroxisome proliferator-activated receptors (PPARs). They are associated with slow improvement in glycemic control over weeks to months in parallel with an improvement in insulin sensitivity and reduction of FFA levels.

Each of these agents varies in important ways with regard to potency, pharmacokinetics, metabolism, binding characteristics, and demonstrated lipid effects. At the same time, all are effective glucose-lowering agents that are remarkably well tolerated with weight gain and fluid retention (and associated edema formation and hemodilution) as the only significant adverse effects. There is no substantial evidence that these newer agents are associated with hepatotoxicity, but this record of safety has been established in the setting of careful liver function test monitoring. Therefore, it is important to continue to recommended that the glitazones not be used in patients with active hepatocellular disease or in patients with unexplained serum alanine aminotransferase (ALT) levels greater than 2.5 times the upper limit of normal as well as to recommend serum ALT monitoring before initiating therapy, every 2 months for the first year and intermittently thereafter.

Whether there are clinically important differences between the two currently available agents is hotly debated, but definitive answers await adequately powered head-to-head studies that are under way. The promise of the glitazone class to reverse or prevent the negative cardiovascular associations of insulin resistance in parallel with its demonstrated effect of improving insulin sensitivity is exciting but unproven and under formal study in a series of randomized prospective clinical trials.
Almost all of the available data with regard to vascular effects of this class come from studies with troglitazone and include the following tantalizing clinical associations: reduced carotid intimal medial thickness, normalization of vascular endothelial function, improvements in dyslipidemia, lower blood pressure, and improved fibroin and coagulation parameters. A second attribute of the glitazones that has generated great enthusiasm is an effect to improve insulin secretory dynamics in subjects with diabetes and IGT. These observations provide hope that glitazone therapy may be useful in preventing diabetes or in halting the progression of established diabetes, thereby reducing the need for additional drug therapy over time. It is critical to recognize that the proven effects of pioglitazone and rosiglitazone to date are limited to improvements in glycemic control and changes in lipid parameters.

The adverse effect that has engendered the greatest concern regarding this class of drugs is weight gain. Careful study indicates that the weight gain is a result of subcutaneous and not visceral fat accumulation and that there is, in fact, a reduction in visceral fat, hepatic fat, and intramyocellular fat. Thus, the weight gain observed with glitazones, while having obvious negative consequences from the cosmetic standpoint, is perhaps less likely to cause significant adverse cardiovascular effects. Both weight gain and fluid retention are more common and severe in patients with the greatest glycemic responses, making expectant management of these adverse effects mandatory. All patients prescribed glitazones should be counseled to redouble lifestyle efforts to minimize weight gain.

With regard to edema, with appropriate caution almost no one should need to withdraw from therapy as a result of fluid retention. The patients most likely to experience edema are those treated with insulin and those with preexisting edema. Thus, women, overweight patients, and those with diastolic dysfunction or renal insufficiency are at greatest risk. It is prudent to teach patients with preexisting edema how to assess pitting pretibial edema at home and suggest that they make a habit of checking nightly. If they note a pattern of increasing edema at home, patients can be instructed to restrict sodium intake, to start a diuretic, or to increase their diuretic by some specified quantity on their own as needed.

In the edematous patient, people treated with insulin, and those at higher risk of fluid retention, it is prudent to initiate therapy with the lowest marketed dose of glitazone. When patients return for their 2-month ALT check, if the glycemic response has been inadequate and significant edema has not developed, increasing the dose of glitazone further with continued expectant management of edema can be accomplished. Most patients with mild edema respond to a low-dose thiazide diuretic (e.g., hydrochlorothiazide [HCTZ] at 25 mg). In patients with more extensive edema, a combination of low-dose thiazide diuretic with moderate-dose loop diuretic is sometimes required. Anecdotal reports suggest that avoidance of nonsteroidal anti-inflammatory agents and dihydropyridine calcium channel blockers can reduce the frequency of edema as an adverse event. It should be noted that fluid retention to the point of congestive heart failure and anasarca has been reported and that in some patients edema is refractory to diuretic therapy.

### Insulin Secretagogues

Currently available insulin secretagogues all bind to the sulfonylurea receptor (SUR1), a subunit of the ATP-sensitive potassium channel (K\(_{ATP}\)) on plasma membrane of pancreatic beta cells. The SUR1 subunit regulates the activity of the channel, and also binds ATP and ADP, effectively functioning as a glucose sensor and trigger for insulin secretion. Sulfonylurea binding as well as increases in intracellular ATP and decreases in ADP as a result of fuel metabolism lead to closing of the channel. The membrane depolarization that ensues causes the opening of voltage-dependent L-type calcium channels. Subsequent calcium influx results in an increase in intracellular calcium, which leads to insulin secretion. Differences in pharmacokinetic and binding properties of the various insulin secretagogues result in the specific responses that each agent produces. The major difference between them seems to be related to duration of action and to fairly subtle variations in their hypoglycemic potential.

### Sulfonylureas

The sulfonylureas have been available since the 1950s. They have a relatively slow onset of action and variable duration of action. There are numerous choices available (Table 29-11), which can be divided into first- and second-generation agents. In general, the second-generation agents are more potent and as a result have fewer adverse effects and drug-drug interactions. Glipizide-GITS (gastrointestinal therapeutic system) and gliclazide are preferred agents as they can be given as a once-daily dose (without additional effect with twice-daily dosing) in the vast majority of patients and involve a relatively low risk of hypoglycemia and weight gain. Glimepiride and glipizide-GR are the most commonly prescribed insulin secretagogues despite the fact that essentially all marketed oral secretagogues have been shown to have a significantly lower hypoglycemic potential.

An unusual characteristic of sulfonylureas is that the maximum marketed dose is generally two to four times higher than the maximally effective dose. There has been concern over the years that sulfonylureas may result in increased arrhythmic cardiovascular events in patients with diabetes as a result of their activity on vascular and cardiac SUR2 receptors with an effect of blunting ischemic preconditioning, a protective autoregulatory mechanism in the heart. There is some evidence to suggest that this may be less likely to occur with glipizide than with gliburide, but it is also a rationale to avoid high-dose sulfonylurea therapy. Sulfonylureas are arguably the most cost-effective glucose-lowering agents and therefore are clearly worthy of their widespread use. In general, limiting the dose to one-fourth maximal, unless higher doses are clearly demonstrated to provide significant benefits in glycemic control, minimizes both costs and adverse events. Small doses of sulfonylurea (e.g., 0.5 to 1 mg of gliburide or 2.5 mg of glipizide-GITS) are remarkably effective, particularly in patients on concomitant insulin-sensitizing therapy, and are almost uniformly well tolerated.

### Repaglinide

Repaglinide is a member of the meglitinide family of insulin secretagogues, distinct from the sulfonylureas. It has a short half-life and a distinct SUR1 binding site. As a result of more rapid absorption, it produces a generally faster and briefer stimulus to insulin secretion. As a result, it is generally taken with each meal and provides better postprandial control and generally less hypoglycemia and weight gain than glyburide. Repaglinide does seem to have a longer residence time on the sulfonylurea receptor and a prolonged effect on fasting glucose despite the fact that its pharmacologic half-life is quite short. Repaglinide is available in 0.5, 1, and 2 mg tablets. The maximal dose is 4 mg with each meal. As is the case with the sulfonylureas, there is a modest glucose-lowering advantage of high doses versus moderate doses of repaglinide.

### Nateglinide

Nateglinide is a derivative of phenylalanine, structurally distinct from both sulfonylureas and the meglitinides. It has a quicker onset and shorter duration of action than repaglinide. Its interaction with SUR1 is fleeting. As a result, its effect in lowering postprandial glucose is quite specific and it has little effect in lowering fasting glucose. This provides advantages (less hypoglycemia) and disadvantages (less overall glucose-lowering effectiveness). Therefore, nateglinide is most appropriately used when fasting glucose levels are modestly elevated in the setting of early diabetes or in combination with insulin sensitizers or long-acting evening insulin. Nateglinide is available as 120-mg tablets and is taken with each meal. A 60-mg tablet is available but is not generally used except in patients with minimal hyperglycemia.

The rationale for stimulating insulin secretion in a way that minimizes fasting hyperinsulinemia and maximizes postprandial control is compelling. Furthermore, these newer agents demonstrate little binding to the vascular smooth muscle and cardiac SUR2 receptors. However, the use in the United States of these newer glinide agents has been modest, in part because of the need for multiple daily doses, greater expense than with sulfonylureas, and lack of head-to-head comparative studies that demonstrate superiority over newer sulfonylureas already perceived as having low potential for producing hypoglycemia and weight gain.

### Glitazones

- **Carbohydrate Absorption Inhibitors:** -Glucosidase Inhibitors

- Glucosidase inhibitors (AGIs) work to inhibit the terminal step of carbohydrate digestion at the brush border of the intestinal epithelium. As a result, carbohydrate absorption is shifted more distally in the intestine and is therefore delayed, allowing the sluggish insulin secretory dynamics characteristic of type 2 diabetes to catch up with carbohydrate absorption. There are two currently available agents, acarbose and miglitol. Their use in the United States has been limited by a number of factors, including the need to administer the medication at the beginning of each meal, flatulence as a common side effect, and only modest reductions in blood glucose. These factors should be balanced against the AGI’s ability to lower postprandial glucose, thereby improving glycemia without increasing weight or hypoglycemic risk. Even though they may potentially lower glucose in everyone, the extent of the lowering is generally modest, calling into question their utility in the
of insulin in type 2 diabetes is designed to supplement endogenous production of insulin both in the basal state to modulate hepatic glucose production and in the postprandial state, in which a surge in insulin release normally facilitates glucose clearance into muscle and fat for storage to allow intraprandial metabolism. Currently, the vast majority of insulin used worldwide is of recombinant human origin. The available formulations largely differ in their pharmacokinetics as reviewed in Table 29-12.

Insulin lispro and insulin aspart are rapid-acting insulin analogues that have an onset of action in 5 to 15 minutes, peak activity in approximately 1 hour, and a duration of activity of approximately 4 hours. Regular insulin is approximately half as fast as the rapid-acting analogue with onset in 30 minutes, a peak at 2 to 4 hours, and a duration of action of 6 to 8 or more hours. Intermediate-acting insulin analogues have been shown to provide a modest reduction in hypoglycemia. Insulin analogues have been shown to provide a modest reduction in hypoglycemia. Insulin analogues are rare, as are chronic skin reactions like lipodystrophy and lipoatrophy. It should be noted that the absolute risk of severe hypoglycemia in patients with type 2 diabetes is relatively small, approximately one third to one tenth as high as in similarly treated patients with type 1 diabetes. This risk can be further minimized with appropriate education of patients and expectant home glucose monitoring at times when unrecognized hypoglycemia is most likely to occur.

<table>
<thead>
<tr>
<th>Generic Name</th>
<th>Maximum Daily Dose</th>
<th>Equivalent Doses (mg)</th>
<th>Duration of Action</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetohexamide 250 mg</td>
<td>1500 mg, divided bid</td>
<td>500</td>
<td>Intermediate 1218 hr</td>
<td>Metabolized by liver to active metabolite (twice as potent as parent compound). Has diuretic activity. Has uricosuric activity.</td>
</tr>
<tr>
<td>Chlorpropamide 100 mg</td>
<td>750 mg qd (500 mg in older patients)</td>
<td>250</td>
<td>Very long 60 hr</td>
<td>70% metabolized by liver to less active metabolites; 30% excreted intact by kidneys. Can potentiate ADH. Disulfiram (Antabuse) like reaction with alcohol occurs in nearly a third of patients.</td>
</tr>
<tr>
<td>Tolazamide 100 mg</td>
<td>1000 mg, divided bid</td>
<td>250</td>
<td>Intermediate 1224 hr</td>
<td>Metabolized by liver to less active and inactive products. Has diuretic activity.</td>
</tr>
<tr>
<td>Tolbutamide 250500 mg</td>
<td>3000 mg, divided bid or tid</td>
<td>1000</td>
<td>Short 612 hr</td>
<td>Metabolized by liver to inactive product.</td>
</tr>
<tr>
<td>Glipizide 5 mg</td>
<td>40 mg, divided bid</td>
<td>5</td>
<td>Intermediate 1224 hr</td>
<td>Metabolized by liver to inactive products that are excreted in the urine and, to a lesser extent, in the bile. Mild diuretic activity.</td>
</tr>
<tr>
<td>Glipizide-GITS 5 mg</td>
<td>20 mg qd</td>
<td>Long &gt; 24 hr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glyburide 2.5 mg</td>
<td>20 mg, divided bid</td>
<td>5</td>
<td>Intermediate 1624 hr</td>
<td>Metabolized by liver to weakly active and inactive products, excreted in urine and bile. Mild diuretic activity. Highest risk of hypoglycemia.</td>
</tr>
<tr>
<td>Micronized glyburide 3 mg</td>
<td>6 mg bid</td>
<td>3</td>
<td>Shorter</td>
<td></td>
</tr>
<tr>
<td>Glimepiride 1 mg</td>
<td>8 mg qd</td>
<td>2</td>
<td>Long &gt; 24 hr</td>
<td>Metabolized to inactive metabolites by liver, excreted in urine and bile.</td>
</tr>
</tbody>
</table>

GITS, gastrointestinal therapeutic system.

Adapted from Facts and Comparisons, drug information monthly update service. St. Louis, JB Lippincott.

Long-acting insulin analogue with distinctive properties. It provides a flat, peakless profile of activity with a duration of action of more than 24 hours in most patients. Premixed insulin formulations provide greater convenience and accuracy of mixing than those mixed by patients. Premixed formulations available in the United States are 70/30 and 50/50 mixtures of NPH and regular insulin, a 75/25 mixture of lispro insulin in its NPH-like formulation with insulin lispro, and a 70/30 mixture of insulin aspart with its NPH-like congener. Premixed insulin provides a profile of activity as expected from the addition of the activities of its components.

Adverse events associated with insulin are well known and include weight gain and hypoglycemia. It is interesting that both fast-acting and long-acting insulin analogues have been shown to provide a modest reduction in hypoglycemia. Insulin allergies are rare, as are chronic skin reactions like lipodystrophy and lipoatrophy. It should be noted that the absolute risk of severe hypoglycemia in patients with type 2 diabetes is relatively small, approximately one third to one tenth as high as in similarly treated patients with type 1 diabetes. This risk can be further minimized with appropriate education of patients and expectant home glucose monitoring at times when unrecognized hypoglycemia is most likely to occur.

Long-acting insulin analogue with distinctive properties. It provides a flat, peakless profile of activity with a duration of action of more than 24 hours in most patients. Premixed insulin formulations provide greater convenience and accuracy of mixing than those mixed by patients. Premixed formulations available in the United States are 70/30 and 50/50 mixtures of NPH and regular insulin, a 75/25 mixture of lispro insulin in its NPH-like formulation with insulin lispro, and a 70/30 mixture of insulin aspart with its NPH-like congener. Premixed insulin provides a profile of activity as expected from the addition of the activities of its components.

Insulin analogues have been shown to provide a modest reduction in hypoglycemia. Insulin analogues are rare, as are chronic skin reactions like lipodystrophy and lipoatrophy. It should be noted that the absolute risk of severe hypoglycemia in patients with type 2 diabetes is relatively small, approximately one third to one tenth as high as in similarly treated patients with type 1 diabetes. This risk can be further minimized with appropriate education of patients and expectant home glucose monitoring at times when unrecognized hypoglycemia is most likely to occur.

Long-acting insulin analogue with distinctive properties. It provides a flat, peakless profile of activity with a duration of action of more than 24 hours in most patients. Premixed insulin formulations provide greater convenience and accuracy of mixing than those mixed by patients. Premixed formulations available in the United States are 70/30 and 50/50 mixtures of NPH and regular insulin, a 75/25 mixture of lispro insulin in its NPH-like formulation with insulin lispro, and a 70/30 mixture of insulin aspart with its NPH-like congener. Premixed insulin provides a profile of activity as expected from the addition of the activities of its components.

Adverse events associated with insulin are well known and include weight gain and hypoglycemia. It is interesting that both fast-acting and long-acting insulin analogues have been shown to provide a modest reduction in hypoglycemia. Insulin allergies are rare, as are chronic skin reactions like lipodystrophy and lipoatrophy. It should be noted that the absolute risk of severe hypoglycemia in patients with type 2 diabetes is relatively small, approximately one third to one tenth as high as in similarly treated patients with type 1 diabetes. This risk can be further minimized with appropriate education of patients and expectant home glucose monitoring at times when unrecognized hypoglycemia is most likely to occur.

Long-acting insulin analogue with distinctive properties. It provides a flat, peakless profile of activity with a duration of action of more than 24 hours in most patients. Premixed insulin formulations provide greater convenience and accuracy of mixing than those mixed by patients. Premixed formulations available in the United States are 70/30 and 50/50 mixtures of NPH and regular insulin, a 75/25 mixture of lispro insulin in its NPH-like formulation with insulin lispro, and a 70/30 mixture of insulin aspart with its NPH-like congener. Premixed insulin provides a profile of activity as expected from the addition of the activities of its components.

Adverse events associated with insulin are well known and include weight gain and hypoglycemia. It is interesting that both fast-acting and long-acting insulin analogues have been shown to provide a modest reduction in hypoglycemia. Insulin allergies are rare, as are chronic skin reactions like lipodystrophy and lipoatrophy. It should be noted that the absolute risk of severe hypoglycemia in patients with type 2 diabetes is relatively small, approximately one third to one tenth as high as in similarly treated patients with type 1 diabetes. This risk can be further minimized with appropriate education of patients and expectant home glucose monitoring at times when unrecognized hypoglycemia is most likely to occur.

Long-acting insulin analogue with distinctive properties. It provides a flat, peakless profile of activity with a duration of action of more than 24 hours in most patients. Premixed insulin formulations provide greater convenience and accuracy of mixing than those mixed by patients. Premixed formulations available in the United States are 70/30 and 50/50 mixtures of NPH and regular insulin, a 75/25 mixture of lispro insulin in its NPH-like formulation with insulin lispro, and a 70/30 mixture of insulin aspart with its NPH-like congener. Premixed insulin provides a profile of activity as expected from the addition of the activities of its components.
A significant challenge in clinical decision making in diabetes is that the increased availability of therapeutic options for antidiabetic therapy is ahead of adequate prospective outcome studies. Currently available clinical trial data have not identified the preferred agents in type 2 diabetes, either as initial therapy or in subsequent care. Each class of drugs and even agents within each class have advantages and limitations, and individual issues may significantly affect the appropriate choice of therapy in particular patients. Table 29-10 (Table Not Available) highlights some of the relative advantages and disadvantages of various agents and classes.

A general approach in the absence of any patient-specific factors is suggested in the algorithm presented in Figure 29-20. A growing body of evidence indicates that the use of metformin as initial therapy in combination with diet and exercise can provide impressive lowering of glucose with essentially no risk of hypoglycemia. Because this agent is available as a generic preparation, relative cost is low, and if the response is judged to be inadequate a thiazolidinedione, sulfonylurea, or glinide can be added. It has been proposed that the use of metformin alone or in combination with a thiazolidinedione may lead to a greater reduction in cardiovascular risk than similarly effective (with respect to glycaemia) approaches that increase insulin levels. At present, the data are not definitive on this point.

Patients with higher levels of glucose (generally FPG > 200 mg/dL) almost always require agents to increase insulin levels. Because insulin, sulfonylureas, and glinides provide much faster improvements in overall control that metformin, glitazones, or AGIs, they are preferred in patients with higher levels of glucose either as monotherapy or as part of initial combined therapy. Starting a patient with a low dose of a glimepiride, glipizide-GITS, or insulin combined with either metformin, glitazone, or AGI is a reasonable initial approach to the poorly controlled condition.

In patients who have reasonable control of fasting and preprandial plasma glucose levels (more than 50% of values less than 130 mg/dL) whose overall control as assessed by HbA1c, is still higher than desired, monitoring may be either inaccurate or ineffective or postprandial plasma glucose (PPG) levels may be elevated. As it can be more difficult to have patients monitor in the postprandial state, it is important to remember that without specific therapy, almost all patients with type 2 diabetes have elevated PPG. Thus, in such patients, targeting presumed PPG elevations with the use of AGIs, glinides, or rapid-acting insulin analogues can theoretically lower average glucose with a lower risk of weight gain and hypoglycemia than with sulfonylureas or long-acting insulin.

The most critical issue in long-term glycemic management is that of continuously reassessing with patients the adequacy of their control, examining glucose monitoring logs and HbA1c values, and refining treatment regimens to achieve optimal control with the lowest dose of the least number of medications. Most patients in specialty care require two or more drugs to achieve recommended targets. Many patients require three or more (particularly if you consider long-acting and short-acting insulin analogues) and many of the three-drug combinations have been evaluated in modest-sized studies and have been shown to be safe and effective. Generally, it is preferred to add agents if there was an improvement in control with the first agent selected and to continue to add agents as needed to achieve goals. Subsequent back-titration to optimize treatment is often possible when glycemic goals are achieved. The selection of initial therapy should be based on mutually (patient and provider) recognized priorities. Increasingly, practitioners are using submaximal doses of agents in combination to increase the ratio of efficacy to adverse effects and in recognition of the potential synergy of sensitizers and secretagogues as well as the value of treatment of postprandial glucose and fasting glucose in combination therapy.

When adding insulin in the management of inadequately controlled type 2 diabetes, some practitioners prefer to stop the oral antidiabetic agents and switch to insulin. Most generally continue the oral agents and add an evening dose of insulin. Classically, bedtime NPH insulin and more lately bedtime insulin glargine have been preferred for initiating insulin therapy. In more overweight patients (>120% of ideal body weight), the use of mixed insulin (or premixed insulin) at supper can help clear glucose elevations after the evening meal, generally the largest meal of the day. This works quite well in most patients, although some experience nocturnal hypoglycemia, which is less common with mixtures employing rapid-acting insulin analogues. There are data suggesting that glargine given at bedtime can similarly provide for lower morning glucose values with less nocturnal hypoglycemia than NPH insulin, particularly in more overweight patients. Many patients eventually require more complex regimens—twice-daily injections, split-mix insulin, multiple injection regimens, and rarely insulin pump therapy. It should be noted that a minority of patients with type 2 diabetes have a better response to insulin administered in the morning than in the evening.

It is important that both patient and health care provider agree on how to reach the goals of therapy. Therefore, biases and concerns of the patient should be addressed when trying to determine which agent should be prescribed. These biases can be elucidated in interviews with patients through discussions of various strategies.

**Strategies**

**Minimal Cost Strategy**

For a large proportion of patients, particularly those who are elderly, drug costs are an overwhelming issue. Diet and exercise can be extremely effective and almost free. The least expensive drugs for the treatment of diabetes are the sulfonylureas; metformin has become available in generic formulations. Thus, a minimum cost strategy could start with a sulfonylurea and progress to the addition of generic metformin or bedtime or presupper insulin and finally two or more insulin injections per day if necessary. In the Veterans Administration Cooperative Study, excellent control was achieved in the context of a comprehensive program of diabetes education using a combination of daytime sulfonylurea and evening insulin. Although insulin is relatively inexpensive, in high doses (1 U/kg or more) the costs begin to rise, conversely insulin for adding metformin or a thiazolidinedione. It should be noted that most pharmaceutical companies have programs to provide no-cost or low-cost medication to the poor. Many of these are listed with links at www.needymeds.com. Furthermore, for increasing numbers of patients, the major driving force in their drug expenses is the number of prescriptions as each is associated with a copayment, providing a rationale for using combination agents.

**Minimum Weight Gain Strategy**

Weight gain associated with the treatment of diabetes is of concern to most clinicians and is often an overriding issue with patients. A strategy to minimize weight gain would emphasize diet and exercise and would almost certainly employ metformin or an AGI as initial therapy with the addition of the other agent if one was inadequate. As sulfonylureas and repaglinide seem to have a modest weight-sparing effect in combination therapy with insulin, one or the other could be added before insulin administration in such a strategy. As discussed earlier, the weight gain associated with thiazolidinediones, although certainly a cosmetic issue, may not be associated with increased cardiovascular risk.

**Minimal Injection Strategy**

Too many patients are determined to avoid insulin injections at any cost. The minimal injection strategy involves sulfonylureas, metformin, AGIs, and thiazolidinediones, which can be added in any order. Insulin, probably as a bedtime or presupper dose to minimize the inconvenience, would be added only if absolutely necessary. The strategy of using thiazolidinediones early in the course of diabetes in the hope that this may reduce the rate of progressive beta cell dysfunction remains unproved. It is important to try to dispel notions that insulin therapy is difficult, ominous, or fraught with peril by highlighting its efficacy and the great strides that have been made in insulin formulations and delivery devices. Most patients require insulin at some point in their lifetime.
Minimal Insulin Resistance Strategy

The possible atherogenic effects of insulin have been widely touted in the lay press and by marketing programs within the pharmaceutical industry. The relationship between circulating insulin levels and cardiovascular risk in nondiabetic populations is incontrovertible but probably related to the presence of insulin resistance rather than the insulin concentrations per se. Furthermore, in essentially all studies of intensive management with insulin, improved outcomes were observed with insulin treatment. There are no clinical data to suggest that exogenous insulin is associated with adverse side effects or long-term complications beyond its hypoglycemic effects and the associated weight gain. In any case, this strategy is analogous to the minimal injection strategy except that the order of introduction of agents is perhaps important. The thiazolidinediones have the greatest efficacy in reducing insulin resistance, metformin is second, and AGIs are third, with nateglinide associated with more specific stimulation of insulin levels after meals than the other insulin secretagogues, which all increase peripheral insulin levels less than injected insulin.

Minimal Effort Strategy

Many patients are capable of making only a minimal effort with regard to their diabetes. Questioning patients about their pill-taking history and their realistic ability to comply with a prescribed frequency of therapy is important. Taking a once-a-day sulfonylurea or thiazolidinedione requires the least effort by the patient. Taking bedtime insulin is actually relatively well accepted by patients to whom this consideration is important. Developing strategies to improve adherence and increase motivation is certainly a long-term goal in this population.

Hypoglycemia Avoidance Strategy

This is another important consideration for many patients. The AGIs have been reported in small studies to reduce "reactive" hypoglycemia. Other oral agents could be added in any order with the exception that insulin secretagogues would be added last, their dose minimized, and glyburide avoided. Nateglinide in particular among the secretagogues is associated with an exceptionally low risk of significant hypoglycemia. The insulin analogues are associated with a lower risk of hypoglycemia than human insulin.

Postprandial Targeting Strategy

Achieving postprandial glucose targets is generally associated with better control than just meeting premeal targets. On the basis of epidemiologic studies, it has been suggested that PPG is more highly correlated with cardiovascular disease risk than fasting glucose levels. Correction for confounding variables such as components of the multiple metabolic syndrome has not been performed, however. Furthermore, there are no outcome studies that have demonstrated the superiority of these approaches in the setting of type 2 diabetes. Control of postprandial glycemia can be achieved only with specific lifestyle efforts and pharmacologic agents, which target postprandial glucose. Postprandial glucose monitoring is helpful in this regard as it reinforces the goals and is the most effective measure to assess the effectiveness of treatment. Techniques that can improve postprandial control include lowering the carbohydrate content of meals, adding fiber, substituting monounsaturated fats for carbohydrates, encouraging physical activity after meals, adding AGIs with meals, and using rapidacting insulin analogues. Nateglinide and repaglinide provide a theoretical advantage in this situation compared with other secretagogues, although formal head-to-head studies have not been completed comparing the glinides with gliclazide and glipizide-GITS.
Prevention of Type 2 Diabetes

The possibility that type 2 diabetes can be prevented in high-risk individuals has been formally tested in a series of large-scale clinical trials. The Da Qing study randomly assigned clinics in an industrial city in China to a dietary intervention, exercise intervention, combined diet and exercise, or no intervention at all. Among the clinics, 577 subjects with IGT were studied. In this study, the interventions were quite modest and conducted largely in group settings. All three interventions led to reductions in the risk of conversion to diabetes of 31% to 46% compared with the control groups.\[^{691}\]

In a Finnish study, a similar number of middle-aged obese subjects with IGT were randomly assigned to a control group that received minimal lifestyle advice or to intensive, individualized instruction on food intake, increased physical activity, and weight reduction. The intensive lifestyle therapy group demonstrated a 58% relative risk reduction compared with the control group in the incidence of diabetes.\[^{692}\]

In the United States, the Diabetes Prevention Program enrolled over 3000 middle-aged, overweight subjects with IGT including substantial representation from high-risk minority groups. The intensive lifestyle group in this study also demonstrated a 58% relative risk reduction in the progression to diabetes.\[^{690}\]

In the Diabetes Prevention Program, there was another arm of the study that evaluated the ability of metformin at 500 mg twice a day to prevent the development of diabetes. It was moderately successful, with a 31% relative risk reduction in the progression of diabetes, although the benefit seemed to be greater in younger, more overweight, and more hyperglycemic subjects. In other studies not yet fully published, other oral antidiabetic agents (troglitazone in the Troglitazone in the Prevention of Diabetes (TRIPOD) study\[^{690A}\] and acarbose in the STOP-NIDDM study\[^{690B}\]) have been reported to reduce the risk of developing diabetes. Patients and families as well as health care professionals are excited about the possibilities of preventing the disease.

The success of the lifestyle interventions is impressive, demonstrating conclusively that with a variety of techniques it is possible for patients to achieve physiologically relevant changes in body weight. Medications overall had less positive impact than lifestyle intervention, although troglitazone did perform remarkably well in diabetes prevention. The questions that arise from these results are how to screen for people at risk and what intervention should be initiated in those with an interest in prevention.

It seems reasonable to screen on the basis of current recommendations as outlined earlier primarily for case finding but also recognizing that patients with abnormal glucose values (fasting greater than 110 mg/dL or IGT with an OGTT) would be ideal candidates for preventive strategies. Certainly, high-risk individuals should be counseled on nutritional approaches to achieve weight loss, instructed to increase physical activity, and observed prospectively to determine whether progression of hyperglycemia has occurred. Treatment for other cardiovascular risk factors should also be considered if they are present. In the absence of outcome studies, it is difficult to recommend drug therapy to prevent diabetes because significant complications are unlikely to develop in the short window of time during which glucose levels increase from a fasting glucose of 110 to 126 mg/dL. An extension phase of the Diabetes Prevention Program that is under way should provide evidence concerning whether prevention or delay in the development of diabetes will prevent death or disability.
Future Directions

The present-day management of type 2 diabetes is significantly more effective and easier for patients than the situation that prevailed even 10 years ago. A better understanding of the barriers to effective diabetes management and how to overcome them would be of great benefit. The epidemic in diabetes and obesity that is under way coupled with the predicted early death and disability that follow threatens to overwhelm our health care system. Practical, cost-effective public health approaches to stem this tide are desperately needed.

Novel pharmaceutical agents including glucagon receptor antagonists, inhibitors of gluconeogenic and glycogenolytic pathways, activators of the insulin signaling pathways, modifiers of lipid metabolism, and antiobesity agents are areas of early pharmaceutical development.

There is tremendous interest in developing novel PPAR modulators that preserve the glucose-lowering effectiveness of the glitazones, enhance the lipid benefits, and mitigate the effects on fluid retention and weight gain. As these PPAR-active agents are thought to exert their action in the nucleus, there is reason to believe that such a goal is achievable. There are dozens of compounds in early stages of development and several already in phase III trials. The major barrier to success in this arena is the lack of well-validated animal models to predict human responses. Novel methods of insulin delivery similarly have generated a great deal of enthusiasm among patients, particularly techniques to deliver insulin orally or by inhalation. There is considerable controversy in the endocrine community in this regard. On the one hand, it seems unlikely that oral insulin delivery would be efficient enough to treat insulin resistance effectively; on the other hand, any delivery into the portal system could be more effective and perhaps have an improved safety profile because of preferential inhibition of hepatic glucose production.

The area of gut hormones offers promise for advances in treatment of type 2 diabetes. Amylin, the second beta cell hormone, is known to act centrally to suppress postprandial glucagon secretion, slow gastric emptying, and increase satiety. Synthetic amylin is in late-phase trials, and although only modestly effective as a glucose-lowering agent, it does seem to be well tolerated and is associated with weight reduction.

GLP-1 is a gut hormone secreted from intestinal cells that has an overlapping but generally more robust profile of action than amylin with the additional effect of preserving functioning beta cell mass. GLP-1 is rapidly degraded in the circulation, and thus inhibitors of the degrading enzyme (dipeptidylpeptidase IV [DP-IV]) as well as DP-IVresistant analogues are being investigated in clinical trials. Additional studies will determine whether glucagon-like peptides provide the next major class of antidiabetic agents.
References


Hellman BL. The variation in blood lipid levels described by various measures of overall and abdominal obesity in Danish men and women aged 35-65 years. Eur J Clin Nutr 1992; 46:597605.


Evans DJ, Hoffmann RG, Kalkhoff RK, Kisselbach AH. Relationship of body fat topography to insulin sensitivity and metabolic profiles in premenopausal women. Metabolism 1984; 33:6875.


References


1473


References


324. Bavenholm PN, Pigon J, Ostenson CG, Efendic S. Insulin sensitivity of suppression of endogenous glucose production is the single most important determinant of glucose tolerance. Diabetes 2001; 50:14491454.


References


147


References


In 1984, Sutherland and co-workers transplanted the tail of the pancreas from nondiabetic identical twins to their twin mates with type 1 diabetes. In contrast to the transplantation of organs such as kidneys, in which the transplants are accepted between identical twins, pancreatic islets but not acinar pancreas were rapidly destroyed. The diabetes of the twin transplant recipients was cured for only a matter of weeks. In retrospect, the results of these transplants were predictable, given the autoimmune nature of type 1A diabetes and similar results in animal models of the disorder. Following this clinical study, type 1 diabetes became one of the most intensively studied autoimmune disorders, and the National Institutes of Health has designated type 1A diabetes a Priority One target for the development of a preventive immunologic vaccine. Knowledge of the immunogenetics and immunopathogenesis of type 1A diabetes is beginning to influence clinical care, greatly influences current clinical research, and will, we hope, lead to disease prevention.
DIFFERENTIAL DIAGNOSIS OF TYPE 1 DIABETES

An expert committee of the American Diabetes Association, with its etiologic diagnostic criteria (Table 30-1), has recommended dividing type 1 diabetes into type 1A (immune-mediated) and type 1B (other forms of diabetes with severe insulin deficiency). At the onset of diabetes, distinguishing type 1A diabetes from type 2 diabetes, let alone type 1B diabetes, is not always a simple task. The best current criterion for diagnosis of type 1A diabetes is the presence of anti-islet autoantibodies measured with highly specific (and reasonably sensitive) autoantibody radioassays.

The presence of autoantibodies with assays defined as positive in less than 1 of 100 control subjects (specificity 99%) is reasonably diagnostic of type 1A diabetes. Non-Hispanic white children presenting with diabetes usually have type 1A diabetes, whereas adults older than 40 years usually have type 2 diabetes. More than 90% of such children presenting with diabetes express one of three commonly measured autoantibodies (see later). In contrast, among black or Hispanic American children, almost one half lack any autoantibody. Most of these children appear to have an early age of onset of type 2 diabetes mellitus, and many have attendant risk factors such as obesity and lack human leukocyte antigen (HLA) alleles associated with type 1A diabetes (see later). Imagawa and co-workers described an unusual form of diabetes. The patients had normal hemoglobin A\textsubscript{1c} (HbA\textsubscript{1c}) despite severe hyperglycemia, suggesting that the diabetes had been present for only a short time. Histologic examination of pancreatic sections demonstrated pancreatitis but no insulitis, and anti-islet autoantibodies were not detected. It is likely that this represents one of the first examples of type 1B diabetes although a fulminant type 1A is possible.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Islet Autoantibodies</th>
<th>Genetics</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 1A diabetes</td>
<td>Autoantibody positive; &gt;90%</td>
<td>30-50% DR3 and DR4</td>
<td>Children: 90% DR3 or DR4; 90% non-Hispanic white type 1A</td>
</tr>
<tr>
<td>Type 1B diabetes</td>
<td>Autoantibody negative</td>
<td>Unknown</td>
<td>Type 1B rare in whites</td>
</tr>
<tr>
<td>Type 2 diabetes</td>
<td>Autoantibody negative</td>
<td>Unknown</td>
<td>If Ab+ likely LADA (latent autoimmune diabetes adults) and then HLA similar to type 1A</td>
</tr>
<tr>
<td>Other forms of diabetes mellitus</td>
<td>Autoantibody negative</td>
<td>MODY mutations, other syndromes</td>
<td></td>
</tr>
</tbody>
</table>

Ab, antibody; HbA\textsubscript{1c}, hemoglobin A\textsubscript{1c}; HLA, human leukocyte antigen; MODY, maturity-onset diabetes of youth.

Obesity does not protect an individual from the development of type 1A diabetes, although it is usually associated with insulin-resistant forms of diabetes. It is also important to realize that individuals can have both insulin resistance and type 1A diabetes, and such autoantibody-positive individuals with both can present with diabetes with high levels of fasting insulin or C peptide but loss of stimulated insulin secretion. With current assays for anti-islet autoantibodies, a subset of children with type 1A diabetes are negative for the autoantibodies. Some children (relatively uncommon), as they progress to diabetes, lose expression of all autoantibodies by the time of diagnosis. Such type 1A autoantibody-negative children typically have HLA alleles associated with type 1A diabetes, are not insulin resistant, may present with ketoacidosis, and, with time, lose C peptide secretion. The diagnosis is not, however, clear at diabetes onset, and perhaps designations such as anti-islet autoantibody positive and negative are more accurate with current laboratory tests.
ANIMAL MODELS OF TYPE 1A DIABETES

In relation to other autoimmune disorders, type 1A diabetes is unusual in having a series of spontaneous animal models of the disease. These animal models provide clues to potential mechanisms of pathogenesis and allow testing of therapies for disease prevention. Tests of new therapies may not be applicable to humans because many more are efficacious in certain animal models than in humans. It should also be recognized that many of the spontaneous animal models are inbred and thus not diabetic at any locus, whereas all humans have different alleles at tens of thousands of loci. Thus, each animal model may or may not provide insights into one of the forms of human diabetes.

Despite the preceding caveats, the animal models are remarkably similar to humans in a number of key immunologic parameters. The most notable include the importance of the major histocompatibility complex (MHC) for disease and the presence of lymphocytic islet invasion followed by specific destruction of islet beta cells. For reasons that are currently unclear, given specific HLA molecules, humans, rats, and mice have a marked propensity for autoimmunity directed at islet beta cells. This propensity may be related to a specific lack of tolerance to a single islet molecule such as insulin, to specific sensitivity of islet beta cells to immune-mediated destruction, or to factors not currently appreciated. It is likely that understanding this propensity will lead to effective therapies.

Polygenic Spontaneous Animal Models

NOD Mouse

The nonobese diabetic (NOD) mouse is the most intensively studied animal model. As with type 1A diabetes of humans, specific HLA class II and class I (see later) molecules are central for disease pathogenesis. The NOD mouse has mutations that cause absence of the I-E (histocompatibility) molecule (similar to human DR) and an unusual I-A (similar to human DQ). The I-A molecule in the NOD mouse is termed I-A<sup>g7</sup> which designates a specific amino acid sequence. HLA class II molecules (in humans there are three, DP, DQ, and DR) function to bind peptides and present these peptides to the T-cell receptor of CD4 (“helper”) T lymphocytes. The genes were termed immune response genes because common variations in their sequences (allelic variation) determine the peptides to which an individual mouse or person can mount a T-cell response. Thus, a central role for these molecules in immune function and autoimmunity is expected. If the lack of I-E expression is corrected in the NOD mouse with introduction of an I-E transgene, diabetes is prevented. If a different I-A sequence is introduced as a transgene into the NOD mouse, diabetes is also prevented. In addition to these class II molecules determining diabetes susceptibility, more than 15 other genetic loci contribute to disease, each with a relatively small contribution, each neither necessary nor sufficient. Thus, inheritance of diabetes in the NOD mouse is polygenic. One manner in which the NOD mouse differs from humans is that more female than male NOD mice develop diabetes.

NOD mice, like humans, produce anti-insulin autoantibodies before the development of diabetes. Autoantibodies usually appear between 6 and 8 weeks of age, and diabetes usually develops after 16 weeks of age. Studies of islet beta cell mass indicate islet beta cell destruction and beta cell regeneration months before the onset of diabetes, although there is convincing evidence of an acceleration of beta cell destruction at disease onset. T cells and not autoantibodies mediate islet beta cell destruction, with clones of T cells reacting with several antigens able to transfer disease. A large number of T-cell clones reacting with insulin and reacting with unknown antigens have been characterized. There is debate about whether any given autoantigen is primary.

Diabetes can be prevented in the NOD mouse with more than 100 different therapies. Most, but not all, of these therapies target the immune system, and a number of these therapies are now in clinical trials in humans. Consistent with diabetes being mediated by T lymphocytes, immunosuppression or genes that block T-cell function prevent disease. Administration of high doses of nicotinamide delays the development of diabetes in NOD mice. Some of the most interesting therapies utilize autoantigens as “vaccines,” and in particular both glutamic acid decarboxylase (GAD) and insulin, when administered to the mice, prevented diabetes. The insulin molecule does not have to be metabolically active, and a dominant insulin peptide, insulin peptide B:9-23, given as a single subcutaneous injection prevented diabetes in 90% of susceptible NOD mice. It is thought that vaccination prevents diabetes by generating T lymphocytes (e.g., T<sub>n2</sub> type, transforming growth factor producing) that target an islet molecule (e.g., insulin) but that produce protective cytokines (e.g., interleukin-4, interleukin-10, transforming growth factor) when they home to the islets.
Oligogenic Animal Models

BB Rat

The BB (biobreeding) rat was the first intensively studied animal model of type 1A diabetes. The diabetes in this model differs from human diabetes in that diabetes-prone BB rats have an autosomal recessive mutation that produces a severe T-cell lymphopenia.\(^4\) One can induce diabetes in a related strain of rat, termed BB diabetes resistant (BB-DR), by administration of a monoclonal antibody that depletes T lymphocytes. As in humans and the NOD mouse, the disease depends upon specific class II alleles (similar to human HLA-DR and HLA-DQ) of the histocompatibility complex, in particular RT1-U. Of note, diabetes can be induced to develop in a series of rat strains with RT1-U (see later). Additional genes segregate to create diabetes susceptibility, but the number of genes is much less than for NOD mice.\(^4\)\(^5\)\(^6\) Prevention of diabetes in BB rats is more difficult than in NOD mice, which may be related to the severe T-cell lymphopenia. For example, insulin administration to BB rats prevented both diabetes and insulitis, but, in contrast to NOD mice, metabolically active insulin and insulin doses that induce hypoglycemia were usually required for prevention.\(^4\)\(^3\)

Long-Evans Tokushima Lean (LETL) Rat

Like BB rats, this strain has the RT-1U alleles and has an oligogenic inheritance of diabetes.\(^6\)\(^7\)
Induced Models of Type 1A Diabetes

Diabetes or insulitis can be induced in several strains of animals with drugs that induce islet destruction and broadly activate immune responses or with specific islet antigens. The drug streptozotocin is directly toxic to islet beta cells. In high doses, it rapidly induces diabetes. In low doses, a more chronic development of diabetes occurs that is likely to have some immunologic derivation. Surprisingly, administration of copolymer of polyinosinic and polycytidylic acids (poly-IC), a simple polynucleotide that activates interferon production when administered to a number of rat strains with the diabetes-susceptible RT1-U alleles, induced insulitis and diabetes. This suggests that many animals are susceptible to diabetes or insulitis given a strong immunologic stimulus. A ubiquitous heat shock protein has been administered to produce a transient form of diabetes in mice. Peptides of this heat shock protein are in clinical trials as a diabetes vaccine.
HISTOPATHOLOGY OF TYPE 1A DIABETES

As in animal models, type 1A diabetes of humans is characterized by selective destruction of the beta cells within islets. The noninsulin-producing cells of the islets remain in patients with long-standing type 1 diabetes, and these remaining islets lacking insulitis and beta cells are termed pseudoatrophic. A remarkable feature of the pancreas of patients with new-onset diabetes is heterogeneity of islet lesions. Within the same section of pancreas, a normal islet with no infiltrate may coexist with an islet containing beta cells with intense infiltration and a pseudoatrophic islet that has no infiltrate. This spottiness of the pathologic process is reminiscent of the destruction of areas of the skin in patients with vitiligo, in which melanocytes are destroyed in patches. Such heterogeneity of lesions may underlie the chronic development of type 1A diabetes in humans.

Islets of patients with type 1A diabetes overexpress class I HLA antigens, relatively rarely express class II HLA molecules on beta cells, express interferon, and up-regulate Fas molecules on all islet cells. The hypothesis that class II HLA expression contributes directly to beta cell autoimmunity is controversial. There is evidence that such expression in animal models does not activate autoimmunity, and it is questioned whether insulin- and class I positive islet cells are not macrophages that have ingested dead beta cells. Antigen presentation requires costimulatory molecules in addition to class II molecules, and beta cells do not express these costimulatory molecules. The specific manner by which the immune system destroys beta cells is not known, and molecules such as Fas may be important because T cells expressing Fas ligand may induce apoptosis of beta cells. Cytokines and CD8 cytotoxic lymphocytes are also likely to contribute to beta cell destruction.

Searches for viral particles and viral ribonucleic acid (RNA) within islets of patients with new-onset diabetes have been unrewarding but newer technologies and concepts should facilitate additional studies. In contrast to the islets of patients with new-onset diabetes, the pancreas of identical twin donors has been described as normal and the pancreas from patients with long-standing diabetes is composed of pseudoatrophic islets without markers of immune activation. A subset of patients with diabetes of several years’ duration still have insulitis, however.
GENETICS OF TYPE 1A DIABETES

It has long been recognized that diabetes is a heterogeneous group of disorders. It is also becoming apparent that type 1A diabetes is heterogeneous. There are probably many genetic forms of type 1A diabetes, with most forms influenced by HLA class II molecules. This group of disorders is likely to be linked by the presence of immunologic abnormalities that foster loss of tolerance to self-antigens. Individuals with specific HLA class I and class II molecules with immune dysfunction are susceptible to "target" islet autoantigens. It is noteworthy that many of the genes underlying diabetes susceptibility are similar in diverse countries, with specific alleles of those genes differing in their frequency. Several monogenic forms of type 1A diabetes can now be identified. It is not clear whether these genetically characterized forms of diabetes should now be included in the group of "Other Defined Causes of Diabetes." For the great majority of patients with type 1A diabetes, most of the genes causing diabetes susceptibility remain to be identified.

Monogenic Forms of Type 1A Diabetes

Autoimmune Polyclinocrine Syndrome Type I (AIRE Gene)

The autoimmune polyendocrine syndrome type I (APS-I) is rare, with an increased incidence in Finland, Sardinia, and among Iranian Jews, but has a worldwide occurrence. The disorders of the syndrome such as type 1 diabetes, mucocutaneous candidiasis, hypoparathyroidism, Addison's disease, and hepatitis (see Chapter 37 for more detailed discussion) identify a unique syndrome, and patients with this group of disorders almost always have mutations of the AIRE (autoimmune regulator) gene on chromosome 21. This gene apparently encodes a deoxyribonucleic acid (DNA) binding protein. The function of the gene is unknown, but its expression in lymphoid tissue and the clinical syndrome suggests an essential role in maintaining self-tolerance. There is considerable variability in the diseases expressed even for siblings with the same mutation. Some of this variability is likely to be influenced by genetic loci other than the AIRE gene. One example is the observation that although 18% of patients with APS-I develop type 1 diabetes, those with the common diabetes-protective HLA allele DQB1*0602 appear to have some protection from diabetes although not from Addison's disease.

X-Linked Polyendocrinopathy, Immune Dysfunction, and Diarrhea (Scurfy Gene)

The syndrome of X-linked polyendocrinopathy, immune dysfunction, and diarrhea (XPID) is associated with overwhelming neonatal autoimmunity, with most children dying in the first few days of life or as infants. In this syndrome lymphocytes invade multiple organs. It is associated with insulitis and beta cell destruction as well as lymphocytic intestinal inflammation with flattened villi and severe malabsorption. It is inherited as an X-linked recessive disease affecting only males, with a frequent clinical history of lack of male births. The disease apparently results from mutations of the scurfy gene, whose function is currently unknown but which, like the AIRE gene of APS-I, is a transcription factor. 

### TABLE 30-2 -- Risk of Type 1A Diabetes

<table>
<thead>
<tr>
<th>Proband with Diabetes</th>
<th>% Childhood Diabetes Mellitus (incidence/yr)</th>
<th>Islet Autoantibody</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>General population (United States)</td>
<td>0.3% (1525/100,000)</td>
<td>3% single Ab</td>
<td>Japanese incidence 1/100,000</td>
</tr>
<tr>
<td>Offspring</td>
<td>1%</td>
<td>0.3% multiple Abs</td>
<td>Incidence increasing in United States as in many European countries (e.g., in Colorado now 25/1,000,000)</td>
</tr>
<tr>
<td>Sibling</td>
<td>3.2%, 6% lifetime</td>
<td>7.4%</td>
<td></td>
</tr>
<tr>
<td>Dizygotic twin</td>
<td>6%</td>
<td>10%</td>
<td></td>
</tr>
<tr>
<td>Mother</td>
<td>2%</td>
<td>5%</td>
<td>Lower risk than offspring of father with diabetes mellitus</td>
</tr>
<tr>
<td>Father</td>
<td>4.6%</td>
<td>6.5%</td>
<td></td>
</tr>
<tr>
<td>Father and mother</td>
<td>10% ?</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>Monozygotic twin</td>
<td>50%</td>
<td>50%</td>
<td>MZT in Japan, 40% risk of diabetes</td>
</tr>
<tr>
<td>Ab, antibody; MZT, monozygotic twin.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Ab, antibody; MZT, monozygotic twin.
Autoantibody

an asterisk (*), followed by two
two molecules are extremely polymorphic in amino acid sequence. Each polymorphic variant of each chain is designated with a gene locus name (e.g., DRB1) followed by

HLA molecules function to present peptides to T lymphocytes. Each molecule is made up of two chains, with each chain encoded by a separate gene. These
terms and a description of the basis for classification it is comprehensible.

The Major Histocompatibility Complex

celiac disease

Associated Autoimmune Disorders

Idiopathic Type 1A Diabetes

Descriptive Genetics

In the United States, the risk of childhood diabetes is approximately 1 in 300. [5] This is 15-fold less than the diabetes risk for a first-degree relative of a patient with type 1 diabetes (Table 30-2). [6] It is 150-fold less than the risk for a monozygotic twin of a patient with type 1 diabetes. [7] Although the population risk of type 1 diabetes in Japan is 15-fold less than in the United States, the risk for an identical twin in Japan is similar to that for an identical twin in the United States. [8] This suggests that when genetic susceptibility is present, either in Japan or in the United States, the diabetes risk is extremely high. Although the risk of diabetes is much greater for relatives of patients with type 1A diabetes, it is important to realize that most (>85%) individuals in whom type 1A diabetes develops do not have a first-degree relative with the disease. The frequency of sporadic cases results in part from the fact that almost 40% of individuals in the general population carry high-risk HLA alleles for type 1A diabetes (see "The Major Histocompatibility Complex").

The highest known incidence of type 1A diabetes is found in Finland and Sardinia. Finland now has an annual incidence approaching 50 per 100,000 children. Over the past four decades the incidence has increased almost threefold, suggesting a dramatic environmental change (either an increase of causative factors or a decrease of protective factors).

Twin Studies

Twin studies of diabetes have an impressive pedigree. The study of monozygotic twins of patients with diabetes by Pyke and co-workers [9] contributed to the recognition of distinct forms of diabetes, initially termed adult-onset and juvenileonset, subsequently termed insulin-dependent and noninsulin-dependent, and now termed type 1 and type 2 diabetes. [10] The concordance rate for monozygotic and dizygotic twins

provides important information regarding genetic factors contributing to a disease because monozygotic twins share all germ line-inherited polymorphisms or mutations whereas dizygotic twins are similar to siblings of patients with a disease and have only one half of genes in common. For a locus that contributes to disease in a recessive manner, only one fourth of dizygotic twins would be homozygous to a sibling with diabetes at that locus but all monozygotic twins would be homozygous for all recessive loci of their diabetic twin mate. Although overall concordance rates of monozygotic twins for type 1 diabetes are calculated, it is likely that type 1 diabetes is heterogeneous and that groups of monozygotic twins may have different genetic etiologies. With such genetic heterogeneity, one would expect different concordance rates for different genetic diseases.

Redondo and co-workers [11] have analyzed prospective follow-up data from a large series of initially discordant monozygotic twins from Great Britain combined with a series from the United States. Progression to diabetes was identical for both series of twins. Of note, there was no length of time of discordance beyond which a monozygotic twin mate did not have a risk of type 1 diabetes. Nevertheless, the hazard rate for development of diabetes decreased as the period of discordance increased. There was also a marked variation in the risk of diabetes relative to the age at which diabetes developed in the index twin. The overall rate of concordance for monozygotic twins was 50%. However, if type 1 diabetes developed in the index twin after age 25, the concordance rate by life table analysis was less than 10% (Fig. 30-1). If diabetes developed in the index twin prior to age 5, the concordance rate was 70% by 40 years of follow-up. This analysis of monozygotic twins suggests genetic heterogeneity but also confirms that a significant subset of monozygotic twins do not progress to diabetes. This suggests that either environmental factors, random factors, or nongerm line-inherited variations (e.g., imprinting, T-cell receptor polymorphisms, somatic mutations) contribute to diabetes risk.

An important unanswered question (given the limited number and size of studies) is whether dizygotic twins of patients with type 1A diabetes have a diabetes risk greater than that of siblings. If the risk is identical, it suggests that environmental factors whose presence is time-dependent (e.g., uncommon infections) may have little influence on the development of diabetes. Dizygotic twins differ from siblings in terms of a greater commonality of environment over time (e.g., common pregnancy). Studies of dizygotic twins suggest that their risk of diabetes may not differ from that of siblings or at most is increased by a factor of 2 compared with the 10-fold increase for monozygotic twins.

Genetic factors influence not only the development of diabetes but also the expression of anti-islet autoantibodies. For identical twins the expression of anti-islet autoantibodies is tightly linked to the eventual progression to overt diabetes, and monozygotic twins have a high prevalence of expression of autoantibodies. Dizygotic twins much less often express anti-islet autoantibodies, and the prevalence is similar to that of siblings. [12]

Associated Autoimmune Disorders

Because type 1A diabetes is an immune-mediated illness that develops in a genetically susceptible individual, it is not surprising that most patients with type 1A have one or more additional autoimmune diseases. The most common associated disorders are thyroid autoimmunity (Graves’ disease or Hashimoto’s thyroiditis) and celiac disease (Table 30-3).

The Major Histocompatibility Complex

The most important loci determining the risk of type 1 diabetes are within the MHC on chromosome 6p21 (Fig. 30-2), in particular HLA class II molecules (DR, DQ, and DP). [13] In addition, standard class I loci (HLA A, B, and C) influence disease, and it is likely that additional loci within the MHC that influence immune function contribute to diabetes risk. [14] Figure 30-2 illustrates the MHC. The nomenclature for alleles of this region is somewhat daunting, but with definitions of several terms and a description of the basis for classification it is comprehensible.

HLA molecules function to present peptides to T lymphocytes. Each molecule is made up of two chains, each chain encoded by a separate gene. These molecules are extremely polymorphic in amino acid sequence. Each polymorphic variant of each chain is designated with a gene locus name (e.g., DRB1) followed by an asterisk (*), followed by two

<table>
<thead>
<tr>
<th>Disease</th>
<th>Autoantibody</th>
<th>Disease Prevalence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thyroiditis or Graves’ disease</td>
<td>25% (peroxidase or thyroglobulin)</td>
<td>4</td>
</tr>
</tbody>
</table>
Celiac disease 12% (transglutaminase) 6
Addison's disease 1.5% (21-hydroxylase) 0.5
Pernicious anemia 21% (parietal cell) 2.6

Digits referring to the serologic specificity (from the time when typing was performed with antibodies), followed by two digits for the specific allele (now determined with DNA-based typing), followed by a single digit to distinguish silent nucleotide polymorphisms (nucleotide differences that do not change amino acid sequence). For example, the designated allele DRB1*0405 has DR4 serologic specificity and is associated with high diabetes risk. For DR alleles, one usually specifies only the DRB chain because the DNA chain is not polymorphic. For the class II molecules (A, B, and C) one also specifies only a single chain because the other chain, \( \alpha \), -microglobulin is minimally polymorphic. There are more than 240 different known alleles of DRB1. Each individual inherits two DRB1 alleles, one from each parent.

Because HLA gene loci are in close proximity to each other on the sixth chromosome, one usually inherits a group of alleles as a unit, and this is termed a haplotype. For example, the alleles A*0101, B*0801, DRB1*0301, and DQA1*0501, DQB1*0201 constitute a common haplotype associated with diabetes risk. When specific alleles of different genes are nonrandomly associated with each other on a haplotype (such as the preceding alleles A1, B8, and DR3), the alleles are said to be in linkage disequilibrium.

Two MHC haplotypes, one inherited from each parent, constitute the MHC genotype. This genotype ultimately determines the MHC-encoded risk for type 1A diabetes. For DQ molecules, both of the chains (DQA and DQB) are polymorphic. This adds an important level of diversity in that the protein chains encoded by the alleles of one haplotype may combine with the chains encoded by the other haplotype. For example, individuals with the highest risk genotype DRB1*0301, DQA1*0501, DQB1*0201 and DRB1*0405, DQA1*0301, DQB1*0302 can produce four different DQ molecules: DQA1*0501, DQB1*0201 and DQA1*0301, DQB1*0302 as expected but also DQA1*0501, DQB1*0302 and DQA1*0301, DQB1*0201. The DQ molecule DQA1*0501, DQB1*0201 is also called DQ2 and DQA1*0301, DQB1*0302 is called DQ8. A common DQ molecule, DQA1*0102, DQB1*0602, provides dominant protection from type 1 diabetes and is termed DQ2.

The major determinants of diabetes susceptibility are DR and DQ molecules, and specific alleles of both DR and DQ can either increase or decrease the risk of diabetes. Table 30-4 summarizes the diabetes risk associated with a number of DR and DQ haplotypes.

In a number of studies, children at birth, either from the general population or relatives of patients with type 1 diabetes, have been HLA typed. The typing is relatively straightforward and either is based on direct DNA sequencing of polymerase chain reaction amplified DNA fragments or utilizes DNA probes that hybridize specifically to different allelic sequences. In Denver, Colorado, 2.4% of newborns have the highest risk DR-DQ genotype for type 1A diabetes, namely DR3-DQ2 with DR4-DQ8 (DR3/4 DQ8/2 heterozygotes). Fifty percent of children younger than 10 years and approximately 30% of older children who develop diabetes have this highest risk genotype. One can estimate that approximately 1 of 16 children with the highest risk HLA genotype from the general population progress to diabetes (versus a population risk of 1 per 300). Alternatively, 15 of 16 children from the general population who are DQB1/DQ2 heterozygotes do not develop diabetes.

Ninety-five percent of individuals who develop diabetes who have either DR3-DQ2 or DR4-DQ8, do as approximately 40% of the general population. The protective haplotype DRB1*1501, DQA1*0102, DQB1*0602 is present in 20% of the general population and less than 3% of patients with type 1A diabetes. A DR allele, DRB1*1401, also appears to provide dominant protection. There are additional high-risk haplotypes that are not common, such as DQA1*0401, DQB1*0402. It was proposed as a simple rule that the presence of aspartic acid at position 57 of the DQ chain and arginine at DQ 52 is associated with diabetes risk. As illustrated before, there are many exceptions to this rule, and knowledge of the complete sequences (allele) rather than dependence on this rule is essential.

**Insulin Locus**

Almost two decades ago, Bell and colleagues discovered that variations in the number of nucleotide repeat elements 5 of the insulin gene were associated with the development of type 1A diabetes. The longest group of repeats was associated with decreased diabetes risk. These studies have been replicated, and the locus of importance is clearly limited to the insulin gene. Of note, the protective insulin gene polymorphism is associated with greater insulin messenger RNA expression within the thymus. Hannan advanced the hypothesis that within lymphoid organs there are peripheral

**Table 30-4 – Diabetes Risk of Representative DR and DQ Haplotypes**

<table>
<thead>
<tr>
<th>DRB1</th>
<th>DQA1</th>
<th>DQB1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0401 or 0403, or 0405</td>
<td>0301</td>
<td>0302 (DQ8)</td>
</tr>
<tr>
<td>0301</td>
<td>0501</td>
<td>0201 (DQ2)</td>
</tr>
<tr>
<td><strong>Moderate Risk</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0801</td>
<td>0401</td>
<td>0402</td>
</tr>
<tr>
<td>0404</td>
<td>0301</td>
<td>0302</td>
</tr>
<tr>
<td>0101</td>
<td>0101</td>
<td>0501</td>
</tr>
<tr>
<td>0901</td>
<td>0301</td>
<td>0303</td>
</tr>
<tr>
<td><strong>Moderate Protection</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0403</td>
<td>0301</td>
<td>0302</td>
</tr>
<tr>
<td>0701</td>
<td>0201</td>
<td>0201</td>
</tr>
<tr>
<td>1101</td>
<td>0501</td>
<td>0301</td>
</tr>
<tr>
<td><strong>Strong Protection</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1501</td>
<td></td>
<td>0102 (DQ6)</td>
</tr>
<tr>
<td>1401</td>
<td>0101</td>
<td>0503</td>
</tr>
<tr>
<td>0701</td>
<td>0201</td>
<td>0303</td>
</tr>
</tbody>
</table>
There is an international effort to define additional genes that contribute to the development of type 1A diabetes. The genes associated with the two monogenic forms of polyendocrinopathy (AIRE gene and XPID gene), HLA genes, and the insulin gene are to date the only clearly identified genes. Much of the effort in searching for relevant genes has analyzed the common inheritance of genetic regions for pairs of diabetic siblings. With such an analysis, a long list of genes for type 1 diabetes, termed iddm loci, have been proposed, each with a specific number (e.g., iddm1 is the MHC, iddm2 the insulin gene, and iddm15 a locus on chromosome 6q).

Such studies are likely to generate at least as many false-positive loci as true loci, and the difficulty in replicating findings attests to the problems. Polymorphisms of the CTLA4 gene (iddm12) contribute to Graves’ disease and apparently to diabetes in some but not all populations.

One locus for which a consensus appears to be developing that it is “real” is the iddm15 locus. This locus is, however, on 6q21 and is loosely linked (30 cM) to the MHC, and it is important to rule out an influence of class I HLA alleles in accounting for iddm15. One locus, iddm17, was identified not through analysis of sibling pairs but by intensive study of a single family with 21 members with type 1A diabetes. It appears that this locus, in combination with high-risk HLA alleles within the initial family studied, creates a risk of diabetes approaching 40%, and there is preliminary evidence that the same locus may influence diabetes risk in a subset of families in the United States. As reflected by this single family, it is not yet clear whether type 1A diabetes is of polygenic origin with small effects of many genes determining risk or oligogenic but heterogeneous. If it is oligogenic, a few (e.g., less than three) genes may determine risk in each family.

In summary, many putative diabetes loci have been identified outside the HLA region, but until actual genes contributing to diabetes risk are identified, these loci are unlikely to contribute to genetic counseling or clinical care. HLA typing is being utilized to define the risk of diabetes at birth. The defined risk can be extremely high, depending on the relationship to a proband with diabetes and the specific HLA genotype. For example, siblings of patients with type 1A diabetes with the DR3-DQ2/DR4-DQ8 genotype appear to have a diabetes risk that exceeds 50%. In contrast, children from the general population with the same HLA DR and DQ genotype have a risk of less than 6%. In addition to influencing diabetes risk, specific HLA genotypes contribute to risk for associated autoimmune disorders. It is remarkable that one third of DR3-DQ2 homozygous patients with type 1A diabetes express transglutaminase autoantibodies and half of them have celiac disease on biopsy.
ENVIRONMENTAL FACTORS

Despite more than three decades of research, there is only one environmental factor clearly associated with type 1A diabetes, namely congenital rubella infection (see later). The association of only one factor is probably related to the long prodromal phase that precedes type 1A diabetes, which makes the discovery of relevant environmental factors particularly difficult. A number of factors can induce type 1A diabetes in animal models, one of the most interesting being the Kilham rat virus infection of BB-DR rats. In this model, which lacks the lymphopenia of the related BB strain, diabetes does not develop unless the animals are infected with Kilham virus (a parvovirus) or injected with poly-IC. Poly-IC mimics double-stranded RNA and induces high levels of interferon. The Kilham virus apparently does not directly infect islets and is thought to be an immune activator similar to poly-IC. Diabetes induction is dependent on specific class II alleles termed RT1-U, and a number of animal strains with RT1-U are susceptible to diabetes. If these animal models are relevant to humans, it may be that many environmental stimuli in a genetically susceptible host activate autoimmunity.

In countries throughout the world the incidence of type 1A diabetes is increasing, particularly for children in whom the disease develops before age 5. This is strong evidence that environmental factors related to diabetes risk have changed over the past four decades. Factors that increase diabetes risk may be increasing or, just as likely, factors that suppress the development of diabetes may be decreasing. For example, in animal models of type 1A diabetes (NOD mouse and BB rat), infections with common viruses usually decrease the development of diabetes.

Infections

Congenital rubella, but not noncongenital infection, greatly increases the development of type 1A diabetes. The children with diabetes usually have high-risk HLA alleles and these children frequently have thyroid autoimmunity. The manner by which this congenital infection increases diabetes development is currently unknown. Hypotheses have ranged from potential molecular mimicry to long-term alteration in T-cell function secondary to the congenital insult.

Enteroviruses are small RNA viruses that frequently infect young children. Initial anecdotal reports that coxsackievirus infections may cause diabetes evaluated children with severe infections with death at diabetes onset. These studies preceded the realization that type 1A diabetes is not an acute disease, and it is likely that viral infection at the onset of type 1 diabetes is most often incidental. At the time of presentation with diabetes, almost all children have elevated HbA1c, reflecting probably months of hyperglycemia preceding diagnosis. A description from Japan of individuals with acute-onset diabetes with normal HbA1c, elevated amylase, and infiltrates within the exocrine but not endocrine pancreas appears to represent a form of type 1B diabetes.

The potential importance of enteroviral infection has been emphasized by studies from Scandinavia in which enteroviral infection was evaluated during pregnancy and in infancy. Infection is usually detected by either changes in antiviral antibodies or detection of enteroviral RNA by molecular techniques. Although some studies have reported increased enteroviral infection during pregnancy in mothers whose children have developed diabetes, others have not. As infants with a genetic risk for the development of diabetes are followed from birth, it becomes possible to analyze prospectively the expression of enterovirus RNA. Studies by Hultyn and colleagues found that enteroviral infection is associated with the appearance of anti-islet autoantibodies. Similar studies from Denver, Colorado, did not find an association. The major difference between the two studies appears to be the lower rate of enteroviral infection in Finnish control subjects compared with Colorado control subjects with rates of infection similar to those in cases.

Other viruses are being evaluated for association with the triggering of autoimmunity. One study from Australia found an association with rotavirus infection. Rotavirus infection is common in young children. The Australian study did not find an increase in rotavirus infection compared with that in control subjects but reported an association of rotavirus infection with increases of anti-islet autoantibodies. Studies from Denver did not indicate an increase in rotavirus infection in infants developing autoantibodies.
Vaccination

It has been claimed that the timing of routine childhood vaccinations influences the development of type 1A diabetes. This is an important health concern if parents alter their family's childhood vaccination because of concern about development of diabetes. A series of studies have been carried out and do not provide evidence that childhood vaccinations influence the development of diabetes.
Diet

A disease such as celiac disease is critically dependent on the ingestion of a specific food, namely the wheat protein gliadin. In addition, a number of dietary modifications altered the development of diabetes in NOD mice and BB rats. Investigators have championed the hypothesis that early introduction of bovine milk increases the development of diabetes. This hypothesis is primarily based on retrospective studies associating early or increased bovine milk ingestion (or less breast-feeding) with an increased risk of type 1A diabetes. Several prospective studies in which infants are observed until the development of anti-islet autoantibodies have failed to find an association or have found a weak association with either breast-feeding or bovine milk ingestion. Pilot studies of an infant formula lacking bovine milk proteins have been initiated in Finland. Preliminary data suggest that such a restricted diet may produce a small decrease of cytoplasmic islet cell autoantibodies but not of GAD65 autoantibodies.

Figure 30-3 (Figure Not Available) Hypothetical stages in the development of type 1A diabetes beginning with genetic susceptibility and ending with complete beta cell destruction. (Modified from Eisenbarth GS. Type 1 diabetes mellitus: A chronic autoimmune disorder. N Engl J Med 1986, with modifications by Jay Skyler, University of Miami.)
NATURAL HISTORY OF TYPE 1A DIABETES

We typically divide the development of type 1A diabetes into a series of stages beginning with genetic susceptibility and ending with essentially complete beta cell destruction (Fig. 30-3). It is, however, likely that both genes and environmental factors influence the course of development of type 1A diabetes during the complete prediabetic period. For instance, injection of immunostimulants such as Freund's adjuvant can prevent progression to diabetes in animals with insulitis. Mathis and co-workers have proposed the existence of "checkpoints" in the development of diabetes, and such checkpoints may have a strong genetic component. As discussed subsequently, type 1A diabetes is quite predictable given specific immunologic, genetic, and metabolic characteristics, and it is such characteristics that set the stage for preventive trials.

Genetic and Immunologic Heterogeneity by Age of Onset

Type 1A diabetes can develop at any age, from the neonatal period to the sixth decade of life. In that identical twins can become concordant 30 years after their twin mate, not all age heterogeneity can be ascribed to different genetic syndromes. Nevertheless, there is an overall correlation between the age at which diabetes develops in one twin or sibling and the age of development of diabetes in his or her relative. As discussed earlier, children in whom type 1A diabetes develops at an early age more often are DR3/4, DQ8/2 heterozygotes. In addition, there is evidence that class I HLA alleles (or other nonclass II genes with the HLA region) may influence the age of diabetes onset (e.g., the A24 allele). At the other end of the age spectrum, there is evidence that the protective HLA allele DQA1*0102, DQB1*0602 is not as protective for young adults as it is for children.

The most characteristic difference related to the age of diabetes onset is the presence of higher levels of insulin autoantibodies in children who develop the disease at an early age (e.g., younger than 5 years). The high levels and frequent positivity of insulin autoantibodies make measurement of IAA (insulin autoantibodies) the best single marker for diabetes development in young children. For children in whom autoantibodies arise in the first 3 years of life, insulin autoantibodies often appear first. In contrast, GAD65 autoantibodies are more often positive in adults developing type 1A diabetes. The correlation of levels of insulin autoantibodies and age of diabetes onset may be related to children with higher levels progressing more rapidly to diabetes. Such rapid progression, however, occurs only if insulin autoantibodies are present with another islet autoantibody (see "Combinatorial Autoantibody Prediction").
Combinatorial Autoantibody Prediction

The most specific anti-islet autoantibody assays are usually set with cutoffs above the 99th percentile of control populations. Thus, with three major autoantibody assays (GAD65, ICA512 [islet cell antibody], and insulin) one would predict that approximately 3 of 100 normal individuals would express one or more of the three autoantibodies. Because approximately 3 of 1000 children develop type 1A diabetes, this suggests that in the great majority of antibody-positive individuals, diabetes never develops or may develop late in life.

A relatively low positive predictive value for single antibodies may be due to methodologic limitations or the presence of autoantibodies identical to those of prediabetics but found in some individuals who do not progress to diabetes. It is likely that both occur. For example, a low positive (but > 99th percentile) autoantibody result of a control subject is often not confirmed on repeated testing. Autoantibodies of prediabetic individuals usually react with multiple epitopes of the ICA512 molecule, whereas false-positive autoantibodies frequently react with only one or no clearly defined epitope of the molecule, suggesting that false-positive and diabetes-associated anti-ICA512 autoantibodies differ. There are, however, individuals, usually adult relatives of patients with type 1A diabetes (often with DQB1*0602), with extremely high levels of GAD65 autoantibodies that react with multiple GAD epitopes with no evidence of progression to diabetes. 142

Assessment of the significance of an autoantibody result (as for any diagnostic test) is improved by taking into account the prior probability of disease. A patient with overt diabetes and expression of a single anti-islet autoantibody has a high probability of type 1A diabetes. An individual from the general population or even a relative expressing a single autoantibody (and remaining with a single autoantibody) has a much lower risk of progressing to type 1A diabetes. 142

Usually, in attempting to improve the specificity of a test, one sacrifices sensitivity. For prediction of type 1A diabetes, because three biochemical autoantibodies are measurable, one can combine the tests with the observation that the presence of two or more autoantibodies is associated with a very high risk of diabetes. 143 144

Approximately 1 of 350 individuals from the general population express two or more of the GAD65, ICA512, or insulin autoantibodies, which approaches population estimates of type 1A diabetes. Among first-degree relatives of patients with type 1A diabetes, two or more autoantibodies indicate a risk over 10 years of more than 90%, whereas a single autoantibody is associated with a risk of less than 20% over 10 years. 143
Metabolic Progression before Hyperglycemia

The intravenous glucose tolerance test aids in evaluating the time to onset of diabetes among individuals expressing anti-islet autoantibodies. Most commonly, glucose is given at 0.5 g/kg over 5 minutes (maximum 35 g, 25 g/dL) and insulin levels are measured before and 1 and 3 minutes after the glucose infusion. Most individuals within a year of overt diabetes have no first-phase insulin secretion after intravenous glucose. The diagnosis of type 1A diabetes usually relies upon the presence of fasting hyperglycemia, but with prospective evaluation many individuals have diabetes by the 120-minute criteria on oral glucose tolerance with nondiagnostic fasting glucose.
C-Peptide Loss after Hyperglycemia

Following the diagnosis of diabetes, levels of C peptide can be utilized to assess remaining beta cell function. C-peptide levels are usually measured in the fasting state or after intravenous glucagon or with a standard meal (e.g., Sustacal). Such measurements are primarily of importance for trials of therapies to alleviate loss of insulin secretion after diagnosis. Determination of C peptide provides the best current measure for assessing the impact of new therapies. As shown in the Diabetes Control and Complications Trial (DCCT), a small amount of remaining C peptide is associated with impressive metabolic benefit.
Type 1A Diabetes with Pregnancy

Approximately 5% of women with gestational diabetes (diabetes diagnosed during pregnancy) have an early form of type 1A diabetes that is discovered during pregnancy. These women express anti-islet autoantibodies and progress to overt diabetes more rapidly after pregnancy.
Latent Autoimmune Diabetes of Adults

Type 1A diabetes can occur at any age. Depending on the population, between 5% and 15% of individuals with what appears to be type 2 diabetes express anti-islet autoantibodies. Multiple studies have demonstrated that such individuals progress relatively rapidly (within 3 years) to insulin-requiring diabetes. The HLA alleles of such individuals reflect that of type 1A diabetes.
Transient Hyperglycemia

A significant number of children are evaluated by endocrinologists for transient hyperglycemia. The usual history is of severe stress associated with hyperglycemia that resolves within days to a month. Such children may be in the "honeymoon" phase of type 1A diabetes or may truly have a transient episode of hyperglycemia. Rarely, diabetes in children is misdiagnosed (e.g., the authors have seen a child with normal HbA1c values for several years who stopped insulin and was subsequently found to have renal glucosuria and not diabetes). Children without severe stress with transient hyperglycemia or with a relative with type 1A diabetes are more likely to have early type 1A diabetes. Absence of anti-islet autoantibodies and a normal intravenous glucose tolerance test are strongly indicative of transient hyperglycemia and not type 1A diabetes. It is not known whether children with transient hyperglycemia are at increased risk for type 2 diabetes later in life.
IMMUNOTHERAPY OF TYPE 1A DIABETES

At the onset of type 1A diabetes, a major clinical research goal is the prevention of further beta cell destruction. At present there is no proven safe and effective therapy to prevent such further destruction or to prevent the development of type 1A diabetes in those at risk (e.g., genetically at risk individuals with anti-islet autoantibodies). A number of clinical trials have been completed and a large number of trials are under way or about to be initiated.

Immunosuppression

The earliest studies of therapies to prevent beta cell destruction utilized immunosuppressive agents. Large trials of cyclosporine indicated that while administered it prevented further loss of C-peptide secretion and improved metabolic function. It did not, however, maintain a nondiabetic state when therapy was instituted after the onset of diabetes, and with discontinuation of the drug individuals rapidly lost C-peptide reserve. The combination of inability to “cure” diabetes and toxicities associated with cyclosporine (in particular nephrotoxicity and concern about increased risk of malignancy) has ruled out its use. Other immunosuppressive agents such as prednisone or azathioprine had relatively little effect. A small study suggested that methotrexate, another common immunotherapeutic agent, is ineffective. Thus, at present, although type 1A diabetes is an immune-mediated disorder, it is not treated with common immunotherapeutic agents. A pilot study of a specific antibody to CD3 has been reported and further essential trials are underway.
Immunologic Vaccination

In animal models (especially the NOD mouse) it is relatively easy to prevent type 1 diabetes. Potentially the most exciting modalities utilize forms of immunologic vaccination. Much of the excitement derives from the specificity of the therapy and relatively low risk compared with immunosuppression and not from demonstrated efficacy in humans. The basic concept behind the bulk of such therapies is the induction of lymphocytes that target a given islet antigen and upon encountering their target antigen (e.g., insulin) produce cytokines that suppress autoimmunity and tissue destruction.

A general class of T lymphocytes termed Th2 T cells produce the cytokine interleukin-4 rather than interferon and interleukin-2 (Th1 T cells) and decrease cell-mediated immune destruction. Induction of a protective immune response may depend upon the route of administration of the given antigen (e.g., oral tolerance) or the utilization of an altered antigen (e.g., altered peptide ligands). For example, insulin given either orally or by subcutaneous injection prevented diabetes in NOD mice. Intact insulin is not necessary because insulin B chain and an immunodominant B:923 peptide of insulin were also effective. The latter molecules have no insulin-like metabolic effect but are able to activate T lymphocytes that target insulin.

The Diabetes Prevention Trial Type 1 is studying both oral insulin and parenteral injections of low doses of insulin. The results of the parenteral trial did not demonstrate a reduction in the risk of developing diabetes. The oral trial will continue for several more years (relatives of patients with type 1A diabetes can be screened for this trial by calling 1-800-HALT-DM1). A peptide of a heat shock protein (p277) is being studied, and trials of the GAD65 molecule are about to be initiated.
Other Therapies

A review by Atkinson and Leiter\textsuperscript{14} pointed out that more than 100 different interventions prevent diabetes in NOD mice. The relative ease of diabetes prevention in this animal model provides the basis for a number of trials initiated in humans. The largest such European trial (ENDIT, European Nicotinamide Trial) utilizes nicotinamide in gram doses. Nicotinamide is able to prevent diabetes induced by the drug streptozotocin and probably acts by preserving nicotinamide adenine dinucleotide levels in islet cells or blocking cytokine-induced destruction. There have been a number of studies of nicotinamide, including a randomized placebo-controlled trial in children and a small trial in at-risk relatives, that found no effect of nicotinamide.\textsuperscript{15,166} Other trials suggest some preservation in adult patients presenting with type 1A diabetes.\textsuperscript{167} The ENDIT trial should provide definitive information concerning the potential effect (or lack of effect) of nicotinamide.
IMMUNOLOGY OF ISLET-PANCREATIC TRANSPLANTATION

Pancreatic transplantation for patients requiring a kidney transplant is an accepted clinical procedure. Patients with a kidney transplant receive immunosuppressive drugs, and results for pancreatic transplantation in this setting have progressively improved. With a successful pancreas transplant, hyperglycemia is "immediately" reversed and there is some evidence of improved long-term outcomes. Nevertheless, the surgery is extensive and there are multiple potential complications associated with the transplant. Diabetes can recur because of either recurrent autoimmunity or more often allograft rejection. It is difficult to monitor specific islet destruction, and with the development of hyperglycemia it is usually not possible to restore euglycemia.

Up until studies from Edmonton the results of islet transplantation have been poor, with less than 10% of patients with type 1A diabetes becoming insulin independent at 1 year. In contrast, with autotransplants of patients with pancreatitis most patients become insulin independent and remain so. The Edmonton group has utilized meticulous islet isolation techniques, transplantation of islets from two pancreases, and an immunosuppressive regimen utilizing the new drug rapamycin. With this regimen more than a dozen patients have had successful transplants, and the longest duration after transplantation has been 18 months. With the Edmonton regimen, insulin antibodies decreased and there was no new development or increase of either GAD65 or ICA512 autoantibodies. With other regimens, marked increases of such autoantibodies appeared to herald loss of islet function.

The Edmonton protocol is now being tested in a series of specialized centers throughout North America and Europe, and the applicability of the technique should be rapidly evaluated. Even if the Edmonton results are reproduced at multiple centers, the number of islets available from cadaveric donors for transplantation is limited. Further research to allow xenogeneic transplantation or production of islets from stem cells is essential.
INSULIN AUTOIMMUNE SYNDROME

The insulin autoimmune syndrome, also termed Hirata syndrome, is rare and typically associated with hypoglycemia. These patients have extremely high concentrations of autoantibodies reacting with human insulin. It is thought that inappropriate (nonregulated) release of autoantibody-bound insulin produces the hypoglycemia. The disease occurs most commonly in Asian individuals. Among 50 Japanese patients with the syndrome and the typical polyclonal anti-insulin autoantibodies, 96% had a DRB1, DR4 allele and 84% (42 of 50) the DRB1*0406 allele. In contrast, patients with monoclonal anti-insulin autoantibodies do not have such a remarkable HLA association. Most patients develop the disease in association with treatment with sulfhydryl-containing medications, in particular methimazole. Treatment usually consists of stopping these medications, and for more than 75% the disease remits.
INSULIN ALLERGY

Mild forms of immune reactivity to insulin are not uncommon. Essentially all patients treated with human insulin produce anti-insulin autoantibodies measurable with sensitive fluid phase radioassays. The levels of these autoantibodies are relatively low and they do not appear to interfere with insulin therapy, although there are reports correlating insulin antibodies with macrosomia. With the introduction of recombinant human insulin replacing animal insulins, symptomatic immune responses to insulin such as immediate hypersensitivity, delayed hypersensitivity, lipoatrophy, and lipohypertrophy have decreased. Allergic reactions can occur with insulin analogues, although this is uncommon. More common perhaps are allergies to protamine used to complex insulin in neutral protamine Hagedorn (NPH) formulations as well as to the lubricants, preservatives, and plastics in bottles, stoppers, syringes, and needles. The usual therapy consists of switching the type or formulation of insulin followed by oral antihistamines for immunoglobulin E-mediated local reactions, followed by insulin desensitization or addition of small amounts of glucocorticoids to the insulin injected for local delayed hypersensitivity reactions.
ANTIINSULIN RECEPTOR AUTOANTIBODIES

Antiinsulin receptor autoantibodies (type B insulin resistance) are associated with both hypoglycemia and insulin resistance. It appears that antiinsulin receptor autoantibodies can act as either an antagonist or agonist. This syndrome is rare and is often associated with nonorgan specific autoimmunity.
CLINICAL PRESENTATION

The peak age of presentation of type 1 diabetes in children is around the age of puberty. The symptoms and signs are related to the presence of hyperglycemia and the resulting effects on fluid and electrolyte balance. They generally include polyuria, polydipsia, polyphagia, weight loss, and blurred vision. Because infection may have precipitated the initial presentation, symptoms of infection may also be present such as fever, sore throat, cough, or dysuria. In children in particular the onset of symptoms may occur over a brief period and families may be able to date their onset with considerable accuracy. Onset of symptoms may also be insidious, particularly in older persons with type 1 diabetes, and may occur over a time frame of weeks or even months.

If onset of type 1 diabetes is associated with ketoacidosis, which is not uncommon, additional symptoms related to this acute metabolic complication of diabetes are also present. These symptoms can include abdominal pain, nausea, and vomiting. Variable effects on mental status may be seen, ranging from slight drowsiness to profound lethargy and even coma if the condition has been untreated for a significant period of time.

Laboratory Findings

Plasma glucose concentrations at presentation are elevated, usually in the range 300 to 500 mg/dL. If the presentation is uncomplicated, the remainder of the fluid and electrolyte measurements may be completely normal. On the other hand, if diabetic ketoacidosis is present, the measurements reflect the presence of an acidosis as well as more severe dehydration. Thus, in diabetic ketoacidosis, the serum sodium value is frequently at the lower limit of normal or even mildly reduced, reflecting the osmotic effect of hyperglycemia and on occasion the presence of vomiting with continued water intake. A sodium value less than 120 mmol/L is usually associated with severe hypertriglyceridemia that can lead to spurious hyponatremia. Despite significant losses of potassium in the urine and total-body potassium deficits, the presence of acidosis usually leads to an elevated serum potassium concentration at the time of the initial presentation. Serum bicarbonate concentrations are usually less than 10 mg/dL, and elevations in serum concentrations of triglyceride and free fatty acids are found. Levels of ketone bodies are also elevated. Because dehydration is invariably present, this leads to increases in the concentrations of blood urea nitrogen and creatinine. In conjunction with the increase in serum glucose, the increases in blood urea nitrogen invariably increase the serum osmolality, often to greater than 300 mmol/kg.
TREATMENT OF TYPE 1 DIABETES

Importance of Tight Glucose Control

The overriding principle in the treatment of the majority of patients with type 1 diabetes is that a health care team that includes a physician, diabetes nurse educator, nutritionist, and other health care professionals as appropriate should work closely with the patient to achieve blood glucose concentrations as close to normal as possible because these are associated with a reduced risk of diabetic complications. Although studies in animal models and epidemiologic studies suggested that tighter glucose control was associated with better long-term outcomes for the diabetic patient in terms of a reduced risk of complications, the most definitive study in this regard has been the Diabetes Control and Complications Trial (DCCT) that was completed in 1993. This landmark study was performed in a total of 1441 patients with type 1 diabetes with no retinopathy at baseline (the primary prevention cohort) and 715 with mild retinopathy (the secondary intervention cohort) who were randomly assigned to intensive therapy or conventional therapy.

Intensive therapy consisted of insulin administration by an external pump or by three or more daily insulin injections. The dosage was adjusted according to the results of self-monitoring of the blood glucose performed at least four times per day as well as in response to dietary intake and anticipated exercise. The goals of intensive therapy were to achieve blood glucose concentrations between 70 and 120 mg/dL before meals, values less than 180 mg/dL after meals, a weekly 3-hour measurement greater than 65 mg/dL, and an HbA1c value within the normal range (6.0% or less). Patients in the intensive treatment group visited their centers each month and had more frequent contacts with a member of the health care team, generally weekly, to review and adjust their regimens.

Conventional therapy consisted of one or two daily injections of insulin, including mixed intermediate and rapid-acting insulins, daily self-monitoring of urine or blood glucose, and education about diet and exercise. The goals of conventional therapy included absence of symptoms of hyperglycemia; absence of ketonuria; maintenance of normal growth, development, and ideal body weight; and freedom from frequent severe hypoglycemia.

The entire cohort of patients was observed for a mean of 6.5 years and 99% of the patients completed the study. Although only 5% of the subjects in the intensive treatment group were able to sustain the goal of a normal HbA1c over time, they nevertheless did have significantly lower average values (approximately 7%) over time than the subjects in the conventional treatment group (approximately 9%). Average capillary blood glucose profiles in the intensive treatment group were 155 ± 30 mg/dL compared with 231 ± 55 mg/dL in the conventional therapy group (P < .0001). These differences in glucose control formed the basis of analyses to determine the effects of lower levels of glycemia on diabetic complications.

When both the primary prevention and secondary intervention cohorts were considered, intensive therapy reduced the risk of proliferative or severe nonproliferative retinopathy by 47% and the need for treatment by photocoagulation by 56%. Intensive therapy reduced the mean adjusted risk of microalbuminuria (defined as urinary albumin excretion of > 40 mg per 24 hours) by 34% in the primary prevention cohort and by 43% in the secondary intervention cohort. The risk of albuminuria was reduced by 56% in the secondary intervention cohort. Intensive therapy reduced the appearance of neuropathy by 69% in the primary prevention cohort and by 57% in the secondary intervention cohort.

There has been considerable controversy regarding potential adverse effects of aggressive insulin therapy in exacerbating the predisposition to macrovascular disease. In this study intensive insulin therapy reduced the development of hypercholesterolemia, defined as a serum concentration of low-density lipoprotein cholesterol greater than 160 mg/dL, by 34% and the risk of macrovascular disease by 41%, although the latter differences were not statistically significant.

Was there an increase in adverse events associated with the intensive treatment regimen? Overall mortality did not differ in the two treatment groups and was actually less than expected on the basis of population-based mortality studies. However, the incidence of severe hypoglycemia was approximately three times higher in the intensive therapy group than in the conventional therapy group (P < .001). Some of the episodes of hypoglycemia were quite severe, resulting in motor vehicle accidents or the need for hospitalization. Severe hypoglycemia occurred more often during sleep and of episodes that occurred while the patients were awake, a significant proportion (approximately one third) were not associated with warning symptoms. In intensively treated subjects predictors of hypoglycemia included a history of severe hypoglycemia, longer duration of diabetes, higher baseline HbA1c, and a lower recent HbA1c.

Weight gain also occurred more frequently in the intensively treated patients. Intensive therapy was associated with a 33% increase in risk of becoming overweight, defined as a body weight more than 120% above the ideal. Five years into the trial, patients being treated intensively had gained a mean of 4.6 kg more than patients receiving conventional therapy. Among subjects in the top quartile of weight gain, changes in plasma lipids, blood pressure, and body fat distribution were observed that were similar to those seen with insulin resistance.
Goals of Treatment

On the basis of these results, the authors of the study recommended that most patients with type 1 diabetes be treated with an intensive treatment regimen under the close supervision of a health care team consisting of a physician, nurses, nutritionist, and behavioral and exercise specialists as needed. However, for certain groups of patients this recommendation may need to be modified because the risk-benefit ratio may not be as favorable as it was in the cohorts with mild or no diabetic complications that were studied in the DCCT. Patients for whom it may be appropriate to be more cautious about instituting intensive treatment regimens include children younger than 13 years, elderly people, and patients with advanced complications such as end-stage renal disease or significant cardiovascular or cerebrovascular disease. It has also been reported that instituting aggressive insulin therapy in subjects with proliferative or severe nonproliferative retinopathy may lead to accelerated progression of retinopathy after the start of intensive therapy. Treatment of the eye disease should be considered before instituting an aggressive insulin regimen. In addition, patients who do not experience warning adrenergic symptoms of hypoglycemia (hypoglycemia unawareness) are at significantly greater risk for severe recurrent hypoglycemia, and this may prevent the safe institution of tight glucose control.

TABLE 30-5 -- Glycemic Control for Nonpregnant Individuals with Diabetes

<table>
<thead>
<tr>
<th>Values</th>
<th>Normal</th>
<th>Goal</th>
<th>Additional Action Suggested</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average preprandial glucose (mg/dL)</td>
<td>&lt;110</td>
<td>90-130</td>
<td>&lt;90/&gt;150</td>
</tr>
<tr>
<td>Average bedtime glucose (mg/dL)</td>
<td>&lt;120</td>
<td>110-150</td>
<td>&lt;110/&gt;180</td>
</tr>
<tr>
<td>Whole blood values</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average preprandial glucose (mg/dL)</td>
<td>&lt;100</td>
<td>80-120</td>
<td>&lt;80/&gt;140</td>
</tr>
<tr>
<td>Average bedtime glucose (mg/dL)</td>
<td>&lt;110</td>
<td>100-140</td>
<td>&lt;100/&gt;160</td>
</tr>
<tr>
<td>Ac1c (%)</td>
<td>&lt;8</td>
<td>&lt;7</td>
<td>&gt;8</td>
</tr>
</tbody>
</table>

From Diabetes Care 2002; 25:S33549, Table 6.

*The values shown in this table are by necessity generalized to the entire population of individuals with diabetes. Patients with comorbid diseases, the very young and older adults, and others with unusual conditions or circumstances may warrant different treatment goals. These values are for nonpregnant adults. Values above or below these levels are not goals, nor are they acceptable in most patients. They are an indication for a significant change in the treatment plan. Additional action suggested depends on individual patients’ circumstances. Such actions may include enhanced diabetes self-management education, comanagement with a diabetes team, referral to an endocrinologist, change in pharmacologic therapy, initiation of or increase in self-monitored blood glucose, or more frequent contact with the patient. Ac1c is referenced to a nondiabetic range of 4.0-6.0% (mean 5.0%, standard deviation 0.5%). Values calibrated to plasma glucose. Measurement of capillary blood glucose.

The American Diabetes Association revises and publishes treatment guidelines annually as the first supplement to Diabetes Care. The recommendations for 2002 are listed in Table 30-5. The goal of therapy is to achieve average preprandial plasma glucose concentrations in the range of 90 to 130 mg/dL, average bedtime plasma glucose values between 110 and 150 mg/dL, and Ac1c values less than 7%. It is recognized that in order to achieve glucose control at this level patients need to monitor glucose levels at least three or four times per day and receive nutritional counseling and training in self-management of the insulin doses and problem solving to allow them to deal with the problems that they encounter in their daily lives. Hospitalization may be necessary for the initiation of therapy.
Team Approach to Treatment

As alluded to previously, because of the complex nature of modern intensive diabetes treatment regimens and the need for regular feedback and modification of the parameters of treatment, it has now become generally accepted that intensive insulin regimens can be instituted more effectively by a health care team rather than a physician alone. Members of the team can include diabetes nurse educators, nutritionists, psychologists, or medical social workers, and others such as exercise physiologists may also be included depending on the needs of a particular patient. A critical aspect of intensive diabetes treatment is the need to monitor continuously the effectiveness of specific components of the regimen and to make adjustments in response to changing life circumstances of the patient.
Pharmacokinetics of Available Insulin Preparations

In the past, insulin for human use was obtained from animal sources—beef and pig. With advances in recombinant DNA technology, it is now possible to produce large quantities of insulin with an amino acid structure identical to that of human insulin using laboratory strains of Escherichia coli bacteria or yeast that have been genetically altered by the addition of the gene for human insulin production. All forms of insulin have identical physiologic effects. They differ in the rapidity of the onset of action, the time from injection to peak action, and the duration of action depending on the chemical nature of the particular insulin preparation. These data are summarized in Table 30–6. The available insulins can be divided into three broad categories on a pharmacokinetic basis.

### Rapid-Acting Insulins

These insulins have an onset of action within an hour or less and are used to reduce the peak of glycemia that occurs after meal ingestion.

1. **Regular insulin.** This form of insulin consists of zinc-insulin crystals dissolved in a clear fluid. The pharmacokinetic profile of regular insulin is related to the fact that after subcutaneous injection it tends to self-aggregate, first into dimers and then into hexamers, which result from the self-association of three dimers. Only the monomeric and dimeric forms can be absorbed to any appreciable degree. The resulting relative delay in the onset and duration of action of regular insulin limits its effectiveness in controlling postprandial glucose.

2. **Insulin lispro** of recombinant DNA origin is a human insulin analogue created when amino acids at positions 28 and 29 on the human insulin B chain are reversed. Insulin lispro was the first insulin analogue to receive approval by the United States Food and Drug Administration. It is chemically Lys(B28), Pro(B29) insulin and is created in a special nonpathogenic laboratory strain of E. coli bacteria that has been genetically altered by the addition of the gene for insulin lispro. The effect of this amino acid rearrangement is to reduce the capacity of the insulin to self-aggregate in subcutaneous tissues, resulting in behavior similar to that of monomeric insulin. This leads to more rapid absorption and shorter duration of action of lispro compared with regular insulin when given by subcutaneous injection. However, lispro is not intrinsically more active and on a molar basis is equipotent to human insulin. When they are given by intravenous injection, the pharmacokinetic profiles of lispro and human regular insulin are similar. Because of its rapid onset of action within 5 to 15 minutes of administration and peak action within 1 to 2 hours, lispro is the first insulin that mimics the time course of the increase in plasma glucose seen after ingestion of a carbohydrate-rich meal.

3. **Insulin aspart** differs from human insulin by substitution of aspartic acid for proline in position 28 on the chain, and this also leads to a more rapid onset and duration of action analogous to those of insulin lispro.

### Intermediate-Acting Insulins

These forms of insulin have a significantly longer delay in their onset and duration of action. In the setting of type 1 diabetes they are generally used in combination with a rapidly acting form of insulin although they may be given before bedtime to limit hyperglycemia overnight and in the early hours of the morning. In general, treatment with monomeric insulin analogues (lispro and aspart) is associated with a lower risk of hypoglycemia, particularly in sleep, than treatment with regular insulin. It is quite easy to document improved glycemic control in the postprandial state. Finally, patients may inject these insulin analogues immediately before or after meals instead of 30 to 60 minutes before meals as is classically recommended with regular insulin, providing greater convenience. These features have been exploited in clinical trials to produce modest improvements in overall control with monomeric insulin analogues versus regular insulin.

#### Intermediate-Acting Insulins

<table>
<thead>
<tr>
<th>Preparation</th>
<th>Onset</th>
<th>Peak</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rapid acting</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regular</td>
<td>½ hr</td>
<td>24 hr</td>
<td>68 hr</td>
</tr>
<tr>
<td>Lispro</td>
<td>15 min</td>
<td>1 hr</td>
<td>34 hr</td>
</tr>
<tr>
<td>Aspart</td>
<td>15 min</td>
<td>1 hr</td>
<td>34 hr</td>
</tr>
<tr>
<td>NPH</td>
<td>13 hr</td>
<td>68 hr</td>
<td>1216 hr</td>
</tr>
<tr>
<td>Lente</td>
<td>14 hr</td>
<td>610 hr</td>
<td>1418 hr</td>
</tr>
<tr>
<td>Ultralente</td>
<td>24 hr</td>
<td>810 hr</td>
<td>1624 hr</td>
</tr>
</tbody>
</table>

### Long-Acting Insulins

These forms of insulin have a significantly longer delay in their onset and duration of action. In the setting of type 1 diabetes they are generally used in combination with a rapidly acting form of insulin although they may be given before bedtime to limit hyperglycemia overnight and in the early hours of the morning.

1. **NPH insulin.** This is a crystalline suspension of insulin with protamine and zinc providing an intermediate-acting insulin with onset of action in 1 to 3 hours, duration of action up to 24 hours, and peak action from 6 to 8 hours.

2. **Lente insulin.** This is an amorphous and crystalline suspension of insulin with zinc resulting in an intermediate pharmacokinetic profile. It generally has a slightly longer onset, peak, and duration of action than NPH insulin.

### Long-Acting Insulins

Even after an overnight fast, the normal pancreas continues to secrete insulin. Investigators have long attempted to develop forms of insulin with pharmacokinetic properties that simulate the basal production of insulin.

1. **Ultralente insulin.** Ultralente insulin is also a preparation of insulin consisting of zinc insulin crystals. It has an onset of action in 2 to 4 hours and a prolonged duration that may be as long as 24 hours. The peak of action is broad and spans 8 to 16 hours. This formulation of insulin is arguably the most variably absorbed and can be associated in some patients with substantial variability in glycemic control.

2. **Insulin glargine.** Insulin glargine is a recombinant human insulin analogue that is long acting. It differs from human insulin in that the amino acid asparagine at position A21 is replaced by glycine and two arginines are added to the C-terminus of the B chain. In the injection solution at pH 4, insulin glargine is completely soluble. However, it has low aqueous solubility at neutral pH. After injection into the subcutaneous tissue, the acidic solution is neutralized, leading to the formation of microprecipitates from which small amounts of insulin glargine are slowly released, resulting in a broad increase in concentration over a 24-hour period with no pronounced peak. It thus simulates the basal production of insulin. In other respects the mechanism of action of glargine insulin is similar to that of aspartic acid for proline in position 28 on the chain, and this also leads to a more rapid onset and duration of action analogous to those of insulin lispro.
of human insulin, and on a molar basis its glucose-lowering effects are similar to those of human insulin when given intravenously. Because this insulin is provided in an acid vehicle, it cannot be mixed with other forms of insulin or intravenous fluids and some patients have greater discomfort with injection at least some of the time. In general, glargine is less variably absorbed than Ultralente, NPH, and Lente insulin and in clinical trials in patients with type 1 diabetes has been associated with a reduced risk of hypoglycemia, particularly nocturnal hypoglycemia. In about 10% of patients, insulin glargine must be taken twice daily to provide 24-hour coverage of basal insulin needs. In a smaller proportion of patients, there may be a modest peak in effect approximately 2 hours after injection.
Approach to the Treatment of Type 1 Diabetes

Levels of glucose control equivalent to those achieved in the intensive treatment group in the DCCT are not possible in the majority of patients unless the insulin regimen utilizes more than two injections of insulin with the patient adjusting the insulin dose depending on the results of self-monitoring of glucose as well as on the basis of dietary intake and physical activity. The reason for this is relatively simple. In patients with little or no endogenous insulin production, the exogenous insulin regimen needs to simulate the multiphasic profile of insulin secretory responses to meals and snacks present in normal subjects if levels of glycemia approaching normal are to be achieved.

A number of regimens have been used to achieve these ends. Three basic approaches are reviewed, although it is clear that other possible approaches may be effective in individual patients. Achieving the glycemic goals of therapy is far more important than the details of the insulin regimen. Nevertheless, one of the following three general approaches to therapy is most likely to lead to the desired outcome.

Combination of Rapid-Acting and Intermediate-Acting Insulin with Breakfast and Dinner and Intermediate-Acting Insulin at Bedtime

The rationale for these regimens is that the rapid-acting insulin (regular, lispro, or aspart) limits the postprandial glycemia after breakfast and dinner, the intermediate-acting insulin (NPH or Lente) administered before breakfast limits glycemia in the afternoon, and the intermediate-acting insulin before dinner limits glycemia in the early hours of the morning. Although such a regimen may be sufficient to achieve glucose targets in some patients, in many individuals the intermediate-acting insulin given before dinner is insufficient to control elevations in blood glucose commonly seen in the early morning (dawn phenomenon). Attempts to increase the dose of intermediate-acting insulin at dinner expose the patient to a greater risk of hypoglycemia in the middle of the night, hence the need for a smaller dose at bedtime to provide sufficient insulin to restrain the dawn phenomenon the following morning while moderating the risk of nocturnal hypoglycemia. This three-injection regimen was the mainstay of therapy in the DCCT but is rapidly being supplanted by regimens that take greater advantage of the availability of insulin analogues.

Combination of Rapid-Acting Insulin Given with Meals and Long-Acting Insulin at Bedtime

This combination of insulins can also simulate the pattern of insulin production that occurs normally. Use of monomeric insulin analogues provides excellent meal coverage. Use of insulin glargine provides excellent control of the fasting plasma glucose. This combination of rapid-acting monomeric insulin analogues with insulin glargine is rapidly supplanting human insulin-based treatment regimens as it seems to be associated with less variability in glycemic control associated with lower risks of hypoglycemia. Human Ultralente can be used as the basal insulin in such a regimen but would generally need to be administered twice daily.

Insulin Administration by an External Insulin Pump

An alternative method of delivering insulin is by an external mechanical pump. This approach involves the administration of a rapidly acting insulin preparation delivered by continuous subcutaneous infusion through a catheter usually inserted into the subcutaneous tissues of the anterior abdominal wall. The pumps deliver insulin as a programmed basal infusion as well as patient-directed boluses given before meals or snacks or in response to elevations in the blood glucose concentration outside the desired range. With currently available pumps, the basal insulin infusion rate (usually approximately 1 U/hour) can be programmed either to continue at a constant rate over the 24-hour period or more commonly to increase and decrease at predetermined times of the day to prevent anticipated excursions in the blood glucose concentration, for example, morning rises in glucose. Newer pumps provide the ability to use multiple basal profiles to deal with recurrent patterns (e.g., menstruation, weekends, activity). Protocols for insulin administration by the pump usually require approximately half the insulin to be administered as a basal infusion and the remainder as premeal boluses.

Insulin administration by an external pump has some advantages over regimens that utilize multiple insulin injections. Only rapidly acting insulin is used in the insulin pump. Consequently, adjustments to the basal insulin infusion rate or changes in the size and timing of the insulin boluses result in more rapid changes in the blood glucose concentration than are possible when adjustments are made to the dose of intermediate-acting or long-acting insulin. This leads to greater flexibility for the patient. It has been suggested that use of lispro insulin may lead to a lower risk of hypoglycemia.

However, there are also disadvantages of insulin pump therapy. There

### Table 30-7 – Sample Algorithm for Premeal Insulin Lispro Supplements

<table>
<thead>
<tr>
<th>Premeal Blood Glucose (mmol/L)</th>
<th>Insulin Dose</th>
<th>Lag Time</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;2.8</td>
<td>Decrease by 2 U</td>
<td>Inject during meal</td>
<td>Include at least 10 g of simple carbohydrate in the meal</td>
</tr>
<tr>
<td>2.8 to 4.4</td>
<td>Decrease by 1 U</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>4.4 to 7.2</td>
<td>No change</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>7.3 to 10.0</td>
<td>Increase by 1 U</td>
<td>10 min</td>
<td></td>
</tr>
<tr>
<td>11.1 to 20.0</td>
<td>Increase by 2 U</td>
<td>15 min</td>
<td></td>
</tr>
<tr>
<td>13.9 to 20.0</td>
<td>Increase by 3 U</td>
<td>20 min</td>
<td>Urinary ketones, especially with CSII</td>
</tr>
<tr>
<td>16.7 to 30.0</td>
<td>Increase by 5 U</td>
<td>25 min</td>
<td>Urinary ketones, especially with CSII; if moderate or large, increase fluid intake, consider additional insulin</td>
</tr>
</tbody>
</table>


CSII, continuous subcutaneous insulin infusion.

*Assumes preprandial target of 4.4 to 7.2 mmol/L (80 to 130 mg/dL). Initial supplements may be based on previous experience or 1500 rule (1500/total insulin dose = the effect of 1 U insulin on blood glucose decrease in mg/dL). Plans should be individualized for each patient. This example would likely be treated with 50 units per day of insulin. Refers to the amount of time between injecting the insulin and eating.*

is a significant initial cost of the pump itself. Furthermore, the tubing, which needs to be changed every 24 to 72 hours, and other supplies are expensive. The risk of infection at the site of insulin administration is significant. Infections occur on average once annually per patient even in the best of practices, and although these can usually be treated by changing the site of infusion and giving a short course of oral antibiotics, if an abscess develops surgical drainage may be necessary. In addition, because only rapidly acting insulin is used, pump failure as a result of mechanical malfunction or catheter-related problems can quickly result in severe
hyperglycemia and even ketoacidosis. Patients treated with insulin pump therapy must monitor glucose frequently and always be alert to the possibility of failure of the infusion system.

Controlled clinical trials have indicated that on average intensive insulin regimens that use multiple insulin injections lead to levels of glucose control similar to those achieved with the insulin pump. On the other hand, there are some patients who never achieve adequate control with multiple daily injections but experience dramatic improvements with pump therapy. According to the Clinical Practice Recommendations of the American Diabetes Association, the insulin pump should be used only by candidates strongly motivated to improve glucose control and willing to work with their health care provider in assuming substantial responsibility for their day-to-day care. They must also understand and demonstrate use of the insulin pump and self-monitoring of blood glucose and be able to use the data obtained in an appropriate fashion.
Algorithms of Insulin Administration

An essential component of intensive regimens of insulin replacement is the need to make regular adjustments to the insulin dose depending on the prevailing blood glucose concentration. Algorithms have been developed to guide these adjustments that aim to simulate the normal feedback control of insulin secretion whereby hyperglycemia stimulates and hypoglycemia inhibits insulin secretion. They all involve frequent monitoring of the blood glucose concentration, generally three or four times per day or more, with increases in the insulin dose if glucose levels are above the target upper level that is judged to be acceptable and reductions in the insulin dose if glucose levels are below the acceptable lower level. Many algorithms are in use; one example is shown in Table 30-7.
Complications of Insulin Therapy

Hypoglycemia

The most serious complication of intensive regimens of insulin replacement is hypoglycemia, and this is generally the factor that limits patients’ ability to achieve tight glucose control. In the DCCT, patients in the intensive treatment group had an approximately threefold greater risk of hypoglycemia than those in the conventional treatment group. Hypoglycemia may be life-threatening, leading to motor vehicle accidents, serious falls with fractures, and seizures. Patients with type 1 diabetes have serious defects in mechanisms responsible for glucose counterregulation, and this is a major underlying reason for the predisposition to hypoglycemia. Glucose counter-regulation is reviewed in detail in Chapter 32.

The risk of hypoglycemia can be reduced if all patients treated with intensive regimens of insulin replacement are carefully educated about recognition of hypoglycemic symptoms and measures that should be taken to avoid more serious hypoglycemia after symptoms are initially experienced. Certain patients, particularly those with long-standing diabetes and autonomic neuropathy, may not subjectively sense symptoms of hypoglycemia even in the presence of low glucose concentrations. Glycemic targets of therapy should be adjusted upward in these subjects because they are at particularly high risk of hypoglycemia. Similarly, patients with advanced end-stage microvascular or macrovascular diabetic complications in whom the benefit of intensive glucose control is likely to be less should not be exposed to the increased risk of hypoglycemia that is inherent in extremely intensive insulin treatment regimens.

In addition to the availability of glucose tablets, hard candy or other sources of a readily absorbable form of carbohydrate, all patients with type 1 diabetes should have emergency glucagon kits at home and at work, assuming that there are people in those settings who can be trained in their use. The administration of 0.5 to 1 mg of glucagon intramuscularly to a severely symptomatic person with hypoglycemia rapidly raises the plasma glucose concentration to an acceptable range and avoids the difficulties and dangers associated with attempting to get a stuporous or disoriented individual to ingest glucose by mouth. Nevertheless, because of occasional failures of glucagon to reverse hypoglycemia fully, friends and family members should always be instructed to call for medical assistance as soon as the injection is provided.

Weight Gain

Improvement in glucose control with a reduction in glycosuria is invariably associated with weight gain as the leakage of calories into the urine is reduced or eliminated. In addition, increased food intake to treat or prevent hypoglycemia can contribute to weight gain. Insulin itself may stimulate appetite. As a result of the combination of all these effects, weight gain is common, particularly with intensive regimens of insulin replacement.

Worsening of Retinopathy

Institution of regimens of tight glucose control has been reported to exacerbate the underlying retinopathy. Thus, if a patient with serious background or proliferative retinopathy presents in poor glucose control, consideration should be given to ophthalmologic treatment of the retinopathy before instituting tight glucose control.

Insulin Allergy

Insulin allergy has become much less common with the use of human insulin. Most manifestations of allergic reactions to insulin consist of local wheal-and-flare reactions at the site of injection. The allergic reaction can be to the insulin itself or to other components of the insulin preparation such as the protamine in NPH insulin. Occasionally, more generalized allergic reactions occur, and even more rarely anaphylactic reactions take place. In general, mild local allergic reactions to insulin can be treated with antihistamines. More severe reactions require desensitization. Admission to the hospital is necessary, and under close supervision of a physician with access to equipment for emergency resuscitation a protocol is followed in which the patient is exposed to gradually increasing amounts of insulin administered according to a set schedule.
ACUTE DIABETIC EMERGENCIES: DIABETIC KETOACIDOSIS

Diabetic ketoacidosis (DKA) is a life-threatening condition in which severe insulin deficiency leads to hyperglycemia, excessive lipolysis, and unrestrained fatty acid oxidation producing the ketone bodies acetone, hydroxybutyrate, and acetoacetate. This results in metabolic acidosis, dehydration, and deficits in fluid and electrolytes. Excess secretion of primarily glucagon as well as catecholamines, glucocorticoids, and growth hormone in combination with insulin deficiency produces hyperglycemia by stimulating glycogenolysis and gluconeogenesis and impairing glucose disposal. DKA is a far more characteristic feature of type 1 than of type 2 diabetes but may be seen in persons with type 2 diabetes under conditions of stress such as occurs with serious infections, trauma, and cardiovascular or other emergencies.

Clinical Presentation

Patients with uncontrolled diabetes present with nonspecific complaints. If the disease follows an indolent course over months to years, patients can manifest profound wasting, cachexia, and prostration similar in degree to those of patients with long-standing malignancy or chronic infection. With significant physical or emotional stress, sudden metabolic decompensation can occur. The cases of DKA that are misdiagnosed usually occur in patients with new-onset diabetes. Polyuria (or at least nocturia) and weight loss are almost always present, although often not reported by the patient. Any patient with severe illness (acute or chronic) or neurologic changes should have glucose and electrolytes measured.

In DKA, metabolic decompensation usually develops over a period of hours to a few days. Patients with DKA classically present with lethargy and a characteristic hyperventilation pattern with deep slow breaths (Kussmaul's respirations) associated with the fruity odor of acetone. They often complain of nausea and vomiting, with abdominal pain being somewhat less frequent. The abdominal pain can be quite severe and may be associated with distention, ileus, and tenderness without rebound but usually resolves relatively quickly with therapy unless there is underlying abdominal pathology. Most patients are normotensive, tachycardic, and tachypneic with signs of mild to moderate volume depletion. Hypothermia has been described in DKA, and patients with underlying infection may not manifest fever. Cerebral edema does occur, generally during therapy. Patients with DKA can be stuporous with obvious profound dehydration and often demonstrate focal neurologic deficits such as Babinski’s reflexes, asymmetric reflexes, cranial nerve findings, paresis, fasciculations, and aphasia.
Laboratory Test Results and Differential Diagnosis

Laboratory tests that would be routinely monitored in the setting of DKA include the following:

1. Hemoglobin, white blood cell and differential count.
2. Glucose, creatinine, blood urea nitrogen (BUN), and serum potassium.
3. Serum sodium, chlorine, bicarbonate, and anion gap.

The sine qua non of DKA is acidosis, and the serum HCO₃⁻ concentration is usually less than 10 mEq/L. The acidosis is due to production and accumulation of ketones in the serum. Three ketones are produced in DKA: ketoacids, -hydroxybutyrate, and acetoacetate, as well as the neutral ketone aceton. Ketones can be detected in serum and urine using the nitroprusside reaction on diagnostic strips for use at the patient's bedside or in the clinical laboratory. This test detects -hydroxybutyrate more effectively than aceton and does not detect an increased concentration of -hydroxybutyrate. Particularly in severe DKA, -hydroxybutyrate is the predominant ketone, and it is possible although unusual to have a negative serum nitroprusside reaction in the presence of severe ketosis. However, under these circumstances the serum HCO₃⁻ is still markedly reduced and the anion gap is increased, indicative of the presence of metabolic acidosis. The urinary -hydroxybutyrate can be measured at many centers and commercially but is not usually readily available. The anion gap is a readily available index for unmeasured anions in the blood (normal < 14 mEq/L): Anion gap = sodium (chloride + bicarbonate) - lactate - protein - ketone.

Most patients with DKA present with an anion gap greater than 20 mEq/L, and some with a gap greater than 40 mEq/L. However, occasional patients have a hyperchloremic metabolic acidosis without a significant anion gap. Patients with DKA almost invariably have large amounts of ketones in their urine. The serum glucose in DKA is usually in the 500 mg/dL range. However, an entity known as euglycemic DKA has been described, particularly in the presence of decreased oral intake or in pregnancy, in which the serum glucose is normal or near normal but the patient requires insulin therapy for the clearance of ketoacidosis. The arterial pH is commonly less than 7.3 and can be as low as 6.5. There is partial respiratory compensation with hypocarbia. Patients are often mildly hyperosmolar, although osmolalities of greater than 330 mOsm/kg are unusual without mental status changes.

Not all patients with hyperglycemia and an anion gap metabolic acidosis have DKA and other causes of metabolic acidosis must be considered in these patients, particularly if the serum or urine ketone measurements are not elevated. The following causes of metabolic acidosis need to be considered in the differential diagnosis of DKA.

1. Lactic acidosis is the most common cause of metabolic acidosis in hospitalized patients and can be seen in patients with uncomplicated diabetes as well as those with DKA. Lactic acidosis usually occurs in the setting of decreased tissue oxygen delivery resulting in the nonoxidative metabolism of glucose to lactic acid. Lactic acidosis complicates other primary metabolic acidoses as a consequence of dehydration or shock, and assessing its relative contribution can be difficult. The presentation is identical to that of DKA. In pure lactic acidosis, the serum glucose and ketones should be normal and the serum lactate concentration should be greater than 5 mM. The therapy of lactic acidosis is directed at the underlying cause and optimizing tissue perfusion.

2. Starvation ketosis results from inadequate carbohydrate availability resulting in physiologically appropriate lipolysis and ketone production to provide fuel substrates for muscle. The blood glucose is usually normal. Although the urine can have large amounts of ketones, the blood rarely does. Arterial pH is normal, and the anion gap is at most mildly elevated.

3. Alcoholic ketoacidosis is a more severe form of starvation ketosis wherein the appropriate ketogenic response to poor carbohydrate intake is increased through as yet poorly defined effects of alcohol on the liver. Classically, these patients are long-standing alcoholics for whom ethanol has been the main caloric source for days to weeks. The ketoacidosis occurs when for some reason alcohol and caloric intake decreases. In isolated alcoholic ketoacidosis, the metabolic acidosis is usually mild to moderate in severity. The anion gap is elevated. Serum and urine ketones are always present. However, alcoholic ketoacidosis produces an even higher ratio of -hydroxybutyrate to acetoacetate than DKA, and negative or weakly positive nitroprusside reactions are common. Respiratory alkalosis associated with delirium tremens, agitation, or pulmonary processes often normalizes the pH but should be evident with careful analysis of acid-base status. Usually, the patient is normoglycemic or hypoglycemic, although mild hyperglycemia is occasionally present. Patients who are significantly hyperglycemic should be treated as if they have DKA. The therapy of alcoholic ketoacidosis consists of thiamine, carbohydrates, fluids, and electrolytes with special attention to the more severe consequences of alcohol toxicity, alcohol withdrawal, and chronic malnutrition. In more severely ill patients in whom alcoholic ketoacidosis is considered a possibility, there is usually another underlying illness such as pancreatitis, gastrointestinal bleeding, hepatic encephalopathy, delirium tremens, or infection complicated by concomitant lactic acidosis.

4. Uremic acidosis is characterized by extremely large elevations in the BUN (often > 200) and creatinine (>10) with normoglycemia. The pH and anion gap are usually only mildly abnormal. The treatment is supportive with careful attention to fluid and electrolytes until dialysis can be performed. Rhabdomyolysis is a cause of renal failure in which the anion gap can be significantly elevated and acidosis can be severe. There should be marked elevation of creatine phosphokinase and myoglobin. It should be noted that mild rhabdomyolysis is not uncommon in DKA, but the presence of hyperglycemia and ketonemia leaves no doubt about the primary etiology of the acidosis.

5. Toxic ingestions can be differentiated from DKA by history and laboratory investigation. Salicylate intoxication produces an anion gap metabolic acidosis usually with a respiratory alkalosis. The plasma glucose is normal or low, the osmolality normal, ketones negative, and salicylates can be detected in the urine or blood. It should be noted that salicylates can cause a false-positive glucose determination when using the cupric sulfate method and a false-negative result when using the glucose oxidase reaction. Methanol and ethylene glycol also produce an anion gap metabolic acidosis without hyperglycemia or ketones but need to be kept in mind primarily because they produce an increase in the measured serum osmolality but not in the calculated serum osmolyte osmolar gap. Their serum levels can also be measured. Isopropyl alcohol does not cause a metabolic acidosis but should be remembered because it is metabolized to aceton, which can produce a positive result in the nitroprusside reaction commonly used for the detection of ketoads. These intoxications must be appropriately treated.

Rare cases of anion gap acidoses have been reported with other ingestions including toluene, iron, hydrogen sulfide, nalidixic acid, papaverine, paraldehyde, strychnine, isoniazid, and outdated tetacycline.

When DKA is considered, the diagnosis can be made quickly with routine laboratory tests. Blood and urine glucose and ketones can be obtained in minutes with glucose oxidase/impregnated strips and the nitroprusside reaction, respectively.

Osmolality

The increase in osmolality that occurs in DKA is must be differentiated from the increase in osmolality seen in hyperosmolar-hyperglycemic nonketotic (diabetic) coma (HHNC). The osmolality can be measured by freezing point depression or estimated using the following formula: Osmolarity (mOsm/L) = 2 × sodium + glucose/18 + urea/2.8 + ethanol/4.6

Patients with DKA not uncommonly present with hyperosmolarity and coma. In HHNC, the osmolality is generally greater than 350 mOsm/L and can exceed 400 mOsm/L. The serum sodium and potassium can be high, normal, or low and do not reflect total-body levels, which are uniformly depleted. The glucose is usually greater than 600 mg/dL, and levels over 1000 mg/dL are quite common. In pure HHNC, there is not a significant metabolic acidosis or anion gap.

It should be remembered that patients often present with combinations of the preceding findings. HHNC can involve mild to moderate ketonemia and acidosis. Alcoholic ketoacidosis can contribute to either DKA or HHNC. Lactic acidosis is common in severe DKA and HHNC. Any patient with hyperglycemia greater than 250 mg/dL and an anion gap metabolic acidosis should be treated by the general principles outlined in the following with special consideration of other possible...
contributing metabolic acidoses.
Therapy

The optimal management of DKA has been the source of considerable controversy over the past half-century. Only recently have prospective studies of various therapeutic approaches been performed. The guidelines we propose rely heavily on prospective studies of DKA by Kitabchi and co-workers. The general approach is to (1) provide necessary fluids to restore the circulation, (2) treat insulin deficiency with continuous insulin, (3) treat electrolyte disturbances, (4) observe the patient closely and carefully, and (5) search for underlying causes of metabolic decompensation.

Fluids

Volume contraction is one of the hallmarks of DKA. It can contribute to acidosis through lactic acid production as well as decreased renal clearance of organic and inorganic acids. It contributes to hyperglycemia by decreasing renal clearance of glucose. If decreased tissue perfusion is significant, it causes insulin resistance by decreasing insulin delivery to the sites of insulin-mediated glucose disposal, namely muscle and adipose tissue, as well as through stimulation of catecholamine and glucocorticoid secretion. Fluid deficits on the order of 5 to 10 L are common in DKA. It should be remembered that the urine produced during the osmotic diuresis of hyperglycemia is approximately half-normal with respect to sodium. Therefore, water deficits are in excess of sodium deficits. Historically, large quantities of isotonic intravenous fluids have been administered rapidly to patients in DKA. Therefore, patients with a history of congestive heart failure, chronic or acute renal failure, severe hypotension, or significant pulmonary disease, early invasive hemodynamic monitoring should be considered.

When there is physical evidence of dehydration that is, hypotension, decreased skin turgor, or dry mucous membranes administer 1 L of normal saline over the first hour and 200 to 500 mL/hour in subsequent hours until hypotension resolves and adequate circulation is maintained. If hypotension is severe with clinical evidence of hyperperfusion and does not respond to crystalloid, therapy with colloid is considered, often in combination with invasive hemodynamic monitoring. If there is no hypotension and no concern about renal failure, administer 1 L of half-normal saline over the first hour.

During that first hour, the laboratory data usually return and can be quite helpful in planning further therapy. Despite the excess of water losses over sodium, the measured sodium is usually low because of osmotic effects of glucose. These osmotic effects can be corrected using a simple formula: Corrected sodium concentration = measured sodium + 0.016(glucose 100) - (sodium/140) 1

Severe hypertglycemia, which is common in severe diabetes, can cause a false decrease in the serum sodium concentration by approximately 1.0 mEq/L at a serum lipid concentration of 460 mg/dL. An estimated water deficit can be calculated using the corrected sodium: Water deficit in liters = 0.6 (weight in kg - corrected sodium) - (sodium/140)

Using these formulas, a 70-kg patient with a measured sodium level of 140 mEq/L and a glucose concentration of 1000 mg/dL would have a calculated water deficit of 4.3 L. If the patient is normotensive after the first liter of fluids, it would be reasonable to aim to replace urinary losses with one-half normal saline and also provide approximately one half the water deficit as 5% dextrose over the first 12 to 24 hours (using the preceding example, 2 L) and the remainder over the subsequent 24 hours. The plan for fluid therapy should be continuously reevaluated in light of the clinical and laboratory response of the patient. When the serum glucose reaches 250 to 300 mg/dL, all fluids should contain 5% dextrose and therapy should be aimed at maintaining the serum glucose in that range for 24 hours to allow slow equilibration of osmotically active substances across cell membranes.

The primary goal of fluid therapy is to maintain an adequate circulation and secondly to maintain a brisk diuresis. Beyond that, pulmonary edema, hyperchloremic metabolic acidosis, and a rapid fall in the serum osmolality should be avoided by frequent monitoring of the patient, glucose, and electrolytes. It has been demonstrated that fluid administration and subsequent continued osmotic diuresis are responsible for a large portion of the initial decline in glucose during therapy.

Insulin

Insulin is the mainstay of therapy of DKA because it is essentially an insulin-deficient state. In the past, high doses of insulin (upward of 50 U/hour) were favored. In later studies, low-dose insulin therapy (0.1 U/kg per hour) has been shown to be as effective as higher doses in producing a decrease in serum glucose and clearance of ketones. Furthermore, low-dose therapy results in a reduction in the major morbidity of intensive insulin therapy, namely hypoglycemia and hypokalemia.

Studies have also shown that intravenous insulin is significantly more effective than intramuscular or subcutaneous insulin in lowering the ketone body concentration over the first 2 hours of therapy. The subcutaneous route is inappropriate for the critically ill patient because of the possibility of tissue hypoperfusion and slower kinetics of absorption. There are numerous studies that attest to the efficacy of intramuscular therapy in severe DKA. In cases in which there is insufficient nursing monitoring or intravenous access to allow safe intravenous administration, intramuscular therapy would be the route of choice.

Lastly, it has been shown that a 10-U intravenous insulin priming dose when insulin therapy is started significantly improves the glycemic response to the first hour of therapy. The rationale is to saturate insulin receptors fully before beginning continuous therapy and to avoid the lag time necessary to achieve steady-state insulin levels. When mixing insulin in normal saline, it does not seem to be necessary to add albumin to prevent insulin adsorption to the infusion set. However, the intravenous tubing should be flushed with the insulin infused before use.

In the rare instances in which the glucose does not decrease at least 10% or 50 mg/dL in an hour, the insulin infusion rate should be increased by 50% to 100% and a second bolus of intravenous insulin administered. As the glucose level decreases, it is usually necessary to decrease the rate of infusion. After the glucose reaches approximately 250 mg/dL, it is prudent to decrease the insulin infusion rate and administer dextrose. It usually takes an additional 12 to 24 hours to clear ketones from the circulation after hyperglycemia is controlled. With resolution of ketosis, the rate of infusion approaches the physiologic range of 0.3 to 0.5 U/kg per day.

When the decision is made to feed the patient, the patient should be switched from intravenous or intramuscular therapy to subcutaneous therapy. Subcutaneous insulin should be administered before a meal and the insulin drip discontinued approximately 30 minutes later. The glucose should be checked in 2 hours and at least every 4 hours subsequently until a relatively stable insulin regimen is determined.

Potassium

Potassium losses during the development of DKA are usually quite high (3 to 10 mEq/kg) and are mediated by shifts to the extracellular space secondary to acidosis and protein catabolism compounded by hyperaldosteronism and osmotic diuresis. Although most patients with DKA or HHNC have normal or even high serum potassium at presentation, the initial therapy with fluids and insulin causes it to fall.

Our approach has been to monitor the electrocardiogram (ECG) for signs of hyperkalemia (peaked T wave, QRS widening) initially and to administer potassium if these are absent and the serum potassium is less than 5.5 mEq/L. If the patient is oliguric, we do not administer potassium unless the serum concentration is less than 4 mEq/L, or there are ECG signs of hypokalemia (U wave), and even then it is done with extreme caution. With therapy of DKA, the potassium level always falls, usually reaching a nadir after several hours. We usually replace potassium at 10 to 20 mEq/hour, one half as potassium chloride and one half as potassium phosphate, and monitor serum levels at least every 2 hours initially as well as follow ECG morphology. Occasionally, patients with DKA who have had protracted courses with vomiting present with hypokalemia and acidosis and may require 40 to 80 mEq/hour by central line to avoid further decreases in the serum potassium.
Like potassium, phosphate is depleted in patients with DKA. Although patients usually present with elevated serum phosphate, the serum level declines with therapy. No well-documented clinical significance of these findings has been determined and no benefit of phosphate administration has been demonstrated, but most authorities recommend phosphate therapy as before and monitoring for its possible complications—hypocalcemia and hypomagnesemia.

**Bicarbonate**

Serum bicarbonate is always low in DKA, but a true deficit is not present because the ketoacid and lactate anions are metabolized to bicarbonate during therapy. The use of bicarbonate in the therapy of DKA is highly controversial. No benefit of bicarbonate therapy has been demonstrated in clinical trials. In fact, in two trials, hypokalemia was more common in bicarbonate-treated patients. There are theoretical considerations against the use of bicarbonate. Cellular levels of 2,3-diphosphoglycerate are depleted in DKA, causing a shift in the oxyhemoglobin dissociation curve to the left and thus impairing tissue oxygen delivery. Acidemia has the opposite effect, and therefore reversing acidosis acutely could decrease tissue oxygen delivery. In addition, there are in vitro data suggesting that pH is a regulator of cellular lactate metabolism and correction of acidosis could increase lactate production. These observations are of questionable clinical relevance, however.

We reserve bicarbonate therapy for use (1) in patients with severe acidosis (pH < 6.9), (2) in the presence of hemodynamic instability if the pH is less than 7.1, or (3) in cases of hyperkalemia with ECG findings. When bicarbonate is used, it should be used sparingly and considered a temporizing measure while definitive therapy with insulin and fluids is under way. Approximately 1 mEq/kg of bicarbonate is administered as a rapid infusion over 10 to 15 minutes with further therapy based on repeated arterial blood gases every 30 to 120 minutes. Potassium therapy should be considered before treatment with bicarbonate as transient hypokalemia is not an uncommon complication of the administration of alkali.

**Monitoring**

It is possible to manage many cases of mild DKA without admission to the intensive care unit, depending on staff availability. We routinely admit patients with DKA to the intensive care unit if they have a pH less than 7.3. If mental status is compromised, prophylactic intubation is considered and nasogastric suctioning is always performed because of frequent ileus and danger of aspiration. If the patient cannot void at will, bladder catheterization is necessary to follow urine output adequately. ECG monitoring is continuous with hourly documentation of QRS intervals as well as T-wave morphology. Initially, serum glucose, electrolytes, BUN, creatinine, calcium, magnesium, phosphate, ketones, lactate, creatine phosphokinase, and liver function tests as well as urinalysis, ECG, upright chest radiograph, complete blood count, and arterial blood gases are obtained. If there is any concern about possible toxic ingestions, toxicology screening is also performed. Subsequently, glucose and electrolytes are measured at least hourly, calcium, magnesium, and phosphate every 2 hours; and BUN, creatinine, and ketones every 6 to 24 hours.

It is often not necessary to monitor arterial blood gases routinely because bicarbonate and anion gap are relatively good indices of the response to therapy. Monitoring venous pH has also been shown to reflect acidemia and response to therapy adequately. Usually, frequent blood work is necessary only for the first 12 hours or so. In the severely ill patient with obvious underlying disease, the course is often more protracted and, particularly when venous access is a problem, early consideration should be given to placement of an arterial line. A flow sheet tabulating these findings as well as mental status, vital signs, insulin dose, fluid, and electrolytes administered, and urine output allows easy analysis of response to therapy. When the acidosis begins to resolve and the response to therapy becomes predictable, it is reasonable to curtail laboratory use. If cardiovascular status is unclear or troublesome, invasive hemodynamic monitoring is an appropriate guide for fluid therapy. The goals should be to achieve hemodynamic stability rapidly and to correct DKA fully in 12 to 36 hours.

**Search for Underlying Causes**

After stabilizing the patient, a careful history and physical examination and a diagnostic strategy should be aimed at determining the precipitating event. In most inner-city practices, the most common cause of DKA is noncompliance with insulin therapy and is usually easily treated. The second most common cause is infection, with viral syndromes, urinary tract infection, pelvic inflammatory disease, and pneumonia predominating. It is often difficult to determine initially whether the patient is infected. Fever can be absent in a significant proportion of patients with diabetic emergencies. The white blood cell count is not uncommonly elevated in the range of 20,000 or higher even in the absence of infection. As a result, cultures should be performed for most patients, and if there is significant concern about infection, empirical broad antibiotic coverage should be considered pending microbiologic findings.

Special consideration should be given to ruling out meningitis in the patient with altered mental status. In this regard, most would perform lumbar punctures in all patients with meningismus and in patients with disproportionate mental status changes. If the index of suspicion is lower, a head CT is performed, and if normal, a lumbar puncture should be performed. The cerebrospinal fluid glucose is not particularly useful in determining whether the fluid is infected, and a cerebrospinal fluid glucose level less than 100 mg/dL is unusual when the serum glucose is greater than 250 mg/dL. The relative frequency of sinus infection (particularly Mucor), foot infection, bacterial arthritis, cholecystitis, cellulitis, and necrotizing fasciitis should also be considered.

Pneumonia can be difficult to diagnose in patients with dehydration because the alveolar edema fluid that shows up as an infiltrate on chest radiographs is often not present but develops along with progressive hypoxia during hydration. To avoid this occurrence, we administer intravenous fluid judiciously to patients we suspect have pneumonia. Pancreatitis and pregnancy are common precipitants and should be especially considered when assessing the abdominal pain that is almost ubiquitous at presentation. Abdominal guarding and tenderness associated with vomiting are common, and rebound is occasionally present. These symptoms and findings usually resolve quickly with therapy in the absence of intra-abdominal pathology. The serum amylase is often elevated without pathologic significance, although lipase is usually more specific. Acute myocardial infarction and stroke as well as thromboembolic phenomena are frequent precipitants and complications of DKA.

The more insulin resistant the patient seems to be, the more likely one is to find a precipitating cause. If a precipitating cause is found, treatment is essential if adequate metabolic control is to be achieved.
Complications and Prognosis

It should now be possible to treat almost all cases of DKA successfully. The most troublesome complication is cerebral edema. It is common particularly in children and can be fatal. In most series, specific causes could not be assigned, although aggressive hydration, particularly with hypotonic fluids, may contribute. In 50% of patients who subsequently had a respiratory arrest, there were premonitory symptoms, and despite early intervention only half of them avoided severe or fatal brain damage. Other complications of life-threatening severity that have been reported include the acute respiratory distress syndrome and bronchial mucous plugging. Arterial and venous thromboembolic events are quite common. Standard prophylactic low-dose heparin is certainly reasonable in patients with DKA, but currently no indication exists for full anticoagulation.
References


41. AwaTa T, Guberski DL, Like AA. Genetics of the BB rat: association of autoimmune disorders (diabetes, insulinis, and thyroiditis) with lymphopenia and major histocompatibility complex class II. Endocrinology 1995; 136:5715735.


References


Chapter 31 - Complications of Diabetes Mellitus

BIOCHEMISTRY AND MOLECULAR CELL BIOLOGY

All forms of diabetes, both inherited and acquired, are characterized by hyperglycemia, a relative or absolute lack of insulin, and the development of diabetes-specific microvascular pathology in the retina, renal glomerulus, and peripheral nerve. Diabetes is also associated with accelerated atherosclerotic macrovascular disease affecting arteries that supply the heart, brain, and lower extremities. Pathologically, this condition resembles macrovascular disease in nondiabetic patients but is more extensive and progresses more rapidly. As a consequence of its microvascular pathology, diabetes mellitus is now the leading cause of new blindness in people 20 to 74 years of age and the leading cause of end-stage renal disease (ESRD).

People with diabetes mellitus are the fastest growing group of renal dialysis and transplant recipients. The life expectancy of patients with diabetic end-stage renal failure is only 3 or 4 years. More than 60% of diabetic patients are affected by neuropathy, which includes distal symmetrical polyneuropathy.

Figure 31-1 Relative risks for the development of diabetic complications at different levels of mean hemoglobin A1c (HbA1c, glycated hemoglobin), obtained from the Diabetes Control and Complications Trial. (Adapted from Skyler J: Diabetic complications: the importance of glucose control. Endocrinol Metab Clin North Am 1996, 25:243254.)

Large, prospective clinical studies show a strong relationship between glycemia and diabetic microvascular complications in both type 1 and type 2 diabetes mellitus. There is a continuous, though not linear, relationship between level of glycemia and the risk of development and progression of these complications. Hyperglycemia and the dyslipidemia induced by insulin resistance both appear to play important roles in the pathogenesis of macrovascular complications.

Shared pathophysiologic features of microvascular complications

In the retina, glomerulus, and vasa nervorum, diabetes-specific microvascular disease is characterized by similar pathophysiologic features.

Requirement for Intracellular Hyperglycemia

Clinical and animal model data indicate that chronic hyperglycemia is the central initiating factor for all types of diabetic microvascular disease. Duration and magnitude of hyperglycemia are both strongly correlated with the extent and rate of progression of diabetic microvascular disease. In the Diabetes Control and Complications Trial (DCCT), for example, type 1 diabetic patients whose intensive insulin therapy resulted in hemoglobin A1c levels 2% lower than those receiving conventional insulin therapy had a 76% lower incidence of retinopathy, a 54% lower incidence of nephropathy, and a 60% reduction in neuropathy.

Although all diabetic cells are exposed to elevated levels of plasma glucose, hyperglycemic damage is limited to those cell types (e.g., endothelial cells) that develop intracellular hyperglycemia. Endothelial cells develop intracellular hyperglycemia because, unlike many other cells, they cannot down-regulate glucose transport when exposed to extracellular hyperglycemia. As illustrated in Figure 31-2, vascular smooth muscle cells, which are not damaged by hyperglycemia, show an inverse relationship between extracellular glucose concentration and subsequent rate of glucose transport measured as 2-deoxyglucose uptake. In contrast, vascular endothelial cells show no significant change in subsequent rate of glucose transport after exposure to elevated glucose concentrations. That intracellular hyperglycemia is necessary and sufficient for the development of diabetic pathology is further demonstrated by the fact that overexpression of the GLUT1 glucose transporter in mesangial cells cultured in a normal glucose milieu mimics the diabetic phenotype, inducing the same increases in collagen type IV, collagen type I, and fibronectin gene expression as diabetic hyperglycemia.
Abnormal Endothelial Cell Function

Early in the course of diabetes mellitus, before structural changes are evident, hyperglycemia causes abnormalities in blood flow and vascular permeability in the retina, glomerulus, and peripheral nerve vasa nervorum. The increase in blood flow and intracapillary pressure is thought to reflect hyperglycemia-induced decreased nitric oxide (NO) production on the efferent side of capillary beds, and possibly an increased sensitivity to angiotensin II. As a consequence of increased intracapillary pressure and endothelial cell dysfunction, retinal capillaries exhibit increased leakage of fluorescein and glomerular capillaries have an elevated albumin excretion rate (AER). Comparable changes occur in the vasa vasorum of peripheral nerve. Early in the course of diabetes, increased permeability is reversible; as time progresses, however, it becomes irreversible.

![Figure 31-2](image1.png)  
**Figure 31-2** Lack of down-regulation of glucose transport in cells affected by diabetic complications. Upper, 2-deoxyglucose (2DG) uptake in vascular smooth muscle cells preexposed to either 1.2, 5.5, or 22 mM glucose. Lower, 2DG uptake in bovine endothelial cells preexposed to either 1.2, 5.5, or 22 mM glucose. (From Kaiser N, Feener EP, Boukobza-Vardi N, et al. Differential regulation of glucose transport and transporters by glucose in vascular endothelial and smooth muscle cells. Diabetes 1993; 42:8089.)

![Figure 31-3](image2.png)  
**Figure 31-3** Overexpression of GLUT1 in mesangial cells cultured in normal glucose mimics the diabetic phenotype. Mesangial cells transfected with either LacZ (MCLacZ) or GLUT1 (MCGT1)-expressing constructs were cultured in 5-mM glucose, and the amount of the indicated matrix components secreted was determined. (From Heilig CW, Concepcion LA, Riser BL, et al. Overexpression of glucose transporters in rat mesangial cells cultured in a normal glucose milieu mimics the diabetic phenotype. J Clin Invest 1995; 96:18021814.)
Increased Vessel Wall Protein Accumulation

The common pathophysiologic feature of diabetic microvascular disease is progressive narrowing and eventual occlusion of vascular lumina, which results in inadequate perfusion and function of the affected tissues. Early hyperglycemia-induced microvascular hypertension and increased vascular permeability contribute to irreversible microvessel occlusion by three processes:

The first is an abnormal leakage of periodic acid-Schiff (PAS)-positive, carbohydrate-containing plasma proteins, which are deposited in the capillary wall and which may stimulate perivascular cells such as pericytes and mesangial cells to elaborate growth factors and extracellular matrix.

The second is extravasation of growth factors, such as transforming growth factor-β1 (TGF-β1), which directly stimulates overproduction of extracellular matrix components, and may induce apoptosis in certain complication-relevant cell types.

The third is hypertension-induced stimulation of pathologic gene expression by endothelial cells and supporting cells, which include glut-1 glucose transporters, growth factors, growth factor receptors, extracellular matrix components, and adhesion molecules that can activate circulating leukocytes. The observation that unilateral reduction in the severity of diabetic microvascular disease occurs on the side with ophthalmic or renal artery stenosis is consistent with this concept.
Microvascular Cell Loss and Vessel Occlusion

The progressive narrowing and occlusion of diabetic microvascular lumina are also accompanied by microvascular cell loss. In the retina, diabetes mellitus induces programmed cell death of Müller cells and ganglion cells, pericytes, and endothelial cells. In the glomerulus, declining renal function is associated with widespread capillary occlusion and podocyte loss, but the mechanisms underlying glomerular cell loss are not yet known. In the vasa nervorum, endothelial cell and pericyte degeneration occur, and these microvascular changes appear to precede the development of diabetic peripheral neuropathy. The multifocal distribution of axonal degeneration in diabetes supports a causal role for microvascular occlusion, but hyperglycemia-induced decreases in neurotrophins may contribute by preventing normal axonal repair and regeneration.

Figure 31-4 Development of retinopathy during posthyperglycemic normoglycemia (“hyperglycemic memory”). Quantitation of retinal microaneurysms and acellular capillaries in normal dogs, dogs with poor glycemic control for 5 years, dogs with good glycemic control for 5 years, dogs with poor glycemic control for 2.5 years (P G), and the same dogs after a subsequent 2.5 years of good glycemic control (P G). (Adapted from Engerman RL, Kern TS. Progression of incipient diabetic retinopathy during good glycemic control. Diabetes 1987; 36:808-812.)
Development of Microvascular Complications During Posthyperglycemic Euglycemia ("Hyperglycemic Memory")

Another common feature of diabetic microvascular disease has been termed hyperglycemic memory, or the persistence or progression of hyperglycemia-induced microvascular alterations during subsequent periods of normal glucose homeostasis. The most striking example of this phenomenon is the development of severe retinopathy in histologically normal eyes of diabetic dogs that occurred entirely during a 2.5-year period of normalized blood glucose that followed 2.5 years of hyperglycemia (Fig. 31-4). Normal dogs were compared to diabetic dogs with either poor control for 5 years, good control for 5 years, or poor control for 2.5 years (P G a) followed by good control for the next 2.5 years (P G b). Hb A1c values for both the good control group and the P G b group were identical to the normal group. Hyperglycemia-induced increases in selected matrix gene transcription also persist for weeks after restoration of normoglycemia in vivo, and a less pronounced, but qualitatively similar, prolongation of hyperglycemia-induced increase in selected matrix gene transcription occurs in cultured endothelial cells. 24

Data from the DCCT study suggest that hyperglycemic memory occurs in patients. In the secondary-intervention cohort, there was no difference in the incidence of sustained progression of retinopathy for the first 3 years, no difference in development of clinical albuminuria for 4 years, and no difference in the rate of change in creatinine clearance during the entire study. For neuropathy, the sural nerve sensory conduction velocity did not differ between the groups for 4 years, and intensive therapy did not slow the rate of decline of autonomic function at all. 25, 26, 27 Even more strikingly, the effects of former intensive and conventional therapy on the occurrence and severity of retinopathy and nephropathy were shown to persist for 4 years after the DCCT, despite nearly identical glycosylated hemoglobin values during the 4-year follow-up (8.2% versus 7.9%, respectively) (Fig. 31-5). 26 Together, these observations from animal and clinical studies imply that hyperglycemia induces prolonged and sometimes irreversible changes in long-lived intracellular molecules that persist and cause continued pathologic function in the absence of continued hyperglycemia.
Genetic Determinants of Susceptibility to Microvascular Complications

Clinicians have long observed that different patients with similar duration and degree of hyperglycemia differed markedly in their susceptibility to microvascular complications. Such observations suggested that genetic differences existed that affected the pathways by which hyperglycemia damaged microvascular cells. The leveling of risk of overt proteinuria after 30 years’ duration of type 1 diabetes at 27% is evidence that only a subset of patients are susceptible to development of diabetic nephropathy.\(^3\)

![Figure 31-6 Familial clustering of diabetic nephropathy. Prevalence of diabetic nephropathy in two studies of diabetic siblings of probands with or without diabetic nephropathy.](image)


A role for genetic determinant of susceptibility to diabetic nephropathy is most strongly supported by the demonstration of familial clustering of diabetic nephropathy. In two studies of families with two or more siblings having type 1 diabetes, if one diabetic sibling had advanced diabetic nephropathy, the other diabetic sibling had a nephropathy risk of 83% or 72%; in contrast, the risk was only 17% or 22% if the index patient did not have diabetic nephropathy (Fig. 31-6).\(^3\)\(^,\)\(^4\)\(^,\)\(^5\) For retinopathy, the DCCT reported familial clustering as well, with an odds ratio of 5.4 for the risk of severe retinopathy in diabetic relatives of positive versus negative subjects from the conventional treatment group.\(^6\)\(^,\)\(^7\)\(^,\)\(^8\)

Numerous associations have been made between various genetic polymorphisms and the risk of various diabetic complications. Examples include the 5’ insulin gene polymorphism,\(^9\) the G2m\(^23\) immunoglobulin allotype,\(^10\) angiotensin-converting enzyme (ACE) insertion/deletion polymorphisms,\(^11\)\(^12\) HLA-DQB1*0201/0302 alleles,\(^13\)\(^14\) polymorphisms of the aldose reductase gene,\(^15\)\(^16\) and a polymorphic CCTTT (n) repeat of NO synthetase (NOS) 2A.\(^17\) In all of these studies, there is no indication that the polymorphic gene actually plays a functional role rather than simply being in linkage disequilibrium with the locus encoding the unidentified relevant genes.
PATHOPHYSIOLOGIC FEATURES OF MACROVASCULAR COMPLICATIONS

Unlike microvascular disease, which occurs only in patients with diabetes mellitus, macrovascular disease resembles that in subjects without diabetes. However, subjects with diabetes have more rapidly progressive and extensive cardiovascular disease (CVD), with a greater incidence of multivessel disease and a greater number of diseased vessel segments than nondiabetic persons. Although dyslipidemia and hypertension occur with great frequency in type 2 diabetic populations, there is still excess risk in diabetic subjects after adjusting for these other risk factors. Diabetes itself may confer 75% to 80% of the excess risk of coronary disease in these diabetic subjects, and it enhances the deleterious effects of the other major cardiovascular risk factors.

In subjects with or without diabetes, atherosclerosis begins with endothelial dysfunction or injury. These endothelial changes or injury induce the secretion of chemokines such as monocyte chemoattractant protein 1 ( MCP-1 ), increase the expression of endothelial adhesion molecules for leucocytes and platelets, and enhance permeability to lipoproteins and other plasma components. As detailed in Chapter 34 , this leads to recruitment of monocyte-macrophages to the subendothelial space and to the infiltration of plasma LDL, which binds to arterial proteoglycan. The retained LDL then undergoes oxidation and is taken up by macrophages.

Activated macrophages and other leukocytes, as well as adherent aggregated platelets, stimulate smooth muscle cell proliferation and elaboration of extracellular matrix, culminating in the formation of a complex lesion filled with prothrombotic material contained by a fibrin cap. Rupture of this fibrin cap by matrix metalloproteinases causes thrombus formation and arterial occlusion. Because macrovascular disease also occurs in nondiabetic subjects, diabetes is thought to accelerate the process by increasing endothelial cell dysfunction and by exacerbating dyslipidemia.

The pathogenesis of endothelial cell dysfunction in diabetic arteries appears to involve both insulin resistance and hyperglycemia. In vitro studies suggest that insulin has both antiatherogenic and proatherogenic effects. One major antiatherogenic effect is the stimulation of endothelial NO production. NO released from endothelial cells is a potent inhibitor of platelet aggregation and adhesion to the vascular wall. Endothelial NO also controls the expression of genes involved in atherosclerosis. It decreases expression of the chemoktractant protein MCP-1, and of surface adhesion molecules such as CD11/CD18, P-selection, vascular cell adhesion molecule-1 ( VCAM-1 ), and intercellular adhesion molecule-1 ( ICAM-1 ). Endothelial cell NO also reduces vascular permeability and decreases the rate of oxidation of low-density lipoprotein (LDL) to its proatherogenic form.

Finally, endothelial cell NO inhibits proliferation of vascular smooth muscle cells. Two major proatherogenic effects of insulin are the potentiation of platelet-derived growth factor (PDGF)-induced vascular smooth muscle cell (VSMC) proliferation and the stimulation of VSMC plasminogen activator inhibitor 1 (PAI-1) production. Since insulin-induced NO production is mediated by the insulin receptor substrate P13 kinase signal transduction pathway, while the effects on smooth muscle cells are mediated by the ras Raf mek kin signal kinase transduction pathway, it has been proposed that pathway-selective insulin resistance in arterial cells may contribute to diabetic atherosclerosis. Recently, evidence of such selective vascular resistance to insulin has been demonstrated in the obese zucker rat.

Hyperglycemia also inhibits arterial NO production, both in vivo and in vitro. Similarly, hyperglycemia potentiates PDGF-induced VSMC proliferation and stimulates endothelial cell PAI-1 production. In addition, hyperglycemia has a variety of other proatherogenic effects on endothelial cells, platelets, and monocyte/macrophages. These include increased expression of MCP-1, up-regulation of adhesion molecules such as ICAM-1 and VCAM-1, potentiation of collagen-induced platelet activation, and increased secretion of collagen type IV and fibronectin.

Both insulin resistance and hyperglycemia have been implicated in the pathogenesis of diabetic dyslipidemia as well. Insulin resistance is associated with a characteristic lipoprotein profile that includes a high very-low-density lipoprotein (VLDL), a low high-density lipoprotein (HDL), and small, dense LDL. Both low HDL and small, dense LDL are each independent risk factors for macrovascular disease. This profile arises as a direct result of increased net free fatty acid (FFA) release by insulin resistant adipocytes and small, dense LDL are each independent risk factors for macrovascular disease. This profile arises as a direct result of increased net free fatty acid (FFA) release by insulin resistant adipocytes.

of cholesteryl ester transfer protein, excess VLDL transfers significant amounts of triglyceride to both HDL and LDL while depleting HDL and LDL of cholesteryl ester. The resultant triglyceride-enriched HDL carries less cholesteryl ester reverse cholesterol transport to the liver, and loss of ApoA1, from these particles reduces the total concentration of HDL available for reverse cholesterol transport. The triglyceride-enriched, cholesteryl ester depleted LDL is smaller and denser than normal LDL, allowing it to penetrate the vessel wall and be oxidized more easily.

Hyperglycemia appears to contribute to diabetic dyslipidemia by causing delayed clearance of postprandial lipoproteins, resulting in elevated levels of atherogenic cholesterol-enriched remnant particles. This remnant clearance defect is caused by a hyperglycemia-induced reduction in expression of the heparin sulfate proteoglycan perlecain on hepatocytes. Perlecain interaction with apoB48 containing lipoprotein remnant particles is necessary for efficient uptake by the LDL receptor-related protein.

The importance of hyperlipidemia in the pathogenesis of diabetic macrovascular disease in patients with type 2 diabetes is underscored by recent studies validating that effective treatment of hyperlipidemia in such patients substantially reduces their risk of CVD. The importance of hyperglycemia in the pathogenesis of diabetic macrovascular disease is suggested by the observation that carotid wall thickness is increased in persons with established diabetes but not in persons with impaired glucose tolerance.

The United Kingdom Prospective Diabetes Study (UKPDS) identified hyperglycemia as an important risk factor for macrovascular disease in type 2 diabetes, and numerous correlational studies show that hyperglycemia is a continuous risk factor for macrovascular disease. Similarly, glycated hemoglobin A1c is an independent risk factor for CVD in type 1 diabetes. The relative importance of hyperglycemia in type 1 patients is suggested by the 41% reduction in macrovascular...
disease \( (P = .06) \) observed in the intensive therapy group of the DCCT.\[\]
Four major hypotheses about how hyperglycemia causes diabetic complications have generated a large amount of data as well as several clinical trials based on specific inhibitors of these mechanisms. Until recently, there was no unifying hypothesis linking these four mechanisms together, nor was there an obvious connection between any of these mechanisms, each of which responds quickly to normalization of hyperglycemia, and the phenomenon of hyperglycemic memory (see earlier).

### Increased Polyol Pathway Flux

#### Aldose Reductase Function

Aldose reductase (alditol:NAD(P)+ 1-oxidoreductase, EC 1.1.1.21) is a cytosolic, monomeric oxidoreductase that catalyzes the NADPH-dependent reduction of a wide variety of carbohydrate compounds including glucose. Triphosphopyridine nucleotide, reduced form of NADP (NADPH) is the cofactor in both this reaction and in the regeneration of glutathione by glutathione reductase. Aldose reductase has a low affinity (high Michaelis constant \([K_m]\)) for glucose, and at the normal glucose concentrations found in nondiabetic patients, metabolism of glucose by this path pathway constitutes a small percentage of total glucose utilization. In a hyperglycemic environment, however, increased intracellular glucose results in increased enzymatic conversion to the polyalcohol sorbitol, with concomitant decreases in NADPH. In the polyol pathway, sorbitol is oxidized to fructose by the enzyme sorbitol dehydrogenase, with NAD+ reduced to NADH.

#### Biochemical Consequences of Increased Polyol Pathway Flux

A number of mechanisms have been proposed to explain the potential detrimental effects of hyperglycemia-induced increases in polyol pathway flux. These include sorbitol-induced osmotic stress, decreased Na+/K+ ATPase activity, increased cytosolic NADH/NAD+, and decreased cytosolic NADPH. Sorbitol does not diffuse easily across cell membranes, and it was originally suggested that this resulted in osmotic damage to microvascular cells. However, sorbitol concentrations measured in diabetic vessels and nerves are far too low to cause osmotic damage.

Another early suggestion was that increased flux through the polyol pathway decreased Na+/K+ ATPase activity. Although this was originally thought to be mediated by polyol-pathway-linked decreases in phosphatidylinositol synthesis, it has been shown to result from activation of protein kinase C (PKC) (see later).

Hyperglycemia-induced activation of PKC increases cytosolic phospholipase A₂ activity, which increases the production of two inhibitors of Na+/K+ ATPase, arachidonate and prostaglandin E₂ (PGE₂)\(^{1,2}\).

More recently, it has been proposed that oxidation of sorbitol by NAD+ increases the cytosolic ratio of NADH/NAD⁺, thereby inhibiting activity of the enzyme glyceraldehyde-3-phosphate dehydrogenase and increasing concentrations of triose phosphate. \(^{1,3}\) Elevated triose phosphate concentrations could increase formation of both methylglyoxal, a precursor of advanced glycation end products (AGEs), and dialglycerol (DAG) (via -glycerol-3-phosphate), thus activating PKC (discussed in subsequent sections). Although increased NADH production is supported by the observation that hyperglycemia increases both lactate concentration and the lactate/pyruvate ratio, there is no direct evidence that the concentrations of NADH and NAD+ , as opposed to NADH and NAD⁺ flux, are altered. In endothelial cells, where aldose reductase activity is low, increased NADH production may also reflect hyperglycemia-induced increased flux through glycolysis \(^{1,2}\) and through the glucuronic acid pathway. \(^{1,2}\)

Other evidence presented in support of this hypothesis includes the observation that administration of pyruvate can prevent diabetes-related endothelial dysfunction in some systems. However, the observed effects of pyruvate on microvascular function may reflect its potent antioxidant properties rather than effects on the NADH/NAD⁺ ratio, because reactive oxygen species (ROS) also partially inhibit glyceraldehyde-3-phosphate dehydrogenase and increase glyceraldehyde-3-phosphate levels. \(^{1,2}\) The source of hyperglycemia-induced ROS is discussed later in this section.

It has also been proposed that reduction of glucose to sorbitol by NADPH consumes the cofactor NADPH. Because NADPH is required for regenerating reduced glutathione (GS), this could induce or exacerbate intracellular oxidative stress. Less reduced glutathione has in fact been found in the lens of transgenic mice that overexpress aldose reductase, and this is the most likely mechanism by which increased flux through the polyol pathway has deleterious consequences. \(^{1,2}\)

#### Hyperglycemia-induced inhibition of glucose-6-phosphate dehydrogenase, the major source of NADPH regeneration, may further reduce NADPH concentration in some vascular cells or neuronal cells. \(^{1,2}\)
Advanced Glycation End-Product Inhibitors

The hydrazine compound aminoguanidine was the first AGE inhibitor discovered, and its effect on diabetic pathology has been investigated in the retina, kidney, nerve, and artery. In the rat retina, diabetes causes a 19-fold increase in the number of acellular capillaries. Aminoguanidine treatment of diabetics prevented excess AGE accumulation and reduced the number of acellular capillaries by 80%. Diabetes-induced pericyte dropout also was markedly reduced by aminoguanidine treatments.

Similar results have been obtained in animal models of diabetic kidney disease. Diabetes increased AGEs in the renal glomerulus, and aminoguanidine treatment prevented this diabetes-induced increase. Untreated diabetic animals developed albuminuria that averaged 30 mg every 24 hours for 32 weeks. This was more than a 10-fold increase above control levels. Untreated diabetic animals also developed the characteristic structural feature of human diabetic nephropathy (i.e., increased fractional mesangial volume). When diabetic animals were treated with aminoguanidine, the increase was completely prevented. A structurally unrelated AGE inhibitor, OPB-9195, also prevented the development and progression of experimental diabetic nephropathy by blocking type IV collagen overproduction and normalizing the expression of TGF-.

In the peripheral nerve of diabetic rats, both motor nerve and sensory NCVs are decreased after 8 weeks of diabetes. Nerve action potential amplitude is decreased by 37% and peripheral nerve blood flow is decreased by 57% after 24 weeks of diabetes. Aminoguanidine treatment prevented each of these abnormalities of diabetic peripheral nerve function.

In a large randomized, double-blind, placebo-controlled, multicenter trial of aminoguanidine in type 1 diabetic patients with overt nephropathy, aminoguanidine lowered total urinary protein and slowed progression of nephropathy, over and above the effects of existing optimal care. In addition, aminoguanidine reduced the progression of diabetic retinopathy (defined as an increase by three or more steps in the Early Treatment Diabetic Retinopathy Study [ETDRS] scale).
Activation of Protein Kinase C

Mechanism of Hyperglycemia-Induced Protein Kinase C Activation

PKCs are a family of at least 11 isoforms, 9 of which are activated by the lipid second-messenger DAG. Intracellular hyperglycemia increases DAG content in cultured microvascular cells and in the retina and renal glomeruli of diabetic animals. Increased de novo synthesis of DAG activates PKC both in cultured vascular cells and in retina and glomeruli of diabetic animals. Increased DAG primarily activates the and isoforms of PKC, but increases in other isoforms have also been found, such as PKC- and PKC-epsilon isoforms in the retina and PKC- and PKC- in the glomerulus of diabetic rats.

Consequences of Hyperglycemia-Induced Protein Kinase C Activation

In early experimental diabetes, activation of PKC- isoforms has been shown to mediate retinal and renal blood flow abnormalities, perhaps by depressing NO production and increasing endothelin-1 activity. Abnormal activation of PKC has been implicated in the decreased glomerular production of NO induced by experimental diabetes and in the decreased smooth muscle cell NO production induced by hyperglycemia. PKC activation also inhibits insulin-stimulated expression of endothelial nitric oxide synthase (eNOS) messenger RNA (mRNA) in cultured endothelial cells. Hyperglycemia increases endothelin 1-stimulated mitogen-activated protein kinase activity in glomerular mesangial cells by activating PKC isoforms. The increased endothelial cell permeability induced by high glucose in cultured cells is mediated by activation of PKC,- however. Activation of PKC by elevated glucose levels also induces expression of the permeability-enhancing factor VEGF in smooth muscle cells.

In addition to affecting hyperglycemia-induced abnormalities of blood flow and permeability, activation of PKC contributes to increased microvascular matrix protein accumulation by inducing the expression of TGF-, , fibronectin, and 1 (IV) collagen in both cultured mesangial cells and in the glomeruli of diabetic rats. This effect appears to be mediated through PKC's inhibition of NO production. Hyperglycemia-induced expression of laminin C1 in cultured mesangial cells is independent of PKC activation, however. Hyperglycemia-induced activation of PKC has also been implicated in the overexpression of the fibrinolytic inhibitor PAI-1 and in the activation of the pleiotropic transcription factor NF- in cultured endothelial cells and vascular smooth muscle cells.
Increased Hexosamine Pathway Flux

A fourth hypothesis about how hyperglycemia causes diabetic complications has recently been formulated, in which glucose is shunted into the hexosamine pathway (Fig. 31-14). In this pathway, fructose-6-phosphate is diverted from glycolysis to provide substrates for reactions that require UDP-N-acetylglucosamine, such as proteoglycan synthesis and the formation of O-linked glycoproteins. Inhibition of the rate-limiting enzyme in the conversion of glucose to glucosamine, glutamine:fructose-6-phosphate amidotransferase, blocks hyperglycemia-induced increases in the transcription of both TGF-β and TGF-β1. This pathway has previously been shown to play an important role in hyperglycemia-induced and fat-induced insulin resistance.

![Figure 31-14](image)

Figure 31-14 Schematic representation of the hexosamine pathway. The glycolytic intermediate fructose-6-phosphate (Fruc-6-P) is converted to glucosamine-6-phosphate (Glc-6-P) by the enzyme glutamine:fructose-6-phosphate amidotransferase (GFAT). Increased donation of N-acetylglucosamine moieties to serine and threonine residues of transcription factors such as Sp1 increases production of such complication-promoting factors as PAI-1 and TGF-β. See text for additional abbreviations. (Adapted from Du XL, Edelstein D, Rossetti L, et al. Hyperglycemia-induced mitochondrial superoxide overproduction activates the hexosamine pathway and induces plasminogen activator inhibitor-1 expression by increasing Sp1 glycosylation. Proc Natl Acad Sci USA 2000; 97:12222-12226.)

The mechanism by which increased flux through the hexosamine pathway mediates hyperglycemia-induced increases in gene transcription has not been clear, but the observation that Sp1 sites regulate hyperglycemia-induced activation of the PAI-1 promoter in vascular smooth muscle cells suggests that covalent modification of Sp1 by N-acetylglucosamine may explain the link between hyperglycemia-induced changes in gene transcription and hexosamine pathway activation. Virtually every RNA polymerase II transcription factor examined has been found to be O-GlcNacylated, and the glycosylated form of Sp1 appears to be more transcriptionally active than the deglycosylated form of the protein. A four-fold increase in Sp1 O-GlcNacylation caused by inhibition of the enzyme O-GlcNac-N-acetylglucosaminidase resulted in a reciprocal 30% decrease in its level of serine/threonine phosphorylation, supporting the concept that O-GlcNacylation and phosphorylation compete to modify the same sites on this protein.

GlcNac modification of Sp1 may regulate other glucose-responsive genes in addition to TGF-β and PAI-1. Glucose-responsive transcription is regulated by Sp1 sites in the acetyl-CoA carboxylase gene, the rate-limiting enzyme for fatty acid synthesis, for example, and it appears that post-translational modification of Sp1 is responsible for this effect. Because virtually every RNA polymerase II transcription factor examined has been found to be O-GlcNacylated, it is possible that reciprocal modification by O-GlcNacylation and phosphorylation of transcription factors other than Sp1 may function as a more generalized mechanism for regulating glucose-responsive gene transcription.

In addition to transcription factors, many other nuclear and cytoplasmic proteins are dynamically modified by O-GlcNAc moieties and may exhibit reciprocal modification by phosphorylation in a manner analogous to Sp1. Thus, activation of the hexosamine pathway by hyperglycemia may result in many changes in both gene expression and in protein function that together contribute to the pathogenesis of diabetic complications.
DIFFERENT PATHOGENIC MECHANISMS REFLECT A SINGLE HYPERGLYCEMIA-INDUCED PROCESS

Although specific inhibitors of aldose reductase activity, AGE formation, and PKC activation each ameliorate various diabetes-induced abnormalities in animal models, there has been no apparent common element linking the four mechanisms of hyperglycemia-induced damage discussed in the preceding section. It has also been conceptually difficult to explain the phenomenon of hyperglycemic memory (discussed in an earlier section) as a consequence of four processes that quickly normalize when euglycemia is restored. These issues have now been resolved by the recent discovery that each of the four different pathogenic mechanisms reflects a single hyperglycemia-induced process: overproduction of superoxide by the mitochondrial electron transport chain.

Hyperglycemia increases ROS production inside cultured bovine aortic endothelial cells. To understand how this occurs, a brief overview of glucose metabolism is helpful. Intracellular glucose oxidation begins with glycolysis in the cytoplasm, which generates NADH and pyruvate. Cytosolic NADH can donate reducing equivalents to the mitochondrial electron transport chain via two shuttle systems, or it can reduce pyruvate to lactate, which exits the cell to provide substrate for hepatic gluconeogenesis. Pyruvate can also be transported into the mitochondria, where it is oxidized by the tricarboxylic acid (TCA) cycle to produce carbon dioxide ($\text{CO}_2$), water ($\text{H}_2\text{O}$), four molecules of NADH, and one molecule of FADH$_2$. Mitochondrial NADH and FADH$_2$ provide energy for adenosine triphosphate (ATP) production via oxidative phosphorylation by the electron transport chain. Electron flow through the mitochondrial electron transport chain is carried out by four inner membrane-associated enzyme complexes, plus cytochrome-c and the mobile carrier ubiquinone. NADH derived from both cytosolic glucose oxidation and mitochondrial TCA cycle activity donates electrons to NADH:ubiquinone oxidoreductase (Complex I). Complex I ultimately transfers its electrons to ubiquinone. Ubiquinone can also be reduced by electrons donated from several FADH$_2$-containing dehydrogenases, including succinate:ubiquinone oxidoreductase (Complex II) and glycerol-3-phosphate dehydrogenase. Electrons from reduced ubiquinone are then transferred to cytochrome-c oxidoreductase (Complex III) by the ubiquinol-cytochrome-c oxidoreductase radical-generating Q cycle. Electron transport then proceeds through cytochrome-c, cytochrome-c oxidase (Complex IV), and finally, molecular oxygen.

Electron transfer through Complexes I, III, and IV generates a proton gradient that drives ATP synthase (Complex V). When the electrochemical potential difference generated by this proton gradient is high, the life of superoxide-generating electron transport intermediates such as ubiquinol is prolonged. There appears to be a threshold value above which superoxide production is markedly increased (Fig. 31-15).

Using inhibitors of both the shuttle that transfers cytosolic NADH into mitochondria, and the transporter that transfers cytochrome c pyruvate into the mitochondria, the TCA cycle was shown to be the source of hyperglycemia-induced ROS in endothelial cells. Overexpression of uncoupling protein 1 (UCP-1), a specific protein uncoupler of oxidative phosphorylation capable of collapsing the proton electrochemical gradient, also prevented the effect of hyperglycemia. These results demonstrate that hyperglycemia-induced intracellular ROS are produced by the proton translocating Q-cycle carrier generated by the mitochondrial electron transport chain. Overexpression of manganese superoxide dismutase, the mitochondrial form of this antioxidant enzyme, also prevented the effect of hyperglycemia. This result demonstrates that superoxide is the reactive oxygen radical produced by this mechanism.

The effect of hyperglycemia-induced mitochondrial superoxide overproduction on polyol pathway flux was evaluated after first determining that sorbitol in these cells was exclusively derived from aldose reductase activity. Sorbitol levels were 2.6-fold higher than baseline (5-mM glucose) when endothelial cells were incubated in 30-mM glucose (Fig. 31-16). Hyperglycemia-induced sorbitol accumulation was completely prevented by UCP-1 and superoxide dismutase (Mn-SOD) (see Fig. 31-16), indicating that mitochondrial superoxide overproduction stimulates aldose reductase activity. This effect appears to reflect the well-described reversible inhibition of glyceraldehyde-3-phosphate dehydrogenase by ROS, which increases glyceraldehyde-3-phosphate levels and the levels of proximal glycolytic metabolites, including glucose (Fig. 31-17).

Next, the effect of hyperglycemia-induced mitochondrial superoxide overproduction on intracellular AGE formation was determined. In bovine aortic endothelial cells, hyperglycemia increases intracellular AGEs primarily, if not exclusively, by increasing the formation of AGE-forming methylglyoxal. Therefore, the effect of UCP-1 and Mn-SOD on hyperglycemia-induced formation of intracellular methylglyoxal-derived AGEs was examined (see Fig. 31-16). Each of these agents completely prevented hyperglycemia-induced formation of intracellular AGEs (see Fig. 31-16), indicating that mitochondrial superoxide initiates intracellular AGE formation. Because methylglyoxal is formed by fragmentation of glyceraldehyde-3-phosphate, this dependency on increased mitochondrial superoxide production also likely reflects increased glyceraldehyde-3-phosphate levels due to inhibition of glyceraldehyde-3-phosphate dehydrogenase by ROS (see Fig. 31-17).

The effect of UCP-1 and Mn-SOD on hyperglycemia-induced activation of PKC was also evaluated (see Fig. 31-16). Each of these agents completely inhibited PKC activation, suggesting that mitochondrial superoxide overproduction initiates the hyperglycemia-induced de novo synthesis of DAG that activates PKC. Most likely this too reflects increased glyceraldehyde-3-phosphate levels due to inhibition of glyceraldehyde-3-phosphate dehydrogenase by ROS (see Fig. 31-17).

Finally, the effect of hyperglycemia-induced mitochondrial superoxide overproduction on the hexosamine pathway was determined. Hyperglycemia induced an increase in hexosamine pathway activity that was completely prevented by UCP-1, Mn-SOD, and azaserine, an inhibitor of the rate-limiting enzyme in the hexosamine pathway.

Hyperglycemia-induced activation of the redox-sensitive pleiotropic transcription factor NF-B was also prevented by inhibition of mitochondrial superoxide overproduction.
A POSSIBLE MOLECULAR BASIS FOR HYPERGLYCEMIC MEMORY

In contrast to the four known hyperglycemia-inducible abnormalities of intracellular metabolism, hyperglycemia-induced mitochondrial superoxide production may provide an explanation for the development of complications during posthyperglycemic normoglycemia (hyperglycemic memory). Hyperglycemia-induced increases in superoxide would not only increase aldose reductase activity, AGE formation, PKC activity, and hexosamine pathway activity but may also induce mutations in mitochondrial DNA (mtDNA), Mitochondria are more vulnerable to mutation because mtDNA contains virtually no introns, lacks protective histones, and has no effective DNA repair mechanism. Defective electron transport complex subunits encoded by mutated mtDNA would eventually cause increased superoxide production at physiologic concentrations of glucose, with resultant continued activation of the four pathways despite the absence of hyperglycemia.

Figure 31-16 Effect of agents that alter mitochondrial electron transport chain function on the three main pathways of hyperglycemic damage. A, Hyperglycemia-induced protein kinase C (PKC) activation. B, Intracellular advanced glycation end-product (AGE) formation. C, Sorbitol accumulation. Cells were incubated in 5-mM glucose, 30-mM glucose alone, and 30-mM glucose plus either agents that uncouple oxidative phosphorylation and reduce the high mitochondrial membrane potential (TTFA, CCCP, UCP-1), or dismutate superoxide (Mn-SOD). See text for additional abbreviations. (From Nishikawa T, Edelstein D, Du XL, et al. Normalizing mitochondrial superoxide production blocks three pathways of hyperglycaemic damage. Nature 2000; 404:787-790.)
PROSPECTS FOR PHARMACOLOGIC INTERVENTION

Aldose Reductase Inhibitors

In vivo studies of polyol pathway inhibition have yielded promising results with neuropathy but disappointing results in other target tissues of diabetic complications. During the course of a 5-year study, nerve conduction velocity (NCV) progressively decreased in untreated diabetic dogs, whereas this decrease was prevented by treatment with an aldose reductase inhibitor (ARI). Positive effects of ARIs on human diabetic neuropathy have been reported. In contrast, aldose reductase inhibition failed to prevent retinopathy in the 5-year study in dogs, nor did it prevent capillary basement membrane thickening in the retina, kidney, or muscles. A 3-year human trial also failed to show any effect on diabetic retinopathy.
Advanced Glycation End-Product Inhibitors

The hydrazine compound aminoguanidine was the first AGE inhibitor discovered, and its effect on diabetic pathology has been investigated in the retina, kidney, nerve, and artery. In the rat retina, diabetes causes a 19-fold increase in the number of acellular capillaries. Aminoguanidine treatment of diabetics prevented excess AGE accumulation and reduced the number of acellular capillaries by 80%. Diabetes-induced pericyte dropout also was markedly reduced by aminoguanidine treatments.

Similar results have been obtained in animal models of diabetic kidney disease. Diabetes increased AGEs in the renal glomerulus, and aminoguanidine treatment prevented this diabetes-induced increase. Untreated diabetic animals developed albuminuria that averaged 30 mg every 24 hours for 32 weeks. This was more than a 10-fold increase above control levels. Untreated diabetic animals also developed the characteristic structural feature of human diabetic nephropathy (i.e., increased fractional mesangial volume). When diabetic animals were treated with aminoguanidine, the increase was completely prevented. A structurally unrelated AGE inhibitor, OPB-9195, also prevented the development and progression of experimental diabetic nephropathy by blocking type IV collagen overproduction and normalizing the expression of TGF-β.

In the peripheral nerve of diabetic rats, both motor nerve and sensory NCVs are decreased after 8 weeks of diabetes. Nerve action potential amplitude is decreased by 37% and peripheral nerve blood flow is decreased by 57% after 24 weeks of diabetes. Aminoguanidine treatment prevented each of these abnormalities of diabetic peripheral nerve function.

In a large randomized, double-blind, placebo-controlled, multicenter trial of aminoguanidine in type 1 diabetic patients with overt nephropathy, aminoguanidine lowered total urinary protein and slowed progression of nephropathy, over and above the effects of existing optimal care. In addition, aminoguanidine reduced the progression of diabetic retinopathy (defined as an increase by three or more steps in the Early Treatment Diabetic Retinopathy Study [ETDRS] scale).
Protein Kinase C Inhibitors

The recent development of a isoform-specific PKC inhibitor has allowed in vivo studies to go forward, because the toxicity of non-selective PKC inhibitors precludes their use. LY333531 inhibits PKC-α and PKC-δ with a half-maximal inhibitory constant (IC₅₀) that is at least 50-fold less than for other PKC isoforms. Treatment with LY333531 significantly reduced PKC activity in the retina and renal glomeruli of diabetic animals. Concomitantly, LY333531 treatment significantly reduced diabetes-induced increases in retinal mean circulation time, normalized diabetes-induced increases in glomerular filtration rate (GFR), and partially corrected urinary AER. Treatment of db/db mice with LY333531 for a longer period also ameliorated accelerated glomerular mesangial expansion. Clinical trials of LY333531 in human diabetic patients are currently in progress.
Future Drug Targets

The recent discovery that each of the four different pathogenic mechanisms discussed in this section reflect a single hyperglycemia-induced process suggests that interrupting the overproduction of superoxide by the mitochondrial electron transport chain would normalize polyol pathway flux, AGE formation, PKC activation, hexosamine pathway flux, and NF-κB activation. Novel compounds that act as superoxide dismutase/catalase mimetics already exist, and these compounds have been shown to normalize hyperglycemia-induced mitochondrial superoxide overproduction. These and the other agents described in this section may have unique clinical efficacy in preventing the development and progression of diabetic complications.
RETNOPATHY, MACULAR EDEMA, AND OTHER OCULAR COMPLICATIONS

Diabetic retinopathy is a well-characterized, sight-threatening, chronic microvascular complication that eventually affects virtually all patients with diabetes mellitus. Diabetic retinopathy is characterized by gradually progressive alterations in the retinal microvasculature, leading to areas of retinal nonperfusion, increased vasopermeability, and pathologic intracellular proliferation of retinal vessels. The complications associated with the increased vasopermeability, termed macular edema, and uncontrolled neovascularization, termed proliferative diabetic retinopathy (PDR), can result in severe and permanent visual loss. Despite decades of research, there is presently no known means of preventing diabetic retinopathy and, despite effective therapies, diabetic retinopathy remains the leading cause of new-onset blindness in working-aged Americans. With appropriate medical and ophthalmologic care, however, more than 90% of visual loss resulting from diabetic retinopathy can be prevented.

Thus, until a cure for diabetes is discovered, the primary clinical care emphasis for the prevention of vision loss is appropriately directed at the early identification, accurate classification, and timely treatment of retinopathy. Emphasis must also be placed on ensuring compliant life-long routine ophthalmologic follow-up of the diabetic patient and optimization of associated systemic disorders.

EPIDEMIOLOGY AND IMPACT

Sixteen million Americans have diabetes mellitus, but only half are aware that they have the disease. Diabetic retinopathy is the leading cause of new cases of legal blindness among Americans between the ages of 20 and 74 years. There is a higher risk of more frequent and severe ocular complications in type 1 diabetes. Approximately 25% of patients with type 1 diabetes have retinopathy after 5 years, with this figure increasing to 60% and 80% after 10 and 15 years, respectively. However, because there are more adult-onset cases than juvenile-onset cases, type 2 disease accounts for a higher proportion of patients with visual loss. The most threatening form of retinopathy (PDR) is present in approximately 25% of type 1 patients with diabetes of 15 years' duration. An estimated 700,000 persons have PDR, 130,000 with high-risk PDR, 500,000 with macular edema, and 325,000 with clinically significant macular edema (CSME) in the United States. An estimated 63,000 cases of PDR, 29,000 high-risk PDR, 80,000 macular edema, 56,000 CSME, and 5000 new cases of legal blindness occur each year as a result of diabetic retinopathy. Blindness has been estimated to be 25 times more common in persons with diabetes than in those without the disease.

Estimates of the medical and economic impact of retinopathy-associated morbidity have been performed using computer simulations that incorporate clinical trial and cost reimbursement data to model effects of applying accepted evaluation and treatment techniques to patients with type 1 and type 2 diabetic retinopathy. The models predict that in the absence of good glycemic control, 72% of patients with type 1 diabetes will develop PDR requiring panretinal photocoagulation (PRP) over their lifetime and that 42% will develop macular edema. If patients with type 1 diabetes receive currently suggested treatment, there is a predicted cost of $966 per person-year of vision saved from PDR and $1120 per person-year of central acuity saved from macular edema as of 1990. Indeed, current estimates are that only 60% of patients in need of retinopathy treatment are receiving appropriate ophthalmic care. If all patients with both type 1 and type 2 diabetes were to receive care according to currently suggested guidelines, annual savings of $624 million and 173,540 person-years of sight would be realized.

The DCCT showed that both the rate of development of any retinopathy as well as the rate of retinopathy progression once it was present were significantly reduced after 3 years of intensive insulin therapy, an effect maintained even 4 years after conclusion of the study. Applying DCCT intensive insulin therapy to all persons in the United States with insulin-dependent diabetes mellitus would result in a gain of 920,000 person-years of sight, although the costs of intensive therapy are three times that of conventional therapy.
PATHOPHYSIOLOGY

A detailed discussion of the pathophysiologic mechanisms underlying diabetic retinopathy and other diabetes-related complications has been presented earlier in this chapter. The earliest histologic effects of diabetes mellitus in the eye include loss of retinal vascular pericytes (supporting cells for retinal endothelial cells), thickening of vascular endothelium basement membrane, and alterations in retinal blood flow. With increasing loss of retinal pericytes, the retinal vessel wall develops outpouchings (microaneurysms) and becomes fragile.

Clinically, microaneurysms and small retinal hemorrhages may not always be readily distinguishable and are evaluated together as "hemorrhages and microaneurysms" (Fig. 31-19A) (Figure Not Available) (see also Color Plate). Rheologic changes occur in diabetic retinopathy resulting from increased platelet aggregation, integrin-mediated leukocyte adhesion, and endothelial damage. Disruption of the blood retinal barrier may ensue, characterized by increased vascular permeability. The subsequent leakage of blood and serum from the retinal vessels results in retinal hemorrhages, retinal edema, and hard exudates (Fig. 31-19A and C) (Figure Not Available). Moderate visual loss follows if the fovea is affected by the leakage.

With time, increasing sclerosis and endothelial cell loss lead to narrowing of the retinal vessels, which decreases vascular perfusion and may ultimately lead to obliteration of the capillaries and small vessels (Fig. 31-19 B) (Figure Not Available). The resulting retinal ischemia is a potent inducer of angiogenic growth factors. Several angiogenic growth factors have been isolated from eyes with diabetic retinopathy, including IGFs, bFGF, hepatocyte growth factor (HGF), and VEGF. These factors promote the development of new vessel growth and retinal vascular permeability. Indeed, inhibition of molecules such as VEGF and their signaling pathways can suppress the development of retinal neovascularization and retinal vascular permeability.

New vessels tend to grow in regions of strong vitreous adhesion to the retina, such as at the optic disc and major vascular arcades (Fig. 31-19 D and E) (Figure Not Available). The posterior vitreous face also serves as a scaffold for pathologic neovascularization, and the new vessels commonly arise at the junctions between perfused and nonperfused retina. When the retina is severely ischemic, the concentration of angiogenic growth factors may reach sufficient concentration in the anterior chamber to cause abnormal new vessel proliferation on the iris and the anterior chamber angle. Uncontrolled anterior segment neovascularization may result in rubeotic glaucoma because the fibrovascular proliferation in the angle of the eye causes blockage of aqueous outflow through the trabecular meshwork.

Proliferating new vessels in diabetic retinopathy have a tendency to bleed, which results in preretinal and vitreous hemorrhages (VHs) (Fig. 31-19 E and F) (Figure Not Available). Although the presence of a large amount of blood in the preretinal space or vitreous cavity per se is not damaging to the retina, these intraocular hemorrhages often cause prolonged visual loss by blocking the visual axis. Membranes on the retinal surface can be induced by blood and result in wrinkling and traction on the retina. Although all retinal neovascularization eventually becomes quiescent, as with most scarring processes there is progressive fibrosis of the new vessel complexes that is associated with contraction. However, in the eye, such forces may exert traction on the retina, leading to tractional retinal detachment and retinal tears that may result in severe and permanent visual loss if left untreated (Fig. 31-19 G and H) (Figure Not Available).

In short, causes of visual loss from complications of diabetes mellitus include retinal ischemia involving the fovea, macular edema or near the fovea, preretinal or vitreous hemorrhages, retinal detachment, and neovascular glaucoma. Visual loss may also result from more indirect effects of disease progression in diabetic patients, such as retinal vessel occlusion, accelerated atherosclerotic disease, and embolic phenomena.

Figure 31-18 Diabetic retinopathy pathogenesis flow chart. The schematic flow chart represents the major preclinical and clinical findings associated with the full spectrum of diabetic retinopathy and macular edema. VEGF, vascular endothelial growth factor.
CLINICAL FEATURES

Risk Factors

Duration of diabetes is closely associated with the onset and severity of diabetic retinopathy. Diabetic retinopathy is rare in prepubescent patients with type 1 diabetes, but nearly all patients with type 1 diabetes and more than 60% of patients with type 2 diabetes develop some degree of retinopathy after 20 years. In patients with type 2 diabetes, approximately 20% have retinopathy at the time of diabetes diagnosis and most have some degree of retinopathy over subsequent decades.

Diabetic retinopathy is the most frequent cause of new-onset blindness among American adults aged 20 to 74 years. In the Wisconsin Epidemiologic Study of Diabetic Retinopathy, approximately 4% of patients younger than 30 years of age at diagnosis and nearly 2% of patients older than 30 years of age at diagnosis were legally blind. In the younger-onset group, 86% of blindness was attributable to diabetic retinopathy. In the older-onset group, where other eye diseases were also common, 33% of the cases of legal blindness were due to diabetic retinopathy.

Lack of glycemic control is another significant risk factor for the onset and progression of diabetic retinopathy. The DCCT demonstrated a clear relationship between hyperglycemia and diabetic microvascular complications, including retinopathy in 1441 patients with type 1 diabetes. In patients monitored 4 to 9 years, the DCCT showed that intensive insulin therapy reduced or prevented the development of retinopathy by 27% as compared with conventional therapy. Additionally, intensive insulin therapy reduced the progression of diabetic retinopathy by 34% to 76% and had a substantial beneficial effect over the entire range of retinopathy severity. This improvement was achieved with an average 10% reduction in Hb A1c from 8% to 7.2%. These results underscore that although intensive therapy does not prevent retinopathy completely, it reduces the risk of retinopathy onset and progression.

Renal disease, as manifested by microalbuminuria and proteinuria, is yet another significant risk factor for onset and progression of diabetic retinopathy. Hypertension is associated with PDR and is an established risk factor for the development of macular edema. Additionally, elevated serum lipid levels are associated with extravasated lipid in the retina (hard exudates) and visual loss.
Clinical Findings

Clinical findings associated with early and progressing diabetic retinopathy include hemorrhages or microaneurysms (H/Ma), cotton-wool spots (CWSs), hard exudates, intraretinal microvascular abnormalities (IRMAs), and venous caliber abnormalities (VCABs), such as venous loops, venous tortuosity, and venous beading (see Fig. 31-19A and C) (Figure Not Available). Microaneurysms are saccular outpoucings of the capillary walls that can leak fluid and result in intraretinal hemorrhages. The intraretinal hemorrhages can be "flame-shaped" or "dot/blot" like in appearance, reflecting the architecture of the layer of the retina in which they occur. IRMAs are either new vessel growth within the retinal tissue itself or shunt vessels through areas of poor vascular perfusion. It is common for IRMAs to be adjacent to CWSs, which are caused by microinfarcts in the nerve fiber layer. VCABs are a sign of severe retinal hypoxia. In some cases of extensive vascular loss, however, the retina may appear free of nonproliferative lesions. Such areas are termed "featureless retina" and are a sign of severe retinal hypoxia.

Vision loss from diabetic retinopathy generally results from persistent, nonclearing vitreous hemorrhage, traction retinal detachment, or diabetic macular edema (DME) (see Fig. 31-18 and Fig. 31-19 (Figure Not Available)). Neovascularization with fibrous tissue contraction can distort the retina and lead to traction retinal detachment. The new vessels may bleed, causing preretinal or vitreous hemorrhage. The most common cause of vision loss from diabetes, however, is macular disease and macular edema. Macular edema is more likely to occur in patients with type 2 diabetes, which represents 90% of the diabetic population. In diabetic macular disease, macular edema involving the fovea or nonperfusion of the capillaries in the central macula is responsible for the loss of vision.

<table>
<thead>
<tr>
<th>TABLE 31-1 – Glossary &amp; Abbreviations Pertinent to Diabetic Eye Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Background Diabetic Retinopathy (BDR):</strong> An outdated term referring to some stages of nonproliferative diabetic retinopathy. Because this terminology is not closely associated with disease progression, it has been replaced by the various levels of nonproliferative diabetic retinopathy.</td>
</tr>
<tr>
<td><strong>Cotton Wool Spot:</strong> A gray or white area lesion in the nerve fiber layer of the retina resulting from stasis of axoplasmic flow as a result of infarction.</td>
</tr>
<tr>
<td><strong>Diabetes Control &amp; Complications Trial (DCCT):</strong> A multicenter randomized, clinical trial designed to address whether intensive insulin therapy could prevent or slow the progression of systemic complications of diabetes mellitus.</td>
</tr>
<tr>
<td><strong>Diabetic Retinopathy (DR):</strong> Retinal pathology related to the underlying systemic disease of diabetes mellitus.</td>
</tr>
<tr>
<td><strong>Diabetic Retinopathy Study (DRS):</strong> The first multicenter randomized clinical trial to demonstrate the value of laser scatter (panretinal) photocoagulation in reducing the risk of visual loss among patients with all levels of diabetic retinopathy.</td>
</tr>
<tr>
<td><strong>Diabetic Retinopathy Vitrectomy Study (DRVS):</strong> A multicenter clinical trial demonstrating the value of early vitrectomy for patients with very advanced diabetic retinopathy.</td>
</tr>
<tr>
<td><strong>Early Treatment Diabetic Retinopathy Study (ETDRS):</strong> A multicenter randomized clinical trial that addressed at what stage of retinopathy scatter (panretinal) photocoagulation was indicated, whether focal photocoagulation was effective for preventing moderate visual loss from clinically significant macular edema, and whether aspirin therapy altered the risks for outcome or treatment of diabetic retinopathy.</td>
</tr>
<tr>
<td><strong>Focal Laser Photocoagulation:</strong> A type of laser treatment used for patients with clinically significant macular edema whose main goal is to reduce vascular leakage either by focal treatment of leaking retinal microaneurysms or by application of therapy in a grid-like pattern.</td>
</tr>
<tr>
<td><strong>Hard Exudate:</strong> Lipid accumulation within the retina as a result of increased vasopermeability.</td>
</tr>
<tr>
<td><strong>High-Risk-Characteristic Proliferative Diabetic Retinopathy (HRC-PDR):</strong> Proliferative diabetic retinopathy of a defined extent, location, and/or clinical findings that is particularly associated with severe visual loss.</td>
</tr>
<tr>
<td><strong>Intraretinal Hemorrhage:</strong> A small to moderate hemorrhage that occurs within the retina.</td>
</tr>
<tr>
<td><strong>Intraretinal Macular Edema (IRME):</strong> Thickening of the retina in the macular region of sufficient extent and location to threaten central visual function.</td>
</tr>
<tr>
<td><strong>Intraretinal Microvascular Abnormalities (IRMAs):</strong> An early vascular abnormality consisting of an outpouching of the retinal microvasculature.</td>
</tr>
<tr>
<td><strong>Neovascular Glaucoma (NVG):</strong> Elevation of intraocular pressure caused by the development of neovascularization in the anterior segment of the eye.</td>
</tr>
<tr>
<td><strong>Neovascularization at the Disc (NVD):</strong> Retina neovascularization occurring at or within 1500 µm of the optic disc.</td>
</tr>
<tr>
<td><strong>Neovascularization Elsewhere (NVE):</strong> Retinal neovascularization that is located more than 1500 µm away from the optic disc.</td>
</tr>
<tr>
<td><strong>Neovascularization of the Iris (NVI):</strong> Neovascularization occurring on the iris (rubeosis iris), usually as a result of extensive retinal ischemia.</td>
</tr>
<tr>
<td><strong>No Light Perception (NLP):</strong> The inability to perceive light.</td>
</tr>
<tr>
<td><strong>Nonproliferative Retinopathy (NPDR):</strong> Severities of diabetic retinopathy that precede the development of proliferative diabetic retinopathy.</td>
</tr>
<tr>
<td><strong>Preproliferative Retinopathy (PPDR):</strong> An outdated term referring to more advanced levels of nonproliferative diabetic retinopathy. Because this terminology is not closely associated with disease progression, it has been replaced by the various levels of nonproliferative diabetic retinopathy.</td>
</tr>
<tr>
<td><strong>Proliferative Diabetic Retinopathy (PDR):</strong> An advanced level of diabetic retinopathy, where proliferation of new vessels occurs on or within the retina.</td>
</tr>
<tr>
<td><strong>Rubeosis Iris:</strong> Retinal neovascularization of the iris (NVI).</td>
</tr>
</tbody>
</table>
Diabetic retinopathy is broadly classified into non-PDR (NPDR) and PDR categories. Macular edema may coexist with either group and is not used in the classification of level of retinopathy. The historical terms background retinopathy and preproliferative diabetic retinopathy have been replaced to reflect the specific characteristics and risk stratification of the prognostically important subgroups in NPDR. Generally, diabetic retinopathy progresses from no retinopathy, through mild, moderate, severe, and very severe nonproliferative disease and eventually on to PDR. Level of NPDR is determined by the extent and location of clinical manifestations of retinopathy. Mild NPDR is characterized by limited microvascular abnormalities such as H/Ma, CWS, and increased vascular permeability. Moderate and severe NPDR are characterized by increasing severity of H/Ma, VCABs, IRMA, and vascular closure.

PDR is characterized by vasoproliferation of the retina and its complications, including new vessels on the optic disc (NVD), new vessels elsewhere on the retina (NVE), preretinal hemorrhage (PRH), vitreous hemorrhage, and fibrous tissue proliferation (FP). On the basis of the extent and location of these lesions, PDR is classified as early PDR or high-risk PDR. Larger areas of these complications as well as new vessels that are near the optic disc are associated with greater risks of visual loss. The level of NPDR establishes the risk of progression to sight-threatening retinopathy and dictates appropriate clinical management and follow-up.
Classification of Diabetic Macular Edema

Diabetic macular edema can be present with any level of diabetic retinopathy. When it involves or threatens the center of the macula, it is called CSME (defined earlier). CSME exists if there is retinal thickening at or within 500 µm of the fovea, hard exudates with adjacent retinal thickening at or within 500 µm of the fovea, or an area of retinal thickening 1500 µm or more in diameter, any part of which is within 1500 µm of the fovea. CSME is a clinical diagnosis that is not dependent on visual acuity or results of ancillary testing such as fluorescein angiography.
Other Ocular Manifestations of Diabetes

All structures of the eye are susceptible to complications of diabetes. The consequence of these changes can range from being unnoticed by both patient and physician, to symtomatic but not sight-threatening, to requiring evaluation to rule out potentially life-threatening underlying causes other than diabetes.

Mononeuropathies of the third, fourth, or sixth cranial nerves may arise in association with diabetes, with the fourth cranial nerve being least likely diabetes-associated.  Nerve palsies present a significant diagnostic challenge because misdiagnosis may result in a life-threatening lesion remaining untreated. In one review of cranial nerve palsies treated in a diabetic patient population in 1967, 42% of mononeuropathies were not diabetic in origin. This finding underscores the danger of routinely attributing mononeuropathies to the diabetic condition itself without carefully ruling out other potential causes. The percentage of all extracranial muscle palsies attributable to diabetes mellitus is estimated at 4.5% to 6%. Mononeuropathies may be the initial presenting sign of new-onset diabetes, and diabetes should therefore be considered in the differential diagnosis of any mononeuropathy affecting the extracranial muscles, even in patients who do not claim a history of diabetes. Diabetes-induced third-, fourth-, and sixth-nerve palsies are usually self-limited and should resolve spontaneously in 2 to 6 months. Palsies may recur or subsequently develop in the contralateral eye.

The optic disc can be affected by diabetes in a variety of ways other than NVD or NVE. Diabetic papillopathy must be distinguished from other causes of disc swelling such as true papilledema from increased intracranial pressure, pseudopapilledema such as optic nerve head drusen, toxic optic neuropathies, neoplasms of the optic nerve, and hypertension. Optic disc pallor can occur following spontaneous remission of proliferative retinopathy or remission following scatter (panretinal) laser photoacoagulation (see Fig. 31-19). Because diabetes poses an increased risk for the development of open-angle glaucoma, the disc pallor following remission of retinopathy or PRP must be considered when evaluating the optic nerve head for glaucoma.

A potentially serious diabetic ocular complication is neovascularization of the iris (NVI). Usually the new iris vessels are first observed at the pupillary border, followed by a fine network of vessels over the iris tissue progressing into the filtration angle of the eye. Closure of the angle by the fibrovascular network results in neovascular glaucoma. Neovascular glaucoma is difficult to manage and requires aggressive treatment. Diabetes is the second leading cause of neovascular glaucoma, accounting for 32% of cases. NVI occurs in 4% to 7% of diabetic eyes and may be present in up to 40% to 60% of eyes with proliferative retinopathy. When possible, scatter (panretinal) laser photoacoagulation is the principal therapy for NVI, although other approaches such as goniotherapy, topical/systemic antiglaucoma medications, and antiglaucomatous filtration surgery are available when needed.

The cornea of the diabetic person is more susceptible to injury and slower to heal after injury than is the nondiabetic cornea. The diabetic cornea is also more prone to infectious corneal ulcers, which can lead to rapid loss of vision, need for corneal transplant, or loss of the eye if not treated aggressively. Consequently, diabetic patients using contact lenses should exercise caution and maintain careful monitoring.

Open-angle glaucoma is 1.4 times more common in the diabetic population than in the nondiabetic population. The prevalence of glaucoma increases with age and duration of diabetes, but medical therapy for open-angle glaucoma is generally effective.

Diabetes effects on the crystalline lens can result in transitory refractive changes, alterations in accommodative ability, and cataracts. Refractive change can be significant and is related to fluctuation of blood glucose levels with osmotic lens swelling. Cataracts occur earlier in life and progress more rapidly in the presence of diabetes. Cataracts are 1.6 times more common in people with diabetes than in those without diabetes. In patients with earlier-onset diabetes, duration of diabetes, retinopathy status, diuretic use, and glycosylated hemoglobin are risk factors. In patients with later-onset diabetes, age of the patient, lower intraocular pressure, smoking, and lower diastolic blood pressure (BP) may be additional risk factors. Diabetic patients undergoing simultaneous kidney or pancreas transplantation are at an increased risk of developing all types of cataract, independent of the use of corticosteroids after transplantation. Both phacoemulsification and extracapsular cataract extraction with intraocular lens implantation are appropriate surgical therapies. The principal determinant of postoperative vision and progression of retinopathy is related to the preoperative presence of diabetic macular edema and level of NPDR.

Other findings with higher frequency among patients with diabetes include xanthelasma, microaneurysms of the bulbar conjunctiva, posterior vitreous detachment, and the rare but frequently fatal orbital fungal infection Mucorales (phycomycosis). Prompt diagnosis and treatment of Mucor is crucial, although the survival rate still remains at only 57%.
MONITORING AND TREATMENT OF DIABETIC RETINOPATHY

Appropriate clinical management of diabetic retinopathy has been defined by results of four major, randomized, multicentered clinical trials (Fig. 31-20: the Diabetic Retinopathy Study [DRS], the Early Treatment Diabetic Retinopathy Study [ETDRS], the Diabetic Retinopathy Vitrectomy Study [DRVS], and the DCCT). These studies have elucidated the progression rates of each level of diabetic retinopathy; defined follow-up intervals; and elucidated the proper delivery, timing, and resulting effectiveness of glycemic control and laser photocoagulation surgery (Figs. 31-21, 31-22, 31-23, 31-24). They have also established guidelines for vitrectomy surgery.

Comprehensive Eye Examination

An accurate ocular examination detailing the extent and location of retinopathy-associated findings is critical for making monitoring and treatment decisions in patients with diabetic retinopathy. As detailed later, most of the blindness associated with advanced stages of retinopathy can be averted with appropriate and timely diagnosis and therapy. Unfortunately, many diabetic patients do not receive adequate eye care at an appropriate stage in their disease. In one study, 55% of patients with high-risk PDR and CSME had never had laser photocoagulation. In fact, 11% of type 1 and 7% of type 2 patients with high-risk PDR necessitating prompt treatment had not been examined by an ophthalmologist within the past 2 years.

Dilated ophthalmic examination is superior to nondilated evaluation because only 50% of eyes are correctly classified as to presence and severity of retinopathy through undilated pupils. Appropriate ophthalmic evaluation entails pupillary dilation, slit lamp biomicroscopy, examination of the retinal periphery with indirect ophthalmoscopy or mirrored contact lens, and sometimes gonioscopy. Because of the complexities of the diagnosis and treatment of PDR and CSME, ophthalmologists with specialized knowledge and experience in the management of diabetic retinopathy are required to determine and provide appropriate surgical intervention. Thus, it is recommended that all patients with diabetes should have dilated ocular examinations by an experienced care provider (ophthalmologist or optometrist) and should be under the direct or consulting care of an ophthalmologist experienced in the management of diabetic retinopathy at least by the time severe diabetic retinopathy or diabetic macular edema is present.
Initial Ophthalmic Evaluation

The recommendation for initial ocular examination in persons with diabetes is based on prevalence rates of retinopathy (see Fig. 31-21). Approximately 80% of type 1 patients have retinopathy after 15 years of disease, but only about 25% have any retinopathy after 5 years. The prevalence of PDR is less than 2% at 5 years and 25% by 15 years. For type 2 diabetes, the onset date of diabetes is frequently unknown, and more severe disease can be observed at diagnosis. Up to 3% of patients first diagnosed after age 30 (type 2) can have CSME or high-risk PDR at the time of initial diagnosis of diabetes. Thus, in patients older than 15 years of age, initial ophthalmic examination is recommended beginning 5 years after the diagnosis of type 1 diabetes mellitus and on diagnosis of type 2 diabetes mellitus (see Fig. 31-21).

Puberty and pregnancy can accelerate retinopathy progression. The onset of vision-threatening retinopathy is rare in children prior to puberty, regardless of the duration of diabetes. However, if diabetes is diagnosed between the ages of 10 and 30, significant retinopathy may arise within 6 years of disease. Diabetic retinopathy can become particularly aggressive during pregnancy in patients with diabetes. In the past, the prognosis for pregnancy in the diabetic patient with microvascular complications was so poor that pregnant diabetic patients were frequently advised to avoid or terminate pregnancies. With recognition of the importance of glycemic control, many diabetic patients in the child-bearing age now experience a safe and satisfying pregnancy and childbirth with minimal risk to both the mother and the baby. There are excellent recent reviews on this subject.

Ideally, patients with diabetes who are planning pregnancy should have a comprehensive eye examination within 1 year prior to conception (see Fig. 31-23). Patients who become pregnant should have a comprehensive eye examination in the first trimester of pregnancy. Close follow-up throughout pregnancy is indicated, with subsequent examinations determined by the findings present at the first-trimester examination. This guideline does not apply to women who develop gestational diabetes, because such individuals are not at increased risk of developing diabetic retinopathy.
Follow-up Ophthalmic Examination

Follow-up ocular examination is determined from the risk of disease progression at any particular retinopathy level (see Fig. 31-22). As described earlier, NPDR is categorized into four levels of severity based on clinical findings compared to stereo fundus photographic standards: mild, moderate, severe, and very severe. Progression of nonproliferative retinopathy to the visually threatening level of high-risk PDR is closely correlated with NPDR level (Table 31-2). Progression rates from each individual NPDR level to any other retinopathy level are also known. These are used to define standard minimal follow-up intervals as detailed in Figure 31-22 and Table 31-3. Patients with no clinically evident diabetic retinopathy and no known ocular problems require annual comprehensive ophthalmic examinations even if totally asymptomatic.
Diabetic Retinopathy

As detailed earlier, the extent and location of neovascularization determine the level of PDR. PDR is best evaluated by dilated examination using slit lamp biomicroscopy, combined with indirect ophthalmoscopy, and/or stereo fundus photography. Without photocoagulation, eyes with high-risk PDR have a 28% risk of severe visual loss within 2 years. This compares with a 7% risk of severe visual loss after 2 years for eyes with PDR but without high-risk characteristics. Severe visual loss is defined as best-corrected acuity of 5/200 or worse on two consecutive visits 4 months apart. The DRS demonstrated that scatter (panretinal) laser photocoagulation was effective in reducing the risk of severe vision loss from PDR by 50% or more. The ETDRS demonstrated that PRP applied when an eye approaches or just reaches high-risk PDR reduces the risk of severe vision loss to less than 4%. Prompt PRP is indicated for all patients with high-risk PDR, often indicated for patients with PDR less than high risk and, on occasion, advisable for patients with severe or very severe NPDR, especially in the setting of type 2 diabetes.

Severe visual loss is defined as best-corrected acuity of 5/200 or worse on two consecutive visits 4 months apart. The DRS demonstrated that scatter (panretinal) laser photocoagulation was effective in reducing the risk of severe vision loss from PDR by 50% or more. The ETDRS demonstrated that PRP applied when an eye approaches or just reaches high-risk PDR reduces the risk of severe vision loss to less than 4%. Prompt PRP is indicated for all patients with high-risk PDR, often indicated for patients with PDR less than high risk and, on occasion, advisable for patients with severe or very severe NPDR, especially in the setting of type 2 diabetes.

In scatter PRP, 1200 to 1800 laser burns are applied to the peripheral retinal tissue, actually focally destroying the outer photoreceptor and retinal pigment epithelium of the retina (see Fig. 31-19) (Figure Not Available). Large vessels are avoided, as are areas of preretinal hemorrhage. The treatment is thought to exert its effect by increasing oxygen delivery to the inner retina and decreasing viable hypoxic growth factor-producing cells. The total treatment is usually applied over two or three sessions, spaced 1 to 2 weeks apart. Follow-up evaluation usually occurs at 3 months.

The response to PRP varies. The most desirable effect is to see a regression of the new vessels, although stabilization of the neovascularization with no further growth may result. This later situation requires careful clinical monitoring. In some cases, new vessels continue to proliferate, requiring additional PRP (see Fig. 31-24). The DRVS, completed in 1989, demonstrated that early PPV in persons with severe fibrovascular proliferation was more likely to result in better vision and less likely to result in poor vision, particularly in patients with type 1 diabetes. PPV is surgery within the eye aimed primarily at removing abnormal fibrovascular tissue, alleviating retinal traction, and removing vitreous opacities such as vitreous hemorrhage. The actual outcome data from this study may not be totally applicable due to the dramatic advances in surgical techniques and the advent of laser endophotocoagulation that have occurred in the intervening years.
Macular Edema

Untreated CSME is associated with an approximately 25% chance of moderate visual loss after 3 years (defined as at least doubling the visual angle, e.g., 20/40 to 20/80). Macular edema is best evaluated by dilated examination using slit lamp biomicroscopy or stereo fundus photography. Focal laser photoacoagulation is generally indicated for patients with CSME (see Fig. 31-19 C [Figure Not Available] and see Fig. 31-22 ). The ETDRS demonstrated that focal laser photoacoagulation for CSME reduced the 5-year risk of moderate vision loss from nearly 30% to less than 15%. In focal laser photoacoagulation, lesions from 300 to 3000 µm from the center of the macula that are contributing to thickening of the macular area are directly photoacoagulated. These lesions are generally identified by fluorescein angiography and consist primarily of leaking microaneurysms. When leakage is diffuse, or microaneurysms innumerable, photoacoagulation may be applied to the macula in a grid configuration, avoiding the fovea region.

Although fluorescein angiography is useful for guiding therapy once CSME has been diagnosed, it is not required for the diagnosis of CSME or PDR because these findings should be clinically evident in most cases (Fig. 31-25). Fluorescein angiography is a valuable test for guiding treatment of CSME, identifying macular capillary nonperfusion, and evaluating unexplained visual loss. However, there are risks associated with fluorescein angiography, including nausea, urticaria, hives, and rarely death (1 in 222,000 patients) or severe medical sequelae (1 in 2000 patients). Thus, fluorescein angiography is not part of the examination of an otherwise normal patient with diabetes, and the procedure is usually contraindicated in patients with known allergy to fluorescein dye or pregnancy.

Follow-up evaluation of focal laser surgery generally occurs after 3 months (Table 31-3, see also Fig. 31-22). In the cases where macular edema persists, further treatment may be necessary. In the presence of macular edema, patients with severe or very severe NPDR should be considered for focal treatment of macular edema whether or not the macular edema is clinically significant because they are likely to require scatter laser photoacoagulation in the near future and because scatter photoacoagulation may exacerbate existing macular edema.

<table>
<thead>
<tr>
<th>TABLE 31-3 – Recommended General Management of Diabetic Retinopathy (DR)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level of DR</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Mild NPDR</td>
</tr>
<tr>
<td>No ME</td>
</tr>
<tr>
<td>ME</td>
</tr>
<tr>
<td>CSME</td>
</tr>
<tr>
<td>Moderate NPDR</td>
</tr>
<tr>
<td>No ME</td>
</tr>
<tr>
<td>ME</td>
</tr>
<tr>
<td>CSME</td>
</tr>
<tr>
<td>Severe NPDR</td>
</tr>
<tr>
<td>No ME</td>
</tr>
<tr>
<td>ME</td>
</tr>
<tr>
<td>CSME</td>
</tr>
<tr>
<td>Very Severe NPDR</td>
</tr>
<tr>
<td>No ME</td>
</tr>
<tr>
<td>ME</td>
</tr>
<tr>
<td>CSME</td>
</tr>
<tr>
<td>PDR &lt; High Risk</td>
</tr>
<tr>
<td>No ME</td>
</tr>
<tr>
<td>ME</td>
</tr>
<tr>
<td>CSME</td>
</tr>
<tr>
<td>High-Risk PDR</td>
</tr>
<tr>
<td>No ME</td>
</tr>
<tr>
<td>ME</td>
</tr>
<tr>
<td>CSME</td>
</tr>
</tbody>
</table>

CSME, clinically significant macular edema; FA, fluorescein angiography; ME, macular edema; NPDR = nonproliferative diabetic retinopathy; Occ, occasionally; OccAF, occasionally after focal; PDR, proliferative diabetic retinopathy.

Courtesy of Lloyd M. Aiello, MD.
Control of Systemic Disorders

In addition to the importance of intensive glycemic control in reducing the onset and progression of diabetic retinopathy as discussed earlier, it is critical for optimal ocular health of diabetic patients that several other systemic considerations be optimized.

Patients with diabetes mellitus commonly suffer from concomitant hypertension. Patients with type 1 diabetes have a 17% prevalence of hypertension at baseline and a 25% incidence after 10 years. There is a 38% to 68% prevalence in type 2 diabetes. In most studies, hypertension is correlated to the duration of diabetes, higher glycosylated hemoglobin level, presence of gross proteinuria, and male gender. Elevated BP exacerbates the development and progression of diabetic retinopathy. The risk of PDR is associated with the presence of hypertension at the baseline visit, higher glycosylated hemoglobin levels, and presence of more severe levels of retinopathy and diffuse macular edema, and more severe levels of retinopathy and more rapid progression of retinopathy when compared with diabetic patients without hypertension. The large randomized, prospective UKPDS in 1148 patients with type 2 diabetes demonstrated a 34% (P = .0004) and 47% (P = .004) reduction in risk of diabetic retinopathy progression and moderate visual acuity loss, respectively, in patients assigned to intensive BP control. These effects were independent of glycemic control and the risk reductions were similar, regardless of whether the hypertension was controlled with ACE inhibitor (captopril) or beta blocker (atenolol). Overall, hypertension appears to be a significant risk factor in the development and progression of diabetic retinopathy and should be rigorously controlled. Until the results of specific trials investigating the BP levels required to minimize end organ damage in patients with diabetes are known, target BP should most likely be maintained as low as safely possible.

Associations between renal and retinal angiopathy are numerous. Proteinuria or microalbuminuria is associated with retinopathy. The presence and severity of diabetic retinopathy are indicators of the risk of gross proteinuria. The presence of proteinuria predicts PDR. Half of all patients with type 1 diabetes mellitus with PDR and 10 or more years of diabetes have concomitant proteinuria. In type 1 diabetes mellitus, the prevalence of PDR increases from 7% at onset of microalbuminuria to 29% 4 years after onset of albuminuria as compared with 3% and 8%, respectively, in patients without persistent microalbuminuria. The Appropriate Blood Pressure Control in Diabetes (ABCD) Trial found both the severity and progression of retinopathy were associated with overt albuminuria. The presence of gross proteinuria at baseline is associated with 95% increased risk of developing macular edema among patients with type 1 diabetes mellitus, and dialysis may improve macular edema in diabetic patients with renal failure.

Despite these associations, the frequent coexistence of retinal and renal microangiopathies and factors such as associated hypertension and disease duration may confound these results. Overall, it is important to carefully consider the renal status of any patient with diabetes mellitus and to ensure that the patient is receiving optimal care in this regard. In addition, rapidly progressive retinopathy, especially in a patient with long history of diabetes mellitus and where retinopathy has been previously stable, should suggest the need for renal evaluation.

Although dyslipidemia is a clear risk factor in diabetic renal disease, the effects of serum lipids on retinopathy and macular edema are less certain. In 2709 ETDRS patients in whom serum levels were measured, elevated total cholesterol, LDL cholesterol, and triglyceride levels were associated with faster development of hard exudates (P < .001, P = .04, and P = .01, respectively), and the risk of moderate visual loss was associated with the extent of exudate. Owing to their salutary effects on cardiovascular morbidity for individuals at risk, lipid-lowering recommendations are currently given to all patients with diabetes and elevated cholesterol levels, irrespective of their retinopathy status.

The precise effects of exercise and physical exertion on diabetic retinopathy remain unknown. Exercise may help reduce the risk of diabetic complications by improving cardiovascular function, increasing HDL, protecting weight control, and increasing insulin sensitivity. In general, exercise and physical activity have not been shown to accelerate diabetic retinopathy, and no studies on exercise and type 1 diabetes show a detrimental effect on the development of PDR. However, physical exercise can have potentially detrimental effects on retinopathy and vision, especially when advanced retinal disease is present. Such adverse effects might be mediated by elevated systolic BP with subsequent vitreous hemorrhage, or decreases in already compromised tissue oxygen concentrations. However, the exact threshold for hemorrhage from the abnormal new vessels in PDR is unknown, with up to 84% of vitreous hemorrhages associated with exercise no more strenuous than walking in some studies. In general, people with PDR should avoid anaerobic exercise and exercise that involves straining, jarring, near-maximal isometric contractions, or Valsalva-type maneuvers such as high-impact aerobics, jogging, or heavy weight training. Beneficial low-risk exercises include stationary cycling, low-intensity machine rowing, swimming, and walking. Because of the beneficial effects, all patients with diabetes should be encouraged to participate in regular physical exercise programs specifically tailored to their individual ocular status.

As conclusively demonstrated by the ETDRS, the use of aspirin in diabetic patients is not associated with an increased risk of hemorrhage and has no demonstrated impact on the progression of retinopathy or macular edema. There is one case report of vitreous hemorrhage associated with retinopathy after thrombolysis. However, approximately 90,000 patients have been involved in clinical trials of thrombolytic agents in myocardial infarction, of which 10% had diabetes mellitus without any reports of ocular complications.

Smoking is a certain risk factor for CVD, progression of albuminuria to proteinuria, and nephropathy in both type 1 and 2 diabetic patients. However, the effects of smoking in diabetic retinopathy are unclear, because some studies have suggested an association whereas others have not. Because smoking has detrimental effects on the cardiovascular system and the development of nephropathy, smoking in patients with diabetes mellitus should be discouraged in the strongest possible terms.

Low hematocrit was an independent risk factor in the ETDRS analysis of baseline risk factors for development of high-risk PDR and of severe visual loss. A cross-sectional study involving 1691 patients revealed a twofold increased risk of any retinopathy in patients with a hemoglobin level of less than 12 g/dL as compared to those with a higher hemoglobin concentration using multivariate analyses controlling for serum creatinine, proteinuria, and other factors. In patients with retinopathy, those with low hemoglobin levels have a fivefold increased risk of severe retinopathy compared with those with higher hemoglobin levels. There have been limited reports of resolution of macular edema and hard exudate with improvement or stabilization of visual acuity in erythropoietin-treated patients after an increase in mean hematocrit. In view of the potential association of low hematocrit and diabetic retinopathy, it is important to ensure that patients with diabetic
Retinopathy and anemia are receiving appropriate management.

In summary, diabetes is clearly a multisystem disease requiring a medical team approach. Even with regard to ocular health, this necessitates the involvement of multiple health care specialists for optimal patient care.
Investigational Approaches

Our understanding of the basic molecular mechanisms underlying the development, progression, and damage from diabetic retinopathy has markedly expanded in recent years. There is considerable evidence to suggest that significant beneficial effects may be achieved using pharmacologic interventions to prevent or delay the development of nonproliferative retinopathy, suppress retinal vascular leakage, and inhibit intraocular neovascularization. Such efforts hold promise for the prevention of visual loss without the retinal damage inherent with current photocoagulation therapies. Indeed, numerous pharmacologic agents are in or approaching clinical trial for indications involving diabetic retinopathy, macular edema, and other ocular disorders. Results from these trials should become available within the next few years and may herald a new era in our therapeutic approach to diabetic ocular complications.
DIABETIC NEPHROPATHY

Diabetic nephropathy is clinically defined by persistent proteinuria greater than 500 mg/24 hours in a person with diabetic retinopathy without other renal disease. The distribution of renal disease due to type 2 diabetes is uneven among racial groups. American Indians, African Americans, and Mexican Americans have a greater incidence than non-Hispanic whites. Genetic predisposition, environmental factors, delayed diagnosis of type 2 diabetes, and subadequate medical care in minority groups contribute in undefined amounts to such disparity.

PATHOLOGY

Kidney injury in diabetes is indistinguishable by diabetes type and affects glomeruli, arterioles, tubules, and interstitium. Glomerular lesions include diffuse and nodular forms of intracapillary glomerulosclerosis (Fig. 31-26 A to C) (see also Color Plate). The diffuse type is characterized by mesangial expansion with increased PAS-positive matrix material, thickening of capillary wall and basement membrane. In type 1 diabetes, an early structural abnormality of diabetic nephropathy is glomerular basement membrane thickening (Fig. 31-27 D). Also noted in early and sequential biopsies is an increase in mesangial fractional volume (mesangial volume per glomerulus). With progression, capillary wall thickening and mesangial widening lead to capillary narrowing and reduced glomerular capillary filtration surface area. In diabetic patients, there is good correlation between the degree of mesangial expansion and the severity of clinical diabetic nephropathy.

Although nodular lesions were first thought to be pathognomonic for diabetic glomerulopathy, they may also be noted by light microscopy in amyloidosis, dysproteinemias (multiple myeloma and heavy chain disease), and glomerulonephritis (mesangial proliferative and membranoproliferative). Typical nodules are well-demarcated, PAS-positive globular structures exaggerating the diffuse lesions, and they occur at the periphery of the glomeruli. Additionally, hyaline deposits (so-called exudative or insudative lesions), consisting of plasma proteins and lipids, are present in arterioles (hyaline arteriosclerosis), capillary walls (fibrin cap), and Bowman's capsules (capsular drops). Hyaline arteriosclerosis or arteriolar hyalinosis, prominent in diabetic nephropathy, affects afferent as well as efferent arterioles. Initially, collections of hyaline material (hyaline drops) aggregate in the wall of juxtaglomerular arterioles. Gradually, these drops increase in size and replace the entire wall structure. Progression of glomerular and arteriolar changes ends in complete sclerosis. Østerby and associates

Despite narrowing, hyaline arteriosclerosis, glomerular sclerosis or occlusion, and interstitial fibrosis,
NEPHROPATHY IN TYPE 1 DIABETES

The natural history of diabetic nephropathy has been extensively studied in type 1 diabetes because it is usually possible to specify the exact time of onset. As first described by Mogensen, there are five distinct stages.\[405\] The course of diabetic nephropathy can be followed by two main variables: proteinuria and GFR (Fig. 31-28).

Stage 1: Glomerular Hyperfiltration and Renal Enlargement

Glomerular hyperfiltration and kidney enlargement typify the first stage. In 1934, Cambier noted that inulin clearances were greater than normal in type 1 diabetes.\[431\] Subsequent studies, using creatinine clearance or radionuclide techniques, validated this finding.\[432\]\[433\]\[434\] At onset of type 1 diabetes, approximately one third of individuals have an elevated GFR that is 20% to 40% higher than that of age-matched normal subjects.\[435\] No single pathophysiologic mechanism totally explains both kidney enlargement and glomerular hyperfiltration. Hyperglycemia, hormonal and vasoactive factors, enhanced renal plasma flow, and elevated transglomerular hydrostatic pressure gradient have been proposed as determinants of diabetic hyperfiltration.\[436\]\[437\]\[438\] With intensive insulin therapy, hyperglycemia decreases and GFR starts to decline within 3 to 8 days and drops further over the next few months.\[435\]\[439\] Thus, hyperfiltration may be an early indicator of individual susceptibility to hyperglycemia-induced renal changes. Hyperfiltration is a predictor of clinical nephropathy in some individuals. Mogensen's initial findings\[440\] that an elevated GFR predicts subsequent glomerulopathy in type 1 patients, previously contested, were confirmed by two different reports.\[441\]\[442\] In an 8-year prospective study of diabetic adolescents, Rudberg and colleagues showed that an increase of the initial GFR significantly predicted nephropathy.\[443\] Recently, Chiarelli and associates, in a 10-year longitudinal study, indicated glomerular hyperfiltration increased the risk of developing microalbuminuria in diabetic children.\[444\] Short-term or long-term intensive insulin treatments did not reduce kidney size in type 1 diabetic individuals.\[445\]\[446\]
Stage 2: Early Glomerular Lesions or Silent Stage with Normal Albumin Excretion

Early glomerular lesions, consisting of glomerular basement membrane thickening and mesangial matrix expansion, characterize the second stage. Those structural changes appear 18 to 36 months after onset of type 1 diabetes and may become prominent after 3.5 to 5 years. During this stage of morphologic changes, microalbuminuria, seen only after exercise or during episodes of very poor metabolic control, may be the only clinically detectable evidence of renal involvement. Otherwise, nephropathy is silent with normal AER (<25 mg/day).
Stage 3: Incipient Diabetic Nephropathy or Microalbuminuria Stage

The third stage, also called incipient diabetic nephropathy, is characterized by persistent and usually increasing microalbuminuria. Hypertension may also be a feature of the microalbuminuric stage. Hyperfiltration and renal enlargement persist, though to a lesser degree. Microalbuminuria, defined as urinary AER greater than 30 mg/24 hours or 20 µg/minute and less than 300 mg/24 hours or 200 µg/minute, represents the first laboratory evidence of diabetic renal disease. Total daily AER varies greatly and is increased by hypertension, strenuous exercise, fever, poor glycemic control, and congestive heart failure (CHF). Therefore, a diagnosis of incipient diabetic nephropathy is made only when microalbuminuria is detected in at least two of three urine specimens over several months.

Hemodynamic abnormalities and variable glomerular charge-selective properties due to loss of negatively charged proteoglycans contribute to variability of microalbuminuria. Measurements of urinary AER can be carried out by 24-hour, overnight, or short-term urine collections. Determinations of albumin:creatinine ratio (30-300 mg/g) or albumin concentration from an early morning urine sample are acceptable for screening, but timed urine collection is more accurate. The prevalence of microalbuminuria varies from 25% to 40% in individuals with type 1 diabetes for 5 to 15 years. Persistent microalbuminuria rarely occurs during the first 5 years of type 1 diabetes or before puberty. Consequently, screening in type 1 subjects should start after 3 years of diabetes duration or with puberty. Microalbuminuria is a sign of renal damage in both types of diabetes that predicts later nephropathy and, ultimately, ESRD. Therefore, detection of microalbuminuria represents a vital step in the early management of diabetic renal disease.
Stage 4: Clinical or Overt Diabetic Nephropathy: Proteinuria and Falling Glomerular Filtration Rate

Albuminuria greater than 300 mg/24 hours, relentless decline of renal function, and hypertension define the fourth stage of diabetic nephropathy. This stage, though variable, usually occurs 15 to 20 years after the onset of type 1 diabetes and after 5 or more years of diagnosed type 2 diabetes. The amount of urinary protein can be as little as 500 mg, but it can reach massive proportions, such as 20 to 40 g/24 hours. Continuing urinary protein loss of this magnitude is associated with increased glomerular pore size. There is a high mortality rate associated with proteinuria. Median survival is 10 years from the onset of proteinuria.

Diagnoses other than diabetic nephropathy should be pursued whenever a nephrotic syndrome develops in a patient with short-term type 1 diabetes or in the absence of retinopathy. Similarly, a diagnosis of diabetic nephropathy is prudently doubted when progressive renal insufficiency in a diabetic patient is not accompanied by macroalbuminuria (AER > 300 mg/24 hours). Percutaneous renal biopsy to clarify the renal disorder is indicated in such instances. In the follow-up of more than 90 patients with insulin-dependent diabetes and diabetic nephropathy, Viberti and co-workers have not encountered a single case of progressive renal failure without macroalbuminuria.

In subjects with type 1 diabetes, the prevalence of arterial hypertension ranges from 65% to 79% when macroalbuminuria is present. Hypertension intensifies the rate of progression of established diabetic renal disease. When structural kidney changes are advanced, GFR will invariably decrease. Without treatment, GFR usually declines in a linear fashion at a rate ranging from 7.5 to 28 mL/minute per year.
Stage 5: End-Stage Renal Disease

After 20 to 30 years of type 1 diabetes, about 30% to 40% of patients progress to ESRD. Recently, the interval between the onset of persistent proteinuria and the final stage of diabetic nephropathy has been lengthened by early and intensive treatment of hypertension and enhanced metabolic control of hyperglycemia.
NEPHROPATHY IN TYPE 2 DIABETES

Although renal structural changes and severity of target organ damage are similar in both types of diabetes, delayed diagnosis has complicated the construction of the natural history of diabetic renal disease in type 2 diabetes. The results of studies about renal hemodynamics and hypertrophy in newly diagnosed and established patients have been inconsistent. For example, Vora and associates reported that 45% of their 110 patients with type 2 diabetes had a GFR higher than 120 mL/minute per 1.73 m² and 16% had frank hyperfiltration (GFR > 140 mL/minute per 1.73 m²). By contrast, Schmitz and colleagues did not observe hyperfiltration in their newly diagnosed patients with type 2 diabetes (GFR 106 ± 14 mL/minute per 1.73 m²).

Fourteen percent to 24% of newly diagnosed patients with type 2 diabetes have microalbuminuria, which is associated with hyperglycemia, elevated BP, smoking, and hyperlipidemia. Microalbuminuria in type 2 diabetes is partially reversed by reduction of hyperglycemia and high BP. In a systematic review of the literature linking microalbuminuria to cardiovascular mortality in individuals with type 2 diabetes, Dinneen and Genstein found that the prevalence of microalbuminuria ranged from 20% to 36% in diabetic patients. There was also a significant association between microalbuminuria and total or cardiovascular mortality. Microalbuminuria raised the overall odds ratio for death to 2.4 and cardiovascular mortality to 2.0 over those without microalbuminuria. For older people, other causes of microalbuminuria should be considered before attributing this abnormality to type 2 diabetes. In a population-based study in southern Wisconsin, the prevalence of overt proteinuria in a 10-year interval was 33%. GFR of diabetic Arizona Pima Indians with macroalbuminuria declined by 35% over a 4-year period. Hypertension is highly characteristic of renal disease in type 2 diabetes, whether the individuals are normoalbuminuric, microalbuminuric, or macroalbuminuric.
Dietary Protein and Lipid Restriction

Rodent studies show that a low-protein diet reduces glomerular hypertension and prevents glomerular injury and albuminuria. In type 1 diabetic subjects with microalbuminuria and glomerular hyperfiltration, short-term dietary protein restriction (0.6 to 0.8 g/kg per day) decreases urinary AER and hyperfiltration. Long-term studies of protein restriction have been criticized for using creatinine clearance or the reciprocal of the serum creatinine instead of inulin clearance or radionuclide techniques to assess renal function. From a meta-analysis of five studies, Pedrini and colleagues concluded that dietary protein restriction delayed progression of diabetic nephropathy in subjects with type 1 diabetes, a finding disputed by Parving. In type 2 diabetes, trials of protein restriction are few and positive results have not been attained.

Abnormal lipid metabolism is highly prevalent in diabetic individuals with nephropathy, especially in those with a nephrotic syndrome. Although rodent studies imply that hyperlipidemia is important in the pathogenesis of glomerular injury, only suggestive data in humans have been reported. Nevertheless, treatment of dyslipidemia is paramount in the overall management of diabetes because it decreases the risk of CVDs.
OTHER DIABETES-ASSOCIATED RENAL DISEASES

Three other renal conditions are associated with diabetes: (1) urinary tract infection, (2) papillary necrosis, and (3) radiocontrast-induced renal failure. In long-standing type 1 diabetes, diabetic women, but not diabetic men, have an increased frequency of urinary tract infection. Diabetic patients are also at risk for emphysematous pyelonephritis, a rare, life-threatening complication of upper urinary tract infections, characterized by gas in the renal parenchyma or perirenal space. Renal ultrasonography or computed tomographic (CT) scanning is necessary to detect upper urinary tract complications early for appropriate treatment. Parenteral antimicrobial therapy and close metabolic control of diabetes with insulin therapy are mandatory, but nephrectomy or drainage is often necessary to preserve life.

Renal papillary necrosis, a severe destruction of renal parenchyma due to impaired blood flow to the inner medulla and papilla of the kidney, although observed in urinary infections, analgesic abuse, and sickle cell disease, is most common in diabetes. Clinical manifestations comprise flank pain, hematuria, chills, fever, and septicemia. Red and white blood cells, bacteria, and fragments of renal papillae can be seen in strained urine. Ureteral obstruction due to papilla fragments should be relieved promptly.

Radiocontrast-induced renal failure occurs more frequently in type 1 and 2 diabetic patients than in nondiabetic patients when the serum creatinine level is higher than 2 mg/dL. The use of radiographic contrast medium in a diabetic patient with renal insufficiency should be limited whenever possible. When angiography is unavoidable, patients should be aware of the risk of acute renal failure. Hydration before and after contrast injection may reduce the risk of nephropathy. A highly encouraging report indicates that pretreatment with acetylcysteine may protect against radiographic contrast media nephropathy.
HYPORENINEMIC HYPOALDOSTERONISM

Hyporeninemic hypoaldosteronism is prevalent in patients with renal insufficiency due to diabetic nephropathy. The syndrome of hyporeninemic hypoaldosteronism is characterized by normal cortisol levels, hyperkalemia, and low plasma renin, angiotensin II, and aldosterone levels. Most patients with the syndrome manifest a coexistent hyperchloremic metabolic acidosis. In diabetes mellitus, low plasma renin activity is due to defective conversion of prorenin to active renin. Increased extracellular volume due to diabetic nephropathy may, in part, diminish renin release. In this case, hyporeninemia is corrected by diuresis.

Clark and co-workers suggest that an increase in ANP contributes to hypoaldosteronism and hyperkalemia in the syndrome of acquired hypoaldosteronism. ACE inhibitors, non-steroidal anti-inflammatory drugs (NSAIDs), cyclosporine, potassium-sparing diuretics, trimethoprim, and pentamidine all can accentuate the hyperkalemia of hyporeninemic hypoaldosteronism.
TREATMENT OF DIABETIC NEPHROPATHY

Experience of the past two decades convincingly demonstrates that intensive control of hyperglycemia and adequate lowering of hypertensive BP are the key components of diabetic nephropathy management. Control of both hyperglycemia and hypertension has modified the natural course of diabetic nephropathy by reversing functional changes and by stabilizing progression of structural abnormalities (Fig. 31-29).

Glycemic Control

Large-scale, prospective trials provide compelling evidence that intensive glycemic control prevents diabetic nephropathy. The DCCT, which strived for near-normoglycemia in patients with type 1 diabetes, showed a 39% reduction in the risk of developing microalbuminuria and a 54% reduction in the occurrence of albuminuria. The UKPDS, comparing intensive blood glucose with conventional therapy in type 2 diabetes, found a 25% risk reduction (7% to 40%; \( P = .0099 \)) in microvascular complications, including progressive nephropathy. Degradation of insulin is compromised in renal failure. Consequently, progressive reductions in oral agent and insulin doses are necessary to minimize hypoglycemia. Metformin, a powerful hypoglycemic drug, is contraindicated in renal insufficiency (serum creatinine > 2.0 mg/dL) because of the risk of fatal lactic acidosis.
Blood Pressure Control

Early diagnosis and intensive treatment of arterial hypertension, whether with ACE inhibitors or other antihypertensive drugs in combination with low-dose diuretics, are essential for diabetic patients with diabetic renal disease. Effective BP control reduces albuminuria, delays progression of nephropathy, postpones renal insufficiency, and improves survival in type 1 and type 2 diabetic patients with diabetic nephropathy. Moreover, risk of fatal or nonfatal cardiovascular events is decreased in diabetic patients when systolic hypertension is treated.

The National Kidney Foundation Hypertension and Diabetes Executive Committees Working Group has lowered the previously recommended BP level of 130/85 to 130/80 mm Hg to optimally preserve renal function and reduce cardiovascular events in diabetic nephropathy. Two or more antihypertensive drugs are usually required to achieve this new target. Additional BP levels lower than 125/75 mm Hg are recommended for people who have proteinuria higher than 1 g/day and renal insufficiency regardless of etiology. Hypoproteinemic diabetic patients with renal insufficiency are susceptible to striking fluid retention, making BP control refractory to usual antihypertensive therapy. Dietary salt restriction and a combination of loop diuretics plus metolazone, a quinazoline diuretic, are necessary. A regimen of furosemide, 80 mg twice a day, and metolazone (Zaroxolyn), 10 mg twice a day, usually is effective as long as the creatinine clearance is higher than 10 mL/minute.

ACE inhibitors are efficient in both types of diabetes. Apart from reducing arterial BP, they are renoprotective due to their ability to decrease intraglomerular pressure.

Several randomized, controlled trials in normotensive type 1 and type 2 diabetic subjects with incipient nephropathy demonstrate that ACE inhibitors reduce microalbuminuria and may even preempt progression to overt nephropathy. Bakris and associates suggest that combining a calcium antagonist with an ACE inhibitor results in a greater decrease in urinary protein excretion and slower GFR decline. In addition, this combination permits equivalent BP reduction by lower doses of both drugs. ACE inhibitors can worsen hyperkalemia in diabetic patients with hyporeninemic hypoaldosteronism and induce temporary renal failure in those with bilateral renal artery stenosis or single-kidney and renal stenosis. Approximately one in five diabetic patients treated with ACE inhibitors discontinued the drug because of persistent nonproductive cough.

A related class of antihypertensive drugs, the angiotensin receptor blockers, block angiotensin II at the receptor level. Whether angiotensin receptor blockers will match the ACE inhibitors in improving the natural history of renal diseases is the subject of two large prospective, randomized trials in type 2 diabetic patients.
Dietary Protein and Lipid Restriction

Rodent studies show that a low-protein diet reduces glomerular hypertension and prevents glomerular injury and albuminuria. In type 1 diabetic subjects with microalbuminuria and glomerular hyperfiltration, short-term dietary protein restriction (0.6 to 0.8 g/kg per day) decreases urinary AER and hyperfiltration. Long-term studies of protein restriction have been criticized for using creatinine clearance or the reciprocal of the serum creatinine instead of inulin clearance or radionuclide techniques to assess renal function. From a meta-analysis of five studies, Pedrini and colleagues concluded that dietary protein restriction delayed progression of diabetic nephropathy in subjects with type 1 diabetes, a finding disputed by Parving. In type 2 diabetes, trials of protein restriction are few and positive results have not been attained.

Abnormal lipid metabolism is highly prevalent in diabetic individuals with nephropathy, especially in those with a nephrotic syndrome. Although rodent studies imply that hyperlipidemia is important in the pathogenesis of glomerular injury, only suggestive data in humans have been reported. Nevertheless, treatment of dyslipidemia is paramount in the overall management of diabetes because it decreases the risk of CVDs.
Uremia Therapy

While renal function relentlessly deteriorates, patients with diabetic renal disorder should be instructed as to the available options in uremia therapy, including home or facility hemodialysis, continuous ambulatory peritoneal dialysis, continuous cyclic peritoneal dialysis, living related or cadaveric kidney transplant, and combined pancreas and kidney transplant (Table 31-4). A small number of uremic diabetic patients, severely debilitated by extensive co-morbid conditions, may choose death instead of renal replacement therapy. As reported by the U.S. Renal Data System (USRDS) in 2000, 531 for 102,942 diabetic patients receiving uremia therapy in 1998, about 74% were treated with facility hemodialysis, but only 7.5% elected peritoneal dialysis. A few (0.7%) adopted home hemodialysis, although this therapy, by consensus, is superior to facility dialysis. Approximately 17% of diabetic ESRD patients had a functioning kidney transplant in the United States in 1998. Kidney transplantation, the best option in uremia therapy for diabetic patients, provides patient survival and rehabilitation that is greater than the best dialytic therapy. 532

Although combined pancreas plus kidney transplantation, when successful, offers exquisite glycemic control and better quality of life, the prospective recipient risks a serious surgical complication rate. 533

<table>
<thead>
<tr>
<th>Variable</th>
<th>Peritoneal Dialysis</th>
<th>Hemodialysis</th>
<th>Kidney Transplant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extensive extrarenal disease</td>
<td>No limitation</td>
<td>No limitation except for hypotension</td>
<td>Excluded in cardiovascular insufficiency</td>
</tr>
<tr>
<td>Geriatric patients</td>
<td>No limitation</td>
<td>Arbitrary exclusion as determined by program</td>
<td></td>
</tr>
<tr>
<td>Complete rehabilitation</td>
<td>Rare, if ever</td>
<td>Very few individuals</td>
<td>Common, so long as graft functions</td>
</tr>
<tr>
<td>Death rate</td>
<td>Much higher than for nondiabetic patients</td>
<td>Much higher than for nondiabetic patients</td>
<td>About the same for nondiabetic patients</td>
</tr>
<tr>
<td>First-year survival</td>
<td>About 75%</td>
<td>About 75%</td>
<td>&gt;90%</td>
</tr>
<tr>
<td>Morbidity during first year</td>
<td>About 15 days in hospital</td>
<td>About 12 days in hospital</td>
<td>Weeks to months hospitalized</td>
</tr>
<tr>
<td>Survival to second decade</td>
<td>Almost never</td>
<td>Fewer than 5%</td>
<td>About 1 in 5</td>
</tr>
<tr>
<td>Progression of complications</td>
<td>Usual and unremitting; hyperglycemia and hyperlipidemia</td>
<td>Usual and unremitting; may benefit from metabolic control</td>
<td>Interdicted by functioning pancreas plus kidney; partially ameliorated by correction of azotemia</td>
</tr>
<tr>
<td>Special advantage</td>
<td>Can be self-performed; avoids swings in solute and intravascular volume level</td>
<td>Can be self-performed; efficient extraction of solute and water in hours</td>
<td>Cures uremia; freedom to travel</td>
</tr>
<tr>
<td>Disadvantage</td>
<td>Peritonitis; hyperinsulinemia; hyperglycemia; hyperlipidemia; long hours of treatment; more days hospitalized than with either hemodialysis or transplant</td>
<td>Blood access a hazard for clotting, hemorrhage, and infection; cyclical hypotension, weakness, aluminum toxicity, amyloidosis</td>
<td>Cosmetic disfigurement, hypertension, personal expense for cytotoxic drugs; induced malignancy; HIV transmission</td>
</tr>
<tr>
<td>Patient acceptance</td>
<td>Variable, usual compliance with passive tolerance for regimen</td>
<td>Variable, often noncompliant with dietary, metabolic, or antihypertensive component of regimen</td>
<td>Enthusiastic during periods of good renal allograft function; exalted when pancreas proffers euglycemia</td>
</tr>
<tr>
<td>Relative cost</td>
<td>Most expensive over long run</td>
<td>Less expensive than kidney transplant in first year, subsequent years more expensive</td>
<td>Pancreas plus kidney engraftment most expensive uremia therapy for diabetic; after first year, kidney transplant alone is lowest cost option</td>
</tr>
</tbody>
</table>

The USRDS reports that diabetic ESRD patients treated with hemodialysis have better survival than those receiving peritoneal dialysis. 534 In contrast, the Canadian-USA (CANUSA) Peritoneal Dialysis Study found superior survival of diabetic ESRD patients treated by peritoneal dialysis in Canada. 535 Overall, survival of diabetic patients receiving uremia therapy, irrespective of modality, is vastly inferior to that of nondiabetic patients. 536 Five-year survival rates are 21% to 31%. Such disparity is the result of cardiovascular complications, cerebrovascular and peripheral vascular disease, plus a greater risk of septicemia.

Diabetic patients with renal failure exhibit symptoms and signs of uremia at a lower serum creatinine than that of nondiabetic patients. As a result, renal replacement therapy is frequently initiated earlier in their deterioration. The end of conservative management is often forced by severe hypervolemia precipitating the start of dialytic therapy. When creatinine clearance falls below 70 mL/minute, individuals with diabetic nephropathy who had been followed by their internists through microalbuminuric and/or macroalbuminuric stages should be referred to nephrologists for pre-ESRD education and planning for management of renal insufficiency (Fig. 31-30).

Delayed referral to renal specialists contributes to late arteriovenous access construction for hemodialysis and increases morbidity, mortality, and utilization of health care resources. 535 Lameir and co-workers, assessing physicians’ referral patterns, note that internists, cardiologists, and endocrinologists refer their patients to nephrologists too late, meaning that urgent hemodialysis must be performed via insertion of a temporary jugular or femoral catheter (Fig. 31-31).

Caring for patients afflicted by diabetic nephropathy is demanding and is better accomplished by a team composed of primary care physician, cardiologist, ophthalmologist, podiatrist, nutritionist, and nurse-educator. Evidence-based data suggest that diabetic nephropathy prevention has been suboptimal in the United States. 535 Constant effort to diagnose type 2 diabetes early, education of diabetic individuals, and primary care providers represent the initial steps of diabetic nephropathy prevention. Better patient management starting in the early phases of diabetic renal disease will prevent or attenuate progression of diabetic nephropathy and will reduce the incidence of cardiovascular complications.
DIABETIC NEUROPATHIES

Diabetic neuropathy (DN) is a common and troublesome complication of diabetes mellitus, leading to great morbidity and mortality and resulting in a huge economic burden for care of the patient with diabetes mellitus. It is the most common form of neuropathy in the developed countries of the world, accounts for more hospitalizations than all the other diabetic complications combined, and is responsible for 50% to 75% of nontraumatic amputations. Diabetic neuropathy is not a single entity but a set of clinical syndromes that affect distinct regions of the nervous system, singly or combined. It may be silent and go undetected while nonetheless exercising its ravages, or it may present with clinical symptoms and signs that are nonspecific and insidious and progress slowly, mimicking those seen in many other diseases.

PREVALENCE

Diabetic neuropathy is a heterogeneous disorder that encompasses a wide range of abnormalities affecting proximal and distal peripheral sensory and motor nerves as well as the autonomic nervous system. For these reasons, it has been difficult to obtain precise estimates of the true prevalence, and reports vary from 10% to 90% in diabetic patients, depending on the criteria and methods used to define neuropathy. Twenty-five percent of patients attending a diabetes mellitus clinic volunteered symptoms; 50% were found to have neuropathy after a simple clinical test such as the ankle jerk or vibration perception threshold (VPT) test; almost 90% tested positive to sophisticated tests of autonomic function or peripheral sensation. Neurologic complications occur equally in type 1 and type 2 diabetes mellitus and additionally in various forms of acquired diabetes.

The major morbidity associated with somatic neuropathy is foot ulceration, the precursor of gangrene and limb loss. DSPN increases the risk of amputation 1.7-fold: 12-fold if there is deformity (itself a consequence of neuropathy) and 36-fold if there is a history of previous ulceration. About 85,000 amputations are performed in the United States each year—one every 2 minutes—and neuropathy is considered to be the major contributor in 87% of cases. It is also the most life-spoiling of the diabetic complications and has tremendous ramifications for the quality of life of the person with diabetes. Once autonomic neuropathy sets in, life can become quite dismal, and the mortality rate approximates 25% to 50% within 5 to 10 years.
NATURAL HISTORY

The natural history of diabetic neuropathy separates patients into two very distinctive entities: (1) those who progress gradually with increasing duration of diabetes mellitus and (2) those who have a relatively explosive onset and experience remission almost completely. Sensory and autonomic neuropathies generally progress, whereas mononeuropathies, radiculopathies, and acute painful neuropathies, although symptoms are severe, are short-lived and tend to recover.

Progression of DSPN is related to glycemic control in both type 1 and type 2 diabetes mellitus. The most rapid deterioration of nerve function occurs soon after the onset of type 1 diabetes mellitus, and within 2 to 3 years there is a slowing of the progress with a shallower slope to the curve of dysfunction. In contrast, slowing of NCVs in type 2 diabetes mellitus may be one of the earliest neuropathic abnormalities and is often present at diagnosis. After diagnosis, slowing of NCVs generally progresses at a steady rate of approximately 1 m/second each year, and the level of impairment is positively correlated with duration of diabetes mellitus. Although most studies have documented that symptomatic patients are more likely to have slower NCVs than patients without symptoms, these do not relate to the severity of symptoms.

In a long-term follow-up study of patients with type 2 diabetes mellitus, electrophyslogic abnormalities in the lower limb increased from 8% at baseline to 42% after 10 years, and a decrease in sensory and motor amplitudes, indicating axonal destruction, was more pronounced than the slowing of the NCVs. An increase of about 2 points in an 80-point clinical symptom and sign scale (neurologic symptom score [NSS] and neurologic impairment score [NIS]) can be expected per year. These scales contain information on motor, sensory, and autonomic signs and symptoms. Using objective measures of sensory function such as the VPT test, the rate of decline in function has been reported as 1 or 2 vibration units per year.

In a 6-year cohort study, elevated VPT (>6.5 vibrations) was found in 62.5% of patients. The risk factors were male sex, age, and increased AER. However, there now appears to be a decline in this rate of evolution. For example, in the nerve growth factor (NGF) study, the VPT at the beginning of the study in the placebo group was identical to that at the end of 1 year. This is particularly important in planning studies on the treatment of DSPN, which have always relied on differences between drug treatment and placebo and have been successful because of the decline in nerve function in the placebo-treated patients. According to the earlier data on rates of change, clinically meaningful loss of VPT and NCV has been estimated to take at least 3 years, dictating a future need to carry out studies over a longer period when considering only large-fiber dysfunction.

We must recognize that DSPN is a disorder in which the prevailing abnormality is loss of axons that electrophysiologically translates to a reduction in amplitudes and not conduction velocities, and changes in NCV may not be an appropriate means of monitoring progress or deterioration of nerve function. It has always been advocated that diabetes mellitus affects the longest fibers first, hence, the increased predisposition in taller individuals. Now it seems that small-fiber involvement may herald the onset of neuropathy and even diabetes mellitus itself. Small-fiber function is not detectable using standard electrodiagnostic methods and requires measurement of sensory, neurovascular, and autonomic thresholds and cutaneous nerve fiber density.

There are few data on the longitudinal trends in small-fiber dysfunction and the mutual concurrence and development of peripheral somatic and autonomic neuropathies. Toyry and associates reported that the development of autonomic and peripheral somatic neuropathies was divergent in patients with type 2 diabetes mellitus, suggesting different pathophysiologic processes for these neuropathies. Much remains to be learned of the natural history of diabetic autonomic neuropathy.

Karamitos and colleagues reported that the progression of diabetic autonomic neuropathy is significant during the 2 years subsequent to its discovery. The mortality for diabetic patients with autonomic neuropathy has been estimated to be 44% within 2.5 years of diagnosing symptomatic autonomic neuropathy. A meta-analysis of 14 longitudinal studies revealed that the mortality rate after 5.8 years of diabetes with asymptomatic autonomic neuropathy was 27%.
CLASSIFICATION

Diabetic neuropathy is not a single entity but a number of different syndromes, ranging from subclinical to clinical manifestations depending on the classes of nerve fibers involved. According to the San Antonio Convention, the main groups of neurologic disturbance in diabetes mellitus include the following:

1. Subclinical neuropathy, determined by abnormalities in electrodiagnostic and quantitative sensory testing without concomitant clinical sign and symptoms.
2. Diffuse clinical neuropathy, which may be proximal or distal and have large symmetrical sensorimotor or small-fiber and autonomic dysfunction.
3. Focal neuropathies, which include mononeuropathies and entrapment syndromes.

The onset of neuropathy may be acute, with pain or, insidious, with chronic pain as well as clinical features of a mixed sensorimotor dysfunction.
PATHOGENESIS

Focal Neuropathies

Mononeuropathies are caused by microscopic vasculitis and subsequent ischemia or infarction of nerve. Focal ischemia results in segmental demyelination followed by remyelination. In diabetic patients, remyelination is defective and delays the repair of focal deficits. Delayed remyelination may be a consequence of diabetes-induced Schwann cell dysfunction. However, in most cases, recovery does occur because adjacent fascicles take over the function of the damaged ones.

It is not clear why diabetic nerves are more susceptible to entrapment syndromes. Upton and McComas suggested that serial constrictions impairing axoplasmic flow may combine to cause nerve dysfunction and that diabetic patients who have severely impaired axoplasmic flow are more frequent. It was thought that mechanical rather than microvascular factors account for pathologic features of compressed nerves. However, the increased prevalence may be related to repeated undetected trauma, susceptibility of diabetic nerves to injury, accumulation of AGEs, or accumulation of fluid or edema within the confined space of the carpal tunnel.
Distal Symmetrical Polyneuropathy

For a detailed discussion of the different theories on pathogenesis of DSPN, the reader is referred to several excellent recent reviews. However, we do review the principal theories on pathogenesis. DSPN is a heterogeneous disease with widely varying pathology, suggesting differences in pathogenic mechanisms for the different clinical syndromes. Recognition of the clinical homologue of these pathologic processes is the first step in achieving the appropriate form of intervention. Figure 31-32 summarizes our current view of the pathogenesis of DSPN. This figure depicts multiple causes, including metabolic, vascular, autoimmune, oxidative stress, and neurohormonal growth factor deficiency.

Metabolic Hypothesis of Nerve Damage

Although there is increasing evidence that the pathogenesis of DSPN consists of several mechanisms, the prevailing theory is that persistent hyperglycemia is the primary factor. Persistent hyperglycemia (or glucose toxicity) or insulin deficiency may precipitate metabolic or vascular events. Metabolic defects include alteration of the poloy or sorbitol pathway, abnormalities in lipid metabolism, deficiencies of dihomo-γ-linolenic acid (GLA) and N-acetyl-l-carnitine (which are significant in diabetes), glycation or AGE formation, increased oxidative stress, and diabetes mellitus-induced growth factor defects. The results of the DCCT endorse the importance of glycemic control in preventing neuropathy. However, it is not unusual that metabolic factors can account for all patients with neuropathy or for the heterogeneity of the clinical syndromes.

Immune Hypothesis of Nerve Damage

Data from several large epidemiologic studies show that neurallogic signs precede the diagnosis of diabetes in up to 11% of diabetes patients. The neuropathogenic process in some patients with diabetes mellitus may be independent of hyperglycemia. A number of autoimmune phenomena have been described in these patients that might induce immune system responses, including antiphospholipid antibodies (PLAs), a family of closely related immunoglobulins that interact with one or more negatively charged phospholipids (constituents of nervous tissues), and sera with high titers of IgG-PLA that inhibit cell growth and differentiation in a neoblastoma cell line. Anti-PLAs have been found in 88% of a diabetic population with neuropathy compared with 32% in diabetes mellitus patients without apparent neurologic complications and 2% in the general population. Because PLAs are associated with a tendency to vascular thrombosis, their presence may provide a link between the immune and vascular theories of causation of neuropathy.

Autoantibodies to the gangliosides, siaio- and asialo-GM1, have been described in diabetes mellitus patients with neuropathy characterized by a slight emphasis on a motor deficit with electrophysiological signs of demyelination. It is argued that anti-GM1 antibodies are not pathogenic but passively reflect cellular destruction. A number of observations, however, suggest that they have pathogenic potential. Indeed, there may be differences in responsiveness based on the distribution of these antibodies among the immunoglobulin classes. As yet, there is no known autoimmune mechanism in the pathogenesis of the disease, but our understanding of autoimmune neuromopathy is constantly being fueled by new evidence. This area of research in diabetes promises to be exciting and fruitful.

Microvascular Hypothesis

The peripheral and autonomic nervous system has an important role in the control and regulation of microvascular function. As a corollary, microvascular perfusion is essential for the integrity of nerves. Microvascular insufficiency due to impaired vasoconstriction and vasodilation to various stimuli has been proposed as a possible cause of DSPN by a number of investigators. The interest in microvascular derangements in DSPN arises from studies suggesting that absolute or relative ischemia may exist in the nerves of diabetes mellitus subjects due to altered function of the endoneural and epineurial blood vessels. Histopathologic studies show the presence of different degrees of endoneurial and epineurial microvasculopathy, mainly vessel basement membrane thickening and obstruction of vasa nervorum.

A number of functional disturbances are found in the microvasculature of the nerves of patients with diabetes mellitus. These include decreased neural blood flow, increased vascular resistance, decreased neural P0.2 and altered vascular permeability characteristics, such as loss of the anionic charge barrier and decreased charge selectivity. Decreased neural blood flow and increased vascular resistance in diabetes mellitus may result from alterations in microvascular reactivity, such as impaired dilator responses to substance P, calcitonin gene-related peptide (CGRP), and reactive hyperemia. Vasomotion, the rhythmic contraction exhibited by arterioles and small arteries, is disordered in diabetes mellitus patients, and the warm thermal sensory threshold correlated significantly with the mean amplitude of vasomotion. This indicates an interaction between C-fiber function and vasomotion, but it is not clear whether the neurologic deficit precedes or follows the loss of normal vascular motility.

It also has been shown that abnormalities of cutaneous blood flow correlated with indices of small fiber neuropathy. Metabolic dysfunction of both central and reflex vasoconstrictor sympathetic nerves, together with abnormal vascular smooth muscle metabolism, may alter vascular tone and increase arteriolar and shunt blood flow, both of which would increase capillary pressure. This might be similar to the mechanism of disordered kidney function in diabetes. In contrast, disordered endothelial and smooth muscle metabolism, resulting in impaired NO generation, or resistance to the vasodilatory actions of NO, would lead to reduced microvascular responses to both flow-mediated vasodilation following ischemia and hyperemia following injury. These defects have now been reported prior to the onset of diabetes mellitus, in family members, and co-segregate with other components of the metabolic syndrome, including hypertension, dyslipidemia with elevated triglyceride levels, and insulin resistance.

Thus, it is possible that ischemia precedes neuropathy or that both conditions are the result of separate processes caused by the same etiologic factors and each accelerates the other's progress; that is, vascular insufficiency causes nerve damage and nerve dysfunction impairs blood flow. Whatever the case, the loss of the neurovascular function reduces the required nutritional delivery to skin and subcutaneous tissue and, coupled with impaired perception, predisposes the limb to injury, ulceration, and infection that may culminate in gangrene.

![Figure 31-32](image-url)
Adult dorsal root ganglia and sympathetic neurons, both of which are affected in DSPN, are dependent on NGF for their maintenance or survival. NGF has been implicated in diverse and widespread activities, including vasodilatation, gut motility, and nociception. Numerous data suggest that a decline in NGF synthesis in diabetes mellitus plays a role in the pathogenesis of DSPN by causing a functional deficit in small fibers. These fibers have a role in pain and thermal sensation. The effect of NGF depletion may be mediated through the down-regulation of neurofilament gene expression or mRNAs that encode the precursor molecules of substance P both shown to be NGF-dependent. Another member of the neurotrophin family, neurotrophin 3, may be important for the survival and function of the large nerve fibers subserving position, vibration, and possibly motor functions.

IGF-I and IGF-II, which are implicated in the growth and differentiation of neurons and IGF receptors, are present in nerve tissues (i.e., neurons, Schwann cells, ganglia) involved in DSPN. IGFs and their binding proteins are regulated by insulin and the glycemic state. One consequence of insulin insufficiency in diabetes mellitus is a reduction in circulating IGF-I concentration. It seems reasonable to hypothesize that abnormal IGF-I and IGF-II metabolism plays a role in some aspect of DSPN. Little is known, however, about the other effects of diabetes mellitus on local expression, synthesis, and transport of these growth factors in nerve tissue.

Laminin, a large heterotrimeric protein present in the basal lamina of nerves, appears to be important in nerve regeneration and its expression. Laminin also exerts antiapoptotic properties against the neurotoxicity of sera of patients with diabetes mellitus.

**Oxidative Stress Hypothesis**

Reactive oxygen species (ROS) can regulate multiple signaling mediators linked with important processes that may involve various components of the nerve, dorsal root ganglia, and the vasa nervorum. These processes include metabolism, immune response, cell-cell adhesion, inflammation, cell proliferation, aging, and cell death. Hyperglycemia generates free radicals primarily through increasing flux through the mitochondrial electron transport chain. Formation of AGEs and interaction with the AGE receptor may also lead to the generation of ROS in some cell types.

A prominent role of these reactive molecules as mediators of cellular processes that lead to endothelial cell dysfunction in diabetes mellitus has been suggested. One of the most extensively investigated redox-sensitive molecular paths is that involving a nuclear factor, NF-B. It is a classic member of the Rel family of transcription factors and is known to regulate a diverse set of cellular functions, such as cell growth, immune response, cell survival and development. There are five members of the family, and they tend to form homo- and heterodimers as well as with other members of the family. High glucose and AGE-mediated activation of NF-B is regarded as a key event in the transformation of the vasculature and accelerated vascular disease as well as smooth muscle dysfunction in diabetes mellitus. It is also potentially reversible with the powerful antioxidant, -lipoic acid.

In summary, DSPN is a heterogeneous disease with widely varying pathology, suggesting differences in pathogenic mechanisms for the different clinical syndromes. Recognition of the clinical homologue of these pathologic processes is the first step in achieving the appropriate form of intervention.
CLINICAL PRESENTATION AND TREATMENT OPTIONS

The spectrum of clinical neuropathic syndromes described in patients with diabetes mellitus includes dysfunction of almost every segment of the somatic peripheral and autonomic nervous system \[568\] \[563\] (Fig. 31-33). Each syndrome can be distinguished by its pathophysiologic, therapeutic, and prognostic features.

### Focal Neuropathies

#### Mononeuropathies

Mononeuropathies occur primarily in the older population, their onset is generally acute and associated with pain, and their course is self-limiting, resolving within 6 to 8 weeks. These are due to vascular obstruction after which adjacent neuronal fascicles take over the function of those infarcted by the clot. \[563\]


**Treatment**

Treatment is predominantly symptomatic for pain. If there is weakness such as of the facial muscles, physical therapy and electrical stimulation may be necessary to prevent the weakness from becoming permanent.

#### Entrapment Syndromes

Entrapment syndromes that start slowly, progress, and persist without intervention must be distinguished from mononeuropathies. Common entrapment sites in diabetes mellitus patients involve median, ulnar, radial, femoral, and lateral cutaneous nerves of the thigh, peroneal nerves, and medial and lateral plantar nerves. Entrapment syndromes are found in one third of patients with diabetes. For example, carpal tunnel syndrome occurs twice as frequently in people with diabetes mellitus compared with a normal healthy population. It is important, therefore, to elicit a detailed history of the distribution of pain and weakness and to perform the equivalent of Tinel's test at various levels of entrapment. If recognized, the diagnosis can be confirmed by electrophysiologic studies.

**Treatment**

The mainstays of nonsurgical treatment are resting the joint traversed by the nerve, aided by the placement of splints in a neutral position for day and night use; diuretics to reduce edema; addition of nonsteroidal anti-inflammatory drugs (NSAIDs), and steroids and local anesthetic injections. Surgical treatment consists of sectioning the constrictive tendon sheath. The decision to proceed with surgery is based on many considerations, including severity of symptoms, appearance of motor weakness, and failure of nonsurgical treatment.
Diffuse Neuropathies

Proximal Motor Neuropathies

For many years, proximal motor neuropathy has been considered to be a component of diabetic neuropathy. Its pathogenesis was not understood, and its treatment was neglected with the anticipation that the patient would eventually recover, albeit over a period of some 1 to 2 years, suffering considerable pain, weakness, and disability. The condition is known by a number of synonyms: proximal neuropathy, femoral neuropathy, diabetic amyotrophy, and diabetic neuropathic cachexia. Proximal motor neuropathy can be clinically identified based on the following common features:

1. Primarily affects the elderly.
2. Gradual or abrupt onset.
3. Begins with pain in the thighs and hips or buttocks, followed by significant weakness of the proximal muscles of the lower limbs with inability to rise from the sitting position (positive Gower's maneuver).
4. Begins unilaterally and spreads bilaterally.
5. Coexists with DSPN.
6. Is characterized by spontaneous or percussion-provoked muscle fasciculation.

Proximal motor neuropathy is now recognized as being secondary to a variety of causes unrelated to diabetes mellitus but that occur more frequently in patients with diabetes mellitus than in the general population. The condition includes patients with chronic inflammatory demyelinating polyneuropathy (CIDP), mononuclear gangliopathy, circulating GM1 antibodies and antibodies to neuronal cells, and inflammatory vasculitis. It was formerly thought to resolve spontaneously in 1.5 to 2 years, but now, if found to be immune-mediated, it can resolve within days of initiation of immunotherapy. The condition is readily recognizable clinically with prevailing weakness of the iliopsoas, obturator, and adductor muscles, together with relative preservation of the gluteus maximus and minimus and hamstrings.

Affected patients have great difficulty rising out of chairs unaidered and often use their arms to assist themselves. Heel or toe standing is surprisingly good. In the classic form of diabetic amyotrophy, axonal loss is the predominant process and the proximal motor neuropathy coexists with DSPN.

Electrophysiologic evaluation reveals lumbosacral plexopathy. In contrast, if demyelination predominates and the motor deficit affects proximal and distal muscle groups, the diagnosis of CIDP, mononuclear gangliopathy of unknown significance, and vasculitis should be considered. Biopsy of the obturator nerve reveals deposition of immunoglobulin, demyelination, and inflammatory cell infiltrate of the vasa nervorum. Cerebrospinal fluid protein content is high, and the lymphocyte count is elevated.

Treatment options include intravenous immunoglobulin for CIDP, plasma exchange for mononuclear gangliopathy of unknown significance, and vasculitis should be considered. Biopsy of the obturator nerve reveals deposition of immunoglobulin, demyelination, and inflammatory cell infiltrate of the vasa nervorum. Cerebrospinal fluid protein content is high, and the lymphocyte count is elevated.

Distal Symmetrical Polyneuropathy

Distal symmetrical polyneuropathy (DSPN) is the most common and widely recognized form of diabetic neuropathy. The onset is usually insidious but occasionally is acute, following stress or initiation of therapy for diabetes mellitus. DSPN may be either sensory or motor and involve small nerve fibers, large nerve fibers, or both.

Small-fiber dysfunction usually occurs early and often is present without objective signs or electrophysiologic evidence of nerve damage. It is manifested by early lower limb symptoms of pain and hyperalgesia in the lower limbs, followed by a loss of thermal sensitivity and reduced light touch and pinprick sensation. DSPN may be accompanied by loss of cutaneous nerve fibers that stain positive for the neuronal antigen protein gene product (PGP) 9.5 (Fig. 31-34) (Figure Not Available) as well as impaired neurovascular blood flow. Small-fiber neuropathies, however, can present in a variety of ways.

Small-Fiber Neuropathy

Acute Painful Polyneuropathy

In some patients, a predominantly small-fiber neuropathy develops, manifested by pain and paresthesias early in the course of diabetes mellitus. It may be associated with the onset of insulin therapy and has been termed insulin neuritis. By definition, it has been present for less than 6 months. Symptoms often are exacerbated at night and are manifested in the feet more than the hands. Spontaneous episodes of pain can be severely disabling. The pain varies in intensity and character. In some patients, the pain has been variably described as burning, lancinating, stabbing, or sharp. Paresthesias or episodes of distorted sensation, such as pins and needles, tingling, coldness, numbness, or burning, often accompany the pain.

The lower legs may be exquisitely tender to touch, with any disturbance of the hair follicles resulting in excruciating pain. Because pain can be exacerbated by repeated contact of the lower limbs with foreign objects, even basic daily activities, such as sitting at a desk, may be disrupted. Pain often occurs at the onset of the disease and is often worsened by initiation of therapy with insulin or sulfonylureas.

It may be associated with profound weight loss and severe depression, termed diabetic neuropathic cachexia. This syndrome occurs predominantly in male patients and may occur at any time in the course of both type 1 and type 2 diabetes mellitus. It is self-limiting and invariably responds to simple symptomatic treatment. Conditions such as Fabry’s disease, amyloid, human immunodeficiency viral infection, heavy metal poisoning (e.g., as with arsenic), and excess alcohol consumption should be excluded. It does overlap with the idiopathic variety of acute, painful small-fiber neuropathy that is also a diagnosis by exclusion.

Chronic Painful Neuropathy

Another variety of painful polyneuropathy is characterized by an onset occurring later (often years) in the course of diabetes mellitus, in which the pain persists for longer than 6 months and becomes debilitating. This condition may result in tolerance to narcotics and analgesics and, finally, to addiction. It is extremely resistant to all forms of intervention and is most frustrating to both patient and physician.

The mechanism of pain in DSPN is not well understood. Interacting pathophysiologic mechanisms at the peripheral and...
central nervous system may be responsible for initiation and maintenance of chronic neuropathic pain (Fig. 31-36). Hyperglycemia may be a factor in lowering the pain threshold. There is a sequence in DSPN, beginning when nerve function is good and there is no pain, hyperalgesia, hyperesthesia, or allodynia. Progression of the condition results in nerve dysfunction and pain, increased sensitivity to painful stimuli, and allodynia.

**Chronic Small-Fiber Neuropathy**

Disappearance of these symptoms may not necessarily reflect nerve recovery but rather nerve death. When patients volunteer the "apparent improvement," progression of the neuropathy must be excluded by careful examination. Pain, however, may persist even with dead nerves. The objective physical features include loss of warm thermal perception, decreased heat pain, cold pain, loss of touch pressure perception, and impairment of blood flow. A foot with these findings is at risk for repeated minor trauma, foot ulceration, infection, and gangrene. Small-fiber neuropathies have profound effects on quality of life and mortality.

**Large-Fiber Neuropathies**

Large-fiber neuropathies may involve sensory nerves, motor nerves, or both (Table 31-6). These tend to be the neuropathies of signs rather than symptoms. Large fibers subserve motor function, vibration perception, position sense, and cold thermal perception. Unlike the small nerve fibers, these are

---

**TABLE 31-5 — Clinical Manifestations of Small-Fiber Neuropathies**

1. Symptoms prominent. Pain is of the C-fiber type. It is burning and superficial and associated with allodynia, i.e., interpretation of all stimuli as painful (e.g., touch).
2. Late in the condition, hypoaesthesia.
3. Defective warm thermal sensation.
4. Defective autonomic function with decreased sweating, dry skin, impaired vasomotion and blood flow, and a cold foot.
5. Remarkable intactness of reflexes, motor strength.
7. Loss of cutaneous nerve fibers using PGP 9.5 staining.
8. Diagnosed clinically by reduced sensitivity to 1.0-g Semmes Weinstein monofilament and pricking sensation using the Waardenberg wheel or similar instrument.
9. Abnormalities in thresholds for warm thermal perception, neuro-vascular function, pain, quantitative sudorimetry, and quantitative autonomic function tests.
10. Risks are foot ulceration and subsequent gangrene.

To repeated midfoot trauma, widening of Lisfranc's joint, and foot deformity that can be relieved by Achilles tendon lengthening. The ataxia and loss of strength increase the susceptibility to falling and fractures. The tendency is increased in aging diabetic patients and requires preventive measures and exercise and strength training.

Most patients with DSPN, however, have a "mixed" variety of neuropathy, with both large-fiber and small-fiber damages. In the case of DSPN, a "glove-and-stocking" distribution of sensory loss is almost universal. Early in the course of the neuropathic process, multifocal sensory loss also might be found. In some patients, severe distal muscle weakness can accompany the sensory loss, resulting in an inability to stand on the toes or heels. Some grading systems use this as a definition of severity.
**DIAGNOSIS AND DIFFERENTIAL DIAGNOSIS OF DISTAL SYMMETRICAL POLYNEUROPATHY**

The 1988 San Antonio conference on DSPN and the 1992 conference of the American Academy of Neurology recommended that at least one parameter from each of the following five categories be measured to classify DSPN:

- Symptom profiles
- Neurologic examination
- Quantitative sensory test (QST)
- Nerve conduction velocity (NCV) study
- Quantitative autonomic function test (QAFT)

The diagnosis of DSPN rests heavily on a careful history for which a number of questionnaires for the neurologic symptom score and the neurologic impairment score have been developed by Boulton, Dyck, Vinik, and others. The initial neurologic evaluation should be directed toward the detection of the specific part of the nervous system affected by diabetes mellitus (Fig. 31-37). Bedside neurologic examination is quick and easy but provides nominal or ordinal measures and contains substantial interindividual and intrindividual variation. The least reliable measure is the neurologic symptom score.

For example, it is useless to measure VPT with a tuning fork other than one that has a frequency of 128 Hz. Similarly, using a 10-g monofilament is good for predicting foot ulceration, as is the Achilles reflex, but both are insensitive to the early detection of neuropathy, and a 1.0-g monofilament increases the sensitivity from 60% to 90%. Sensory function must be evaluated on both sides of the feet and hands if one wants to be sure not to miss entrapment syndromes. Tinel's sign not only is useful for carpal tunnel problems but also can be applied to the ulnar notch, the head of the fibula, and below the medial tibial epicondyle for ulnar, peroneal and medial plantar entrapments, respectively. The QST and QAFT are objective indices of neurologic functional status. Combined, these tests cover vibratory, proprioceptive, light touch, pain, thermal, and autonomic function. Developments of a number of relatively inexpensive devices allow suitable assessment of somatosensory function, including vibration, thermal, light touch, and pain perception. These types of instruments allow for cutaneous sensory functions to be assessed noninvasively, and their measurements are correlated with specific neural fiber function.

**TABLE 31-6 — Clinical Presentation of Large-Fiber Neuropathies**

1. Impaired vibration perception (often the first objective evidence) and position sense.
2. Depressed tendon reflexes.
3. A delta type deep-seated gnawing, dull, like a toothache in the bones of the feet, or even crushing or cramp-like pain.
4. Sensory ataxia (waddling like a duck).
5. Wasting of small muscles of feet with hammer toes (intrinsic minus feet and hands) with weakness of hands and feet.
7. Increased blood flow (hot foot).
8. Risk is Charcot's neuroarthropathy.

function and peripheral skin blood flow induced by autonomic neuropathy. Subclinical neuropathy is diagnosed on the basis of the following:

1. Abnormal electrodiagnostic tests with decreased NCV or decreased amplitudes;
2. Abnormal QST for vibration perception, light touch, thermal warming, and cooling thresholds;
3. QAFT revealing diminished heart rate variation with deep breathing, Valsalva maneuver, and postural testing.

Biopsy of nerve tissue may be helpful for excluding other causes of neuropathy and in the determination of predominant pathologic changes in patients with complex clinical findings as a means of dictating choice of treatment. Skin biopsy has some clinical advantages in diagnosis of small-fiber neuropathies by quantification of PGP 9.5, when all other measures, including electromyography, are negative and there are no objective physical signs. Diabetes mellitus as the cause of neuropathy is diagnosed by exclusion of various other causes of neuropathy.
MANAGEMENT OF DISTAL SYMMETRICAL POLYNEUROPATHY

Once DSPN is diagnosed, therapy can then be instituted with the goal of both ameliorating symptoms and preventing the progression of neuropathy. Successful management of these syndromes may eventually be geared to the individual pathogenic processes. At present, however, control of hyperglycemia and meticulous foot care (Table 31-7) are the mainstays of therapy.

Management Aimed at Pathogenetic Mechanisms

Control of Hyperglycemia

Retrospective and prospective studies have suggested a relationship between hyperglycemia and the development and severity of DSPN. In the Steno trial, followed 4400 diabetic patients over 25 years and showed an increase in prevalence of clinically detectable DSPN from 12% of patients at the time of diagnosis of diabetes mellitus to almost 50% after 25 years. The highest prevalence occurred in those people with poorest diabetes control.

The DCCT research group reported significant effects of intensive insulin therapy on progression of neuropathy. In type 1 diabetes, the prevalence rates for clinical or electrophysiologic evidence of neuropathy were reduced by 64% in those treated by intensive insulin therapy after 5 years of follow-up.

The results of the DCCT study support the necessity for strict glycemic control, but the effect of insulin therapy significantly reduced the prevalence of clinical neuropathy by 61% (7% in intensive insulin therapy group versus 17% in conventional therapy group). In a 12-month study of zenarestat, perception, whereas placebo-treated patients showed deterioration in most of the parameters measured.

In a placebo-controlled, double-blind study of tolrestat, 219 diabetes mellitus patients with symmetrical polyneuropathy, as defined by at least one pathologic finding, were treated for 1 year. Patients who received tolrestat showed significant improvement in autonomic function tests as well as in vibration perception, whereas placebo-treated patients showed deterioration in most of the parameters measured.

ARIs reduce the flux of glucose through the polyol pathway, inhibiting tissue accumulation of sorbitol and fructose and preventing reduction of redox potentials.

In a placebo-controlled, double-blind study of tolrestat, 219 diabetes mellitus patients with symmetrical polyneuropathy, as defined by at least one pathologic finding, were treated for 1 year. Patients who received tolrestat showed significant improvement in autonomic function tests as well as in vibration perception, whereas placebo-treated patients showed deterioration in most of the parameters measured.

ARIs reduce the flux of glucose through the polyol pathway, inhibiting tissue accumulation of sorbitol and fructose and preventing reduction of redox potentials.

In a placebo-controlled, double-blind study of tolrestat, 219 diabetes mellitus patients with symmetrical polyneuropathy, as defined by at least one pathologic finding, were treated for 1 year. Patients who received tolrestat showed significant improvement in autonomic function tests as well as in vibration perception, whereas placebo-treated patients showed deterioration in most of the parameters measured.

Alpha-Lipoic Acid

Lipoic acid (1,2-dithiolane-3-pentanoic acid), a derivative of octanoic acid, is present in food and is also synthesized by the liver. It is a natural cofactor in the pyruvate dehydrogenase complex where it binds acyl groups and transfers them from one part of the complex to another. It is effective in ameliorating both the somatic and autonomic neuropathies in diabetes mellitus.

Lipoic acid is currently undergoing extensive trials in the United States as an antidiabetic agent and as a therapy for DSPN.

Gamma-Linolenic Acid

Linoleic acid, an essential fatty acid, is metabolized to GLA, which serves as an important constituent of neuronal membrane phospholipids and also as a substrate for...
prostaglandin formation, seemingly important for preservation of nerve blood flow. In diabetes mellitus, conversion of linoleic acid to ω-linolenic acid and subsequent metabolites is impaired, possibly contributing to the pathogenesis of DSPN. A multicenter double-blind, placebo-controlled trial using GLA for 1 year demonstrated significant improvements in both clinical measures and electrophysiologic tests.

Aminoguanidine

Animal studies using aminoguanidine, an inhibitor of the formation of AGEs and a free radical scavenger, show improvement in NCV in streptozotocin-induced DSPN in rats. Controlled clinical trials to determine its efficacy in humans have been discontinued because of toxicity. However, successors to aminoguanidine and other drugs hold promise for this approach.

Human Intravenous Immunoglobulin

Immune intervention with human intravenous immunoglobulin (IVIg) has become appropriate in some patients with forms of peripheral DSPN that are associated with signs of antineuronal autoimmunity. Treatment with immunoglobulin is well tolerated and is considered safe, especially with respect to viral transmission. The major toxicity of IVIg has been an anaphylactic reaction, but the frequency of these reactions is now low and confined mainly to patients with immunoglobulin (usually immunoglobulin A) deficiency. Patients may experience severe headache due to aseptic meningitis, which resolves spontaneously. In some instances, it may be necessary to combine treatment with prednisone and/or azathioprine. Relapses may occur requiring repeated courses of therapy.

Neurotrophic Therapy

In animal models of diabetes mellitus, the evidence now suggests that decreased expression of NGF and its receptor trk A reduces retrograde axonal transport of NGF and diminishes support of small unmyelinated neurons and their neuropeptides, such as substance P and CGRP both potent vasodilators. Furthermore, recombinant human NGF (rhNGF) administration restores these neuropeptide levels toward normal and prevents the manifestations of sensory neuropathy in animals.

In a 15-center, double-blind, placebo-controlled study of the safety and efficacy of rhNGF in 250 subjects with symptomatic small-fiber neuropathy, rhNGF improved the neurologic impairment score of the lower limbs as well as small nerve fiber function cooling threshold (A delta fibers) and the ability to perceive heat pain (C-fiber) compared with placebo. These results were consistent with the postulated actions of NGF on trk A receptors present on small-fiber neurons. This finding led to two large multicenter studies conducted in the United States and the rest of the world. Regrettably, rhNGF was not found to have beneficial effects over and above placebo. The reason for this dichotomy has not been resolved, but this has somewhat dampened the enthusiasm for growth factor therapy of DSPN. Nonetheless, several new agents are able to bring about nerve growth, proliferation, and differentiation in vitro and have neurotrophic potential; these agents are now being evaluated in early phase II studies.
Management Aimed at Symptoms

Pain Control

Control of pain constitutes one of the most difficult management issues in DSPN. In essence, simple measures are tried first [Fig. 31-38]. If no distinction is made for pain syndromes, the numbers needed to treat (NNT) in DSPN to reduce pain by 50% is 1.4 for optimal-dose tricyclic antidepressants, 1.9 for dextromethorphan, 3.3 for carbamazepine, 3.4 for tramadol, 3.7 for gabapentin, 5.9 for capsaicin, 6.7 for selective serotonin reuptake inhibitors, and 10.0 for nortriptyline. If pain is divided according to its derivation from different nerve fiber type (A delta versus C-fiber), however, different types of pain respond to different therapies (see Fig. 31-38), as described next.

C-Fiber Pain

Initially, when there is ongoing damage to the nerves, the patient experiences the pain of the burning, lancinating, dysesthetic type often accompanied by hyperalgesia and allodynia. Because the peripheral sympathetic nerve fibers are also small unmyelinated C-fibers, sympathetic blocking agents (clonidine) may improve the pain. Loss of sympathetic regulation of sweat glands and arteriovenous shunt vessels in the foot creates a favorable environment for bacteria to penetrate, multiply, and wreak havoc with the foot. These fibers use the neuromodulation systems P as their neurotransmitter, and depletion of axonal substance P (capsaicin) often leads to amelioration of the pain. However, when the destructive forces persist, the individual becomes pain free and develops impaired warm temperature and pain thresholds. Disappearance of pain in these circumstances should be hailed as a warning that the neuropathy is progressing.

Capsaicin

Capsaicin is extracted from chili peppers. A simple, cheap mixture is formed by adding 1 to 3 teaspoons of cayenne pepper to a jar of cold cream and applying to the area of pain. The capsaicinoid receptor, or vanilloid receptor (VR1), is present on C and A delta fibers. When it is activated, it produces desensitization or degeneration of the sensory afferents.

VR1 is essential for selective modalities of pain sensation and for tissue injury-induced thermal hyperalgesia. Prolonged application of capsaicin depletes stores of substance P and, possibly, other neurotransmitters from sensory nerve endings. This reduces or abolishes the transmission of painful stimuli from the peripheral nerve fibers to the higher centers. Care must be taken to avoid eyes and genitals, and gloves must be worn. Because of capsaicin’s volatility, it is safer to cover the affected areas with plastic wrap. There is an initial exacerbation of symptoms that is followed by relief in 2 to 3 weeks.

Clonidine

There is an element of sympathetic-mediated C-fibertype pain that can be overcome with clonidine (α2-adrenergic agonist) or phentolamine. Clonidine can be applied topically, but the dose titration may be more difficult. Unresponsive patients are treated as outlined in Figure 31-38.

A Delta Fiber Pain

A delta fiber pain is a more deep-seated, dull, and gnawing ache, which often does not respond to the previously described measures. A number of different agents have been used for pain associated with these fibers with varying success.

Insulin

Continuous intravenous insulin infusion without resort to blood glucose lowering may be useful in some patients. A response with reduction of pain usually occurs within 48 hours, and the insulin infusion can be discontinued. If this measure fails, several medications are available that may abolish the pain.

Tramadol and Dextromethorphan

There are two possible targeted therapies. Tramadol is a nonopioid centrally acting analgesic for use in treating moderate to severe pain. It has recently been reported to provide pain relief in DSPN. Another spinal cord target for pain relief is the excitatory glutaminergic N-methyl-D-aspartate (NMDA) receptor. Blockade of NMDA receptors is believed to be one mechanism by which dextromethorphan exerts analgesic efficacy. An accomplished pharmacist can procure a sugar-free solution of dextromethorphan.

Antidepressants

Clinical trials have focused on interrupting pain transmission using antidepressant drugs that inhibit the reuptake of norepinephrine or serotonin. This central action accentuates the effects of these neurotransmitters in activation of endogenous pain-inhibitory systems in the brain that modulate pain transmission cells in the spinal cord. Side effects, including dysautonomia and dry mouth, can be troublesome. Switching to nortriptyline may lessen some of the anticholinergic effects of amitriptyline.

Carbamazepine

Several double-blind, placebo-controlled studies have demonstrated carbamazepine to be effective in the management of pain in DSPN. Toxic side effects may limit its use in some patients. However, it is useful for those patients with lightning or shooting pain.

Gabapentin

Gabapentin, which is structurally related to the neurotransmitter -aminobutyric acid, is an effective anticonvulsant whose mechanism is not well understood yet holds additional promise as an analgesic agent in painful neuropathy. In a multicenter study in the United States, gabapentin monotherapy appeared to be...
efficacious for the treatment of pain and sleep interference associated with DSPN. It also exhibited positive effects on mood and quality of life.

Transcutaneous Electrical Nerve Stimulation (TENS)

TENS occasionally may be helpful and certainly represents one of the more benign therapies for painful neuropathy. Care should be taken to move the electrodes around to identify sensitive areas and obtain maximal relief.

Analgesics

Analgesics are rarely of much benefit in the treatment of painful neuropathy, although they may be of some use on a short-term basis for some of the self-limited syndromes, such as painful diabetic third nerve palsy. Use of narcotics in the setting of chronic pain generally is avoided because of the risk of addiction.
MANAGEMENT OF LARGE-FIBER NEUROPATHIES

Patients with large-fiber neuropathies are uncoordinated and ataxic. As a result, they are more likely to fall than are non-neuropathic age-matched people. It has been demonstrated that high-intensity strength training in older people increases muscle strength in a variety of muscles. More important, the strength training results in improved coordination and balance quantifiable with backward tandem walking. Thus, it is vital to embark on a program of strength training and improvement of balance.

<table>
<thead>
<tr>
<th>TABLE 31-8 -- Management of Large-Fiber Neuropathies</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Gait and strength training.</td>
</tr>
<tr>
<td>2. Pain management as detailed in text.</td>
</tr>
<tr>
<td>3. Orthotics should be fitted with proper shoes for the deformities.</td>
</tr>
<tr>
<td>4. Tendon lengthening for Achilles tendon shortening.</td>
</tr>
<tr>
<td>5. Bisphosphonates may be given for osteopenia.</td>
</tr>
<tr>
<td>6. Surgical reconstruction and full-length casting as necessary.</td>
</tr>
</tbody>
</table>
AUTONOMIC NEUROPATHIES

Diabetic autonomic neuropathy may involve any system in the body. Involvement of the autonomic nervous system can occur as early as the first year after diagnosis, and major manifestations are cardiovascular, gastrointestinal, and genitourinary system dysfunction (Table 31-9). Reduced exercise tolerance, edema, orthostatic supine or nocturnal hypertension, and intolerance to heat due to defective thermoregulation are a consequence of autonomic neuropathy.

Common peripheral autonomic function tests are the quantitative sudomotor axon reflex test and the sympathetic skin response, which in most hands are not reliable and sufficiently variable to be of no use clinically. Defective blood flow in the small capillary circulation is found with decreased responsiveness to mental arithmetic, cold pressor, hand grip, and heating. The defect is associated with a reduction in the amplitude of vasomotion and resembles premature aging. There are differences in the glabrous and hairy skin circulations. In hairy skin, a functional defect is found prior to the development of neuropathy and is correctable with antioxidants. The clinical counterpart is a dry, cold skin; loss of sweating; and development of fissures and cracks that are portals of entry for organisms leading to infectious ulcers and gangrene. Silent myocardial infarction, respiratory failure, amputations, and sudden death are hazards for diabetic patients with cardiac autonomic neuropathy. Therefore, it is vitally important to make this diagnosis early so that appropriate intervention can be instituted.

Management of Autonomic Neuropathy

Prevention and Reversibility of Autonomic Neuropathy

It is now become clear that strict glycemic control: a stepwise, progressive management of hyperglycemia, lipid levels, and BP; and the use of antioxidants and ACE inhibitors reduce the odds ratio for autonomic neuropathy to 0.32 (Fig. 31-39). It has also been shown that mortality is a function of loss of beat-to-beat variability with myocardial infarction. This mortality can be reduced by 35% with acute administration of insulin.

Kendall and coworkers reported that successful pancreas transplantation improves epinephrine response and normalizes hypoglycemia symptom recognition in patients with longstanding diabetes mellitus and established autonomic neuropathy. Burger and associates showed that a reversible metabolic component of cardiac autonomic neuropathy exists in patients in the early stages of the neuropathy.

TABLE 31-9 – Clinical Features of Autonomic Neuropathies

<table>
<thead>
<tr>
<th>Cardiovascular</th>
<th>Resting tachycardia</th>
<th>Orthostatic hypotension</th>
<th>Silent myocardial infarction, congestive heart failure, and sudden death</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gastrointestinal</td>
<td>Gastroparis</td>
<td>Diarrhea, constipation</td>
<td></td>
</tr>
<tr>
<td>Genitourinary</td>
<td>Bladder dysfunction</td>
<td>Erectile dysfunction</td>
<td></td>
</tr>
<tr>
<td>Peripheral</td>
<td>Gustatory sweating</td>
<td>Pupillary abnormalities</td>
<td>Disturbed neurovascular flow</td>
</tr>
<tr>
<td>Metabolic</td>
<td>Edema</td>
<td>Hypoglycemia unawareness, hypoglycemia unresponsiveness</td>
<td></td>
</tr>
</tbody>
</table>

and associates reported that successful pancreas transplantation improves epinephrine response and normalizes hypoglycemia symptom recognition in patients with longstanding diabetes mellitus and established autonomic neuropathy. Burger and associates showed that a reversible metabolic component of cardiac autonomic neuropathy exists in patients in the early stages of the neuropathy.

Postural Hypotension

The syndrome of postural hypotension is posture-related dizziness and syncope. Patients who have type 2 diabetes mellitus and orthostatic hypotension are hypovolemic and have sympathoadrenal insufficiency; both factors contribute to the pathogenesis of orthostatic hypotension. Postural hypotension in the patient with diabetic autonomic neuropathy can present a difficult management problem. Elevating the BP in the standing position must be balanced against preventing hypotension in the supine position.

Supportive Garments

Whenever possible, attempts should be made to increase venous return from the periphery by means of total body stockings. However, leg compression alone is less effective, presumably reflecting the large capacity of the abdomen relative to the legs. Patients should be instructed to put the stockings on while lying down and to avoid removing them until returning to the supine position.

Drug Therapy

Some patients with postural hypotension may benefit from treatment with 9-fluorohydrocortisone. Unfortunately, symptoms do not improve until edema occurs, and there is a significant risk for development of CHF and hypertension. If fluorohydrocortisone does not work satisfactorily, various adrenergic agonists and antagonists may be used. If the adrenergic receptor status is known, therapy can be guided to the appropriate agent. Metoclopramide may be helpful in patients with dopamine excess or increased sensitivity to dopaminergic stimulation. Patients with 2-adrenergic receptor excess may respond to the 2-antagonist yohimbine. Those few patients in whom receptors are increased may be helped with propranolol.

2-Adrenergic receptor deficiency can be treated with the 2-agonist, clonidine, which in this setting may paradoxically increase BP. One should start with small doses and gradually
increase the dose. If the preceding measures fail, midodrine, an \( \alpha \)-adrenergic agonist, or dihydroergotamine in combination with caffeine may help. A particularly refractory form of postural hypotension occurs in some patients postprandially and may respond to therapy with octreotide given subcutaneously in the mornings.

**Gastrophy**

Gastrointestinal motor disorders are common and widespread in patients with type 2 diabetes mellitus, regardless of symptoms. \(^{(10)}\) and there is a poor correlation between symptoms and objective evidence of functional or organic defects. The first step in management of diabetic gastroparesis consists of multiple, small feedings. The amount of fat should be decreased because it tends to delay gastric emptying. Maintenance of glycemic control is important. \(^{(33)}\) Metoclopramide may be used; when gastroparesis is severe, it is important to administer it intravenously, in liquid form, or as a suppository. When emptying becomes normal, the oral form may be resumed. Tachyphylaxis is common, and withdrawal from the drug periodically restores responsiveness.

Domperidone has been effective in some patients, \(^{(10)}\) although probably no more so than metoclopramide. Erythromycin, given as either a liquid or a suppository, may also be helpful. Erythromycin acts on the motilin receptor, “the sweeper of the gut,” and shortens gastric emptying time. \(^{(10)}\) If medications are unsuccessful and severe gastroparesis persists, jejunostomy placement into normally functioning bowel may be needed.

**Enteropathy**

Enteropathy involving the small bowel and colon can produce both chronic constipation and explosive diabetic diarrhea, making treatment of this particular complication difficult.

**Diet**

Patients with poor digestion may benefit from a gluten-free diet. The physician should be aware that certain fibers in the neuropathic patient can lead to bezoar formation because of bowel stasis in the gastroparetic or constipated state.

**Antibiotics**

Stasis of bowel contents with bacterial overgrowth may contribute to the diarrhea. Treatment with broad-spectrum antibiotics is the mainstay of therapy, including tetracycline or trimethoprim and sulfamethoxazole. Metronidazole appears to be the most effective and should be continued for at least 3 weeks.

**Cholestyramine**

Retention of bile may occur and can be highly irritating to the gut. Chelation of bile salts with cholestyramine 4 g three times daily mixed with fluid may offer relief of symptoms.

**Cholestyramine plus Atropine**

Diphenoxylate plus atropine may help control diarrhea; toxic megacolon can occur, however, and extreme care should be used.

**Somatostatin**

In refractory cases, small doses of octreotide can be helpful in controlling diarrhea.

**Cystopathy**

Patients with neurogenic bladder should be instructed to palpate the bladder, and if they cannot initiate micturition when the bladder is full, they should be advised to use Credé’s maneuver to start the flow of urine. Parasympathomimetics such asbethanechol are sometimes helpful, although often they do not help fully empty the bladder. Extended sphincter relaxation can be achieved with an \( \alpha \)-adrenergic blocker, such as doxazosin. \(^{(28)}\) Self-catheterization can be particularly useful in this setting, with the risk of infection generally being low.

**Sildenafil** (Viagra), an orally active selective inhibitor of phosphodiesterase type 5, can be used in diabetic ED. \(^{(57)}\) Most diabetic patients require 50 or 100 mg, and treatment should never be started without an evaluation of cardiac function. Patients with diabetic neuropathy are notorious for silent myocardial ischemia and poor ejection fractions. \(^{(57)}\) and cardiovascular function must be evaluated before the prescription is written. The concomitant use with nitrates or nitrates is contraindicated because of the profound hypotension that may occur. \(^{(57)}\)

**Gustatory Sweating**

Gustatory sweating is more common than previously believed, and topically applied glycopyrrolate, an antimuscarinic compound, is effective treatment in reducing both the severity and frequency of sweating of the head and neck region while eating food that triggers this reflex. \(^{(57)}\)
DIABETIC HEART DISEASE

IMPACT OF CARDIOVASCULAR DISEASE IN PATIENTS WITH DIABETES MELLITUS

It is widely underappreciated that CVD is the leading cause of mortality in patients with diabetes mellitus. Approximately 75% of the cardiovascular deaths attributed to diabetes are directly related to coronary artery disease. The economic burden of CVD in patients with diabetes far exceeds that of ESRD. Despite the well-recognized benefits of tight glycemic control in reducing the risk of microvascular complications of diabetes, as evidenced from the results of large randomized clinical trials, a similar corollary for macrovascular complications has not been firmly established.

As the treatment of diabetic microvascular disease improves, an even greater increase in the incidence and prevalence of diabetic macrovascular disease can be expected. The steady increase in both the prevalence of overt diabetes and impaired glucose tolerance in the United States further highlights the need for a comprehensive and aggressive approach to cardiovascular risk factor management in these patients. The "graying" of populations in most Westernized countries will also likely lead to a substantially larger social and economic burden from the cardiovascular complications of diabetes.
CORONARY HEART DISEASE MORBIDITY AND MORTALITY IN TYPE 1 AND TYPE 2 DIABETES

The last decades have been witness to substantial declines in coronary heart disease (CHD) mortality in the general population in the United States. However, significantly less improvement in CHD mortality has been seen in diabetic men and women during this same period. More than 90% of all patients with diabetes have type 2 diabetes, and it is this population (most middle-aged and elderly) that has been evaluated in the majority of studies of CHD risk. In these studies, the excess morbidity and mortality associated with diabetes and elevated glucose remained even after adjustment for traditional CHD risk factors.

Data from the Framingham Study showed a twofold to threefold elevation in the risk of clinically evident atherosclerotic disease in patients with type 2 diabetes compared to those without diabetes. Diabetic men in the Multiple Risk Factor Intervention Trial (MRFIT) study had an absolute risk of CHD death more than three times higher than that in the nondiabetic cohort, even after adjustment for established risk factors. Seminal work from Finland showed that patients with type 2 diabetes without a previous myocardial infarction have as high a risk of myocardial infarction over 7 years as nondiabetic patients with a previous myocardial infarction. (Fig. 31-40).

The case fatality rate following a myocardial infarction is also substantially higher in patients with diabetes. In women, diabetes mitigates the cardioprotective effects of the premenopausal period, and women with diabetes have a CHD mortality rate as high as that in diabetic men.

The risk of CHD has been evaluated in small subsets of patients with type 1 diabetes. In the Framingham study, the cumulative coronary artery disease mortality in patients with type 1 diabetes was approximately four times that of nondiabetic patients by age 55 years. Similar to patients with type 2 diabetes, the first deaths related to coronary artery disease in patients with type 1 diabetes generally occur by the fourth decade of life, and the cumulative mortality increases at a similar rate in both groups in the subsequent 20 years. The rise in coronary artery disease mortality with age in patients with type 1 diabetes is substantially higher in patients with nephropathy. In these patients, the risk of coronary artery disease development can be as much as 15 times higher than in patients without persistent proteinuria. Thus, persistent proteinuria is a strong predictor of the development of coronary artery disease in this population. These findings are consistent with a pathophysiologic viewpoint that proteinuria is a marker of generalized vascular damage that in the coronary vasculature emerges as a predisposition to CVD.
AGGREGATION OF TRADITIONAL CORONARY HEART DISEASE RISK FACTORS IN DIABETES MELLITUS

It is now well established that a number of traditional CHD risk factors (e.g., hypertension, dyslipidemia, obesity, insulin resistance) tend to occur together in patients with diabetes. Approximately 50% of patients with diabetes have hypertension, and more than 30% have hypercholesterolemia at the time of diagnosis. As in nondiabetic patients, these risk factors independently predict the risk of CVD mortality. However, even in the presence of one or more concomitant risk factors, diabetes increases the CVD death rate. It also appears that diabetes interacts synergistically with other risk factors to more sharply increase risk as the number of total risk factors increases.

Data from the UKPDS confirm the importance of risk factor aggregation in diabetic patients. In this large population of patients with newly diagnosed type 2 diabetes, the development of coronary artery disease during follow-up was significantly associated with increased concentrations of LDL cholesterol, decreased concentrations of HDL cholesterol, increased levels of Hg A1c and systolic BP, and a history of smoking as measured at baseline.
PLASMA GLUCOSE AND INSULIN RESISTANCE AS INDEPENDENT RISK FACTORS FOR ATHEROSCLEROSIS

Hyperglycemia may be responsible for the high excess risk of CHD in patients with diabetes that cannot be accounted for by the interaction of multiple risk factors alone. This association appears to be graded and continuous, without a clear threshold below which the relationship ends. One study showed that mortality from all causes, CVD, and ischemic heart disease increases progressively across quintiles of fasting blood glucose levels in patients with type 2 diabetes. [Fig. 31-42] Other data suggest a dose-response relationship between hyperglycemia and CVD mortality in diabetes, with patients with the highest levels of fasting blood glucose having a CVD mortality rate almost five times higher than patients with the two lowest levels combined. [730]

The continuum of CVD risk with rising glucose levels has also been identified in patients with type 1 diabetes [731] and in subjects without clinically overt diabetes but with varying levels of glucose tolerance. [732] In addition to its association with clinical outcomes, objective assessment of carotid artery intimal media thickness has highlighted the deleterious effects of hyperglycemia at the level of the vessel wall. [733] [734] [735] [736] [737]

Almost all patients with diabetes and the concomitant CVD risk factors of hypertension, obesity, and dyslipidemia also have insulin resistance. [738] The clustering of these risk factors in a single patient has been characterized as a syndrome, described by a variety of names such as syndrome X, insulin resistance syndrome, and cardiovascular dysmetabolic syndrome. Because insulin resistance typically precedes the development of hyperglycemia, and patients with type 2 diabetes have an elevated risk of CHD at diagnosis, it has been hypothesized that hyperinsulinemia is the underlying link between hyperglycemia and

![Figure 31-41](https://example.com/image.png)

**Figure 31-41** Age-adjusted cardiovascular disease (CVD) death rates by presence of number of risk factors for men with and without diabetes at baseline screened for the Multiple Risk Factor Intervention Trial. In the presence of diabetes the cardiovascular death rate steeply rises at any level of concomitant risk factors. SBP, systolic blood pressure. (From Stamler J, Vaccaro O, Neaton JD, et al. Diabetes, other risk factors, and 12-year cardiovascular mortality for men screened in the Multiple Risk Factor Intervention Trial. Diabetes Care 1993; 16:434-444.)

CVD in these patients [739] This has been supported by data from a number of studies showing that hyperinsulinemia is an independent predictor of CVD risk. Furthermore, in patients spanning the spectrum of glucose tolerance, from normal to hyperglycemic to diabetic, insulin resistance positively correlates with atherosclerosis as assessed by carotid intima-to-media thickness. [740]

Data from a long-term follow-up of nondiabetic men without heart disease at baseline have suggested that endogenous insulin itself, rather than insulin resistance, may directly increase CVD risk. Subjects with clinically overt CHD had significantly elevated fasting insulin levels at baseline compared with subjects without CHD, even after adjustment for the presence of traditional risk factors. [741]
THE ROLE OF GLYCEMIC CONTROL

The UKPDS confirmed the positive association between plasma glucose levels and CHD risk for Hb A1c levels above 6.2% in patients with diabetes. CHD risk increased by 11% with each 1% elevation in Hb A1c (Table 31-10). The question remains, however, whether intensive glycemic control can modify the cardiovascular risk profile of patients with diabetes.

Earlier studies, such as the DCCT and the smaller Veterans Affairs (VA) study, did not show a reduction in cardiovascular end points with intensive metabolic control. However, the conclusions that can be drawn from these studies are limited. The DCCT followed a relatively young (mean age, 27 years) population of patients with type 1 diabetes for more than 6 years. Although the study group was relatively large (n = 1441), at the end of follow-up few events had occurred. The 41% reduction between intensive and conventional therapy in the risk of cardiovascular and peripheral vascular disease was not statistically significant. In the VA study, intensive blood glucose control in patients with type 2 diabetes did not significantly reduce cardiovascular end points. Like the DCCT, this study lacked adequate power to detect a difference in macrovascular events between treatment groups, given the small number of events in each group. In addition, the patient population was small and followed for a relatively short period.

The UKPDS was larger and adequately powered to detect a difference between groups in macrovascular events. In this study, 3867 patients with newly diagnosed type 2 diabetes were randomly assigned to intensive glucose control (diet plus oral therapy or insulin) or to conventional treatment and were followed for 10 years. The study population had relatively few CHD risk factors and a low background rate of CHD. Patients treated with intensive glycemic control had a lower rate of myocardial infarction than did patients treated with conventional treatment, a finding that showed a trend for statistical significance (P = .052). As in the DCCT, intensive therapy in UKPDS significantly improved the rate of microvascular disease.
DYSLIPIDEMIA AND ITS TREATMENT IN PATIENTS WITH DIABETES MELLITUS

Dyslipidemia is the most well-characterized risk factor for increasing atherosclerosis in patients with type 2 diabetes. Moreover, there are a number of features of dyslipidemia in patients with diabetes that are uniquely associated with this metabolic derangement and that appear to increase the predisposition to atherogenesis. Although patients with diabetes tend not to have marked elevations in plasma LDL levels, this lipid fraction differs qualitatively from that in nondiabetic patients with dyslipidemia.

For example, LDL particles in patients with diabetes are generally smaller and more dense than typical LDL particles. These small, dense LDL particles are more susceptible to oxidation, particularly in the setting of poor glucose control. Other evidence suggests that glycation of LDL may be enhanced in diabetes, impairing recognition of the lipoprotein by its hepatoreceptor and extending its half-life. Conversely, levels of the cardioprotective lipid fraction, HDL cholesterol, are decreased in patients with diabetes. The HDL cholesterol of these patients may also be less effective at protecting LDL from oxidative stress, one of the proposed mechanisms for HDL’s cardioprotective effects.

Undoubtedly, the key feature of diabetic dyslipidemia is an increase in the production of VLDL by the liver in response to elevations in FFAs. Although insulin mediates the uptake of FFAs by striated muscle, reducing the levels presented to the liver, insulin resistance produces the opposite effect, increasing the levels of FFAs available to the liver. Dysmetabolic syndrome, with its characteristic abdominal obesity, also increases the delivery of FFAs to the liver. In addition, reduced lipoprotein lipase activity in type 2 diabetes leads to an accumulation of triglyceride-rich lipoproteins in the plasma of these patients. Triglyceride-rich lipoproteins also play a role in the reduced levels of HDL cholesterol, by increasing the transfer of cholesterol from these particles.

A number of landmark trials have proved that lowering lipid levels produces major clinical benefits in terms of reducing cardiovascular events in patients with and without a history of CHD at baseline. These findings have now been extended to the population of subjects with type 2 diabetes and dyslipidemia. For example, even though LDL levels are frequently within the average range in these patients, treatment with hydroxymethylglutaryl coenzyme A reductase inhibitors (statins) has been shown to improve outcomes. In the Cholesterol and Recurrent Events (CARE) trial, diabetic patients treated with pravastatin had a significant 25% reduction in the incidence of CHD death, nonfatal myocardial infarction, coronary artery bypass graft surgery, and revascularization procedures. In the Long-term Intervention with Pravastatin in Ischaemic Disease (LIPID) study, patients with diabetes experienced a 19% reduction in major CHD (fatal CHD and nonfatal myocardial infarction).

In a post hoc subgroup analysis of secondary prevention in a large cohort of patients with diabetes, impaired glucose tolerance, or normal glucose tolerance, simvastatin normalized associated elevations in total cholesterol and triglycerides across the range of glucose values. Treatment also significantly reduced major coronary events and revascularizations in patients with diabetes and major coronary events, revascularizations, and total and coronary mortality in patients with impaired glucose tolerance. Treatment with a fibrinolytic agent may also be beneficial in patients with diabetes, because these agents address the low HDL cholesterol and high triglyceride levels typically associated with diabetes.

In the VA High-Density Lipoprotein Cholesterol Intervention Trial (VA-HIT), men given gemfibrozil, many of whom had a lipoprotein profile characteristic of insulin resistance, had lower rates of coronary events and strokes. A fibrinolytic agent in combination with a statin may be the optimal approach in patients with diabetes and CHD who have hypercholesterolemia in association with elevated triglyceride and reduced HDL cholesterol levels.

The importance of dyslipidemia in contributing to cardiovascular risk in patients with diabetes is reflected in the new guidelines of the National Cholesterol Education Program Adult Treatment Panel (NCEP-ATP III). For the first time diabetes is considered a CHD risk equivalent, meaning that patients with diabetes have a risk of CHD that is similar to that in patients with clinically manifest CHD (i.e., a greater than 20% risk of an event in the following 10 years). In addition, the presence of multiple CHD risk factors, dysmetabolic syndrome, and mixed hyperlipidemia (high triglyceride and low HDL cholesterol levels) all should be taken into account when estimating a patient’s global risk.

Diabetic patients are candidates for cholesterol-lowering therapy if the LDL cholesterol level is higher than 3.36 mmol/L (130 mg/dL) (with the goal of reducing LDL cholesterol to < 2.57 mmol/L [100 mg/dL]), although many clinicians consider it prudent to approach therapy more aggressively by instituting drug treatment if the LDL cholesterol level is higher than 2.57 mmol/L (100 mg/dL). Studies are under way to determine whether even lower goals (e.g., < 2.07 mmol/L [80 mg/dL]) provide additional benefit. However, regardless of the drug regimen employed, patients with diabetes should maintain light glycemic control, which in itself can help reverse the dyslipemic profile prevalent in diabetes. As with all patients, lifestyle modification, including weight reduction and regular exercise, remains an important cornerstone of treating dyslipidemia in patients with diabetes.
SIGNATURE FEATURES AND TREATMENT OF HYPERTENSION IN DIABETIC PATIENTS

It has been estimated that up to 50% of patients with newly diagnosed diabetes also have high BP. As with dyslipidemia, hypertension interacts with diabetes to amplify the risk of cardiac mortality (see Fig. 31-41). Although the etiology of hypertension is multifactorial, the insulin-resistant state is one factor postulated to predispose patients to the development of hypertension. In addition to its negative effects on the cardiovascular system, high BP is a key contributor to the development of microvascular disease in diabetes. Based on the guidelines of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (JNC VI), BP should be reduced to be less than 130/85 mm Hg in patients with diabetes and to an even lower goal, less than 125/75 mm Hg, in patients with concomitant renal insufficiency.

Results of the most recent clinical trials underscore the rewards of aggressive treatment of hypertension in patients with diabetes, although none of these studies achieved mean BP reductions to currently recommended targets. In the Systolic Hypertension in the Elderly Study, diuretic-based therapy resulted in a 34% reduction in cardiovascular events in the cohort of patients with diabetes. Use of a long-acting dihydropyridine calcium channel blocker in the Systolic Hypertension in Europe (Syst-Eur) study resulted in substantial reductions in total mortality (55%), cardiovascular mortality (76%), and cardiovascular events (69%) in the diabetic subgroup, greater benefits than were seen in the subgroup without diabetes.

In the Heart Outcomes Prevention Evaluation (HOPE) study, in which almost 40% of patients had diabetes and one other cardiovascular risk factor, ramipril reduced the primary outcome by 24% and total mortality by 25% in this cohort. Even in normotensive patients with diabetes, some benefit was seen, with a 2- to 4-mm Hg drop in BP with ACE inhibitor therapy.

Other rigorously designed studies, such as the UKPDS and the Hypertension Optimal Treatment (HOT) study, suggest even greater benefits from tight BP control in patients with diabetes. In the UKPDS, hypertensive patients with type 2 diabetes achieving BP control to a mean of 144/82 mm Hg showed a 21% reduction in the rate of myocardial infarction compared with patients achieving less tight BP control (154/87 mm Hg). A 44% reduction in the risk of fatal and nonfatal stroke was also seen in the group assigned to tight BP control. Achieving diastolic BP reduction to the low 80s in the HOT was particularly beneficial for patients with diabetes treated with felodipine-based therapy.
ACUTE CORONARY SYNDROMES IN DIABETES MELLITUS

The case fatality rate from myocardial infarction is nearly twice as high in patients with diabetes as in nondiabetic patients. This excess risk is seen both during the acute phase of myocardial infarction and in the early and late postinfarction period. A number of mechanisms have been responsible for worse outcomes in patients with diabetes, including the following:

1. Increased risk of CHF due to maladaptive remodeling of the left ventricle.
2. Increased risk of sudden death due to sympathovagal imbalance as a consequence of autonomic neuropathy.
3. Increased likelihood of early reinfarction due to impaired fibrinolysis.
4. Extensive underlying coronary artery disease.
5. Changes in myocardial cell metabolism, including a shift from glucose oxidation to FFA oxidation, with less generation of ATP at any level of oxygen consumption.
6. Associated cardiomyopathy.

Collective data provide strong evidence that a variety of treatment modalities can improve outcomes from myocardial infarction in patients with diabetes (Table 31-11). In terms of interventions, patients with diabetes experiencing an acute myocardial infarction respond as favorably to fibrinolytic therapy as do nondiabetic patients. In terms of medical therapy, excellent glycemic control is an essential component of overall management. Glucose levels at hospital admission have been independently correlated with early and late mortality after myocardial infarction in both patients with and without diabetes mellitus.

TABLE 31-11 — Management of Acute Coronary Syndrome in Patients with Diabetes

| 1. Excellent glucose control (insulin-glucose infusion) |
| 2. Stop sulfonylureas |
| 3. Add ACE inhibitors within 24 hr |
| 4. Cardioselective beta blockers |
| 5. Fibrinolytic therapy |
| 6. CABG gives better outcome than PTCA |

ACE, angiotensin-converting enzyme; CABG, coronary artery bypass graft; PTCA, percutaneous transluminal coronary angioplasty.

Studies such as the Diabetes and Insulin-Glucose Infusion in Acute Myocardial Infarction (DIGAMI) study have assessed the impact of intensive glycemic control in patients with diabetes during the acute phase of myocardial infarction. Patients in this study were randomized to either intensive insulin therapy (insulin-glucose infusion for 24 hours, followed by subcutaneous insulin injection for 3 months) or standard glycemic control. The intensive insulin regimen lowered blood glucose level during the first hour after admission and at discharge compared with conventional therapy. One-year mortality was significantly reduced with the insulin infusion compared with control, a difference that was maintained after 3.4 years of follow-up.

Although the mechanisms behind this benefit are not entirely clear, experimental data suggest that strict glycemic control may improve myocardial cell metabolism by increasing the availability of glucose as a substrate for ATP generation and reducing the formation of FFAs, thereby shifting cardiac metabolism from FFA oxidation to glycolysis and glucose oxidation. Intensive glycemic control may also reverse the impaired fibrinolysis that is typically seen in patients with diabetes.

Although widely used for glycemic control in patients with type 2 diabetes and proven to prevent microvascular complications, sulfonylureas have been implicated in an increase in cardiovascular mortality, particularly in patients undergoing revascularization for acute myocardial infarction. The UKPDS did not show a deleterious effect of these agents on the incidence of sudden death or myocardial infarction over 10 years of follow-up. Nonetheless, concerns persist surrounding the use of sulfonylureas in the setting of myocardial injury, owing to their blockade of ATP-sensitive potassium channels. In the heart, these channels are involved in ischemic preconditioning and coronary vasodilatation. Not all sulfonylureas have high specificity for vascular or myocardial channels, however, and it is not clear whether these agents as a class pose an increased material risk in patients with diabetes who experience an acute myocardial infarction.

ACE inhibitors dramatically reduce mortality following a myocardial infarction in patients with diabetes, ostensibly through their effects on reducing infarct size and limiting ventricular remodeling. In addition to these hemodynamic benefits, ACE inhibitors may improve outcomes in diabetes by improving endothelial function, improving fibrinolysis, and decreasing insulin resistance.

In a retrospective analysis of the GROPO Italian per lo Studio della Sopravvivenza nell’Infarto Miocardico (GISSI-3) study, lisinopril administration within 24 hours of hospital admission substantially reduced both 6-week and 6-month mortality in patients with diabetes compared with the nondiabetic group. Similarly, a subgroup analysis from the Trandolapril Cardiac Evaluation Study (TRACE) showed that patients with diabetes suffering an anterior myocardial infarction treated with trandolapril had greatly improved outcomes over 5 years compared with patients without diabetes, including a nearly 50% reduction in the risk of sudden death, reinfarction, and progression of CHF.

In contrast to ACE inhibitors, beta blockers are less widely accepted in the treatment of acute coronary syndrome in patients with diabetes, primarily over concerns about their effects in masking the signs of hypoglycemia and inhibiting the metabolic response to hypoglycemia. Older, noncardioselective beta blockers may also have adversely affected the lipid profile, further adding to the concerns of clinicians. More recent data with cardioselective beta blockers suggest these agents have less tendency to negatively impact metabolic indices. Clinical trial data confirm that beta blockers reduce the rates of mortality and reinfarction in patients with myocardial infarction in the presence of diabetes. In fact, their effects in patients with diabetes appear to exceed those seen in nondiabetic patients. A large review of data from more than 45,000 patients, 26% of whom had diabetes, showed that beta blocker therapy was associated with a lower 1-year mortality rate in patients with diabetes than in those without diabetes, with no evidence of an increase in diabetes-related complications.

Postulated mechanisms for the benefit of beta blockers in patients with diabetes include dampening of the sympathetic nervous system overactivity that arises as a consequence of autonomic neuropathy. Beta blockers may also reduce FFA levels and thereby reduce myocardial oxygen requirements. Carvedilol, although noncardioselective, is a beta blocker that decreases insulin resistance and also has antioxidant effects, both of which may be of particular benefit in patients with type 2 diabetes.

Aspirin is a cornerstone of therapy for the primary or secondary prevention of acute coronary syndrome in patients with type 1 and type 2 diabetes who do not have contraindications to its use. Aspirin significantly lowers the risk of myocardial infarction without increasing the risk of vitreous or retinal bleeding, even in patients with retnopathy. Enteric-coated aspirin, 81 to 325 mg/day, is currently recommended by the American Diabetes Association. The benefits of this therapy are likely associated with its effects in reversing the enhanced platelet aggregation evident in patients with either type 1 or type 2 diabetes.

Newer adjunctive therapies, such as the platelet glycoprotein IIb/IIIa receptor antagonists that antagonize platelet action, have also been assessed in patients with diabetes and insulin-glucose infusion (?).
diabetes presenting with unstable angina or non-Q-wave infarction. Overall, these agents appear to work equally well, or perhaps slightly better, in patients with diabetes as in nondiabetic patients.

In the Platelet Receptor Inhibition in Ischemic Syndrome Management in Patients Limited by Unstable Signs and Symptoms (PRISM-PLUS) study, the addition of tirofiban to heparin therapy reduced the 7-day composite end point compared with heparin alone. This effect was greater in patients with diabetes than in patients without diabetes.\[789\]

In one study of patients undergoing percutaneous transluminal coronary angioplasty (PTCA), glycoprotein IIb/IIIa antagonist therapy was associated with fewer acute events but a higher rate of target-vessel revascularization in the long term in the diabetic cohort compared with the nondiabetic cohort.\[790\] In another trial, however, in which stents were used, the rate of target vessel revascularization at 6 months was significantly decreased with the addition of a glycoprotein IIb/IIIa antagonist compared with placebo.\[791\]

Results of the Bypass Angioplasty Revascularization Investigation (BARI) showed that coronary bypass graft surgery provides better outcomes than PTCA in patients with diabetes, possibly as a result of addressing the extensive coronary vascular disease in these patients.\[792\] This study did not employ the use of stents or glycoprotein IIb/IIIa inhibitors, two modalities that, when used together, appear to improve outcomes after PTCA in patients with diabetes.
CARDIOMYOPATHY IN PATIENTS WITH DIABETES MELLITUS

Diabetes is associated with a fourfold increase in the risk for CHF, even after adjustment for other cardiovascular risk factors such as age, BP, cholesterol level, obesity, and history of coronary artery disease. Patients with diabetes experience higher rates of CHF than do nondiabetic patients following an acute myocardial infarction, regardless of the size of the infarct zone. These findings suggest that diabetes itself causes deleterious effects on the myocardium, leading to poorer outcomes.

A number of key structural, functional, and metabolic factors have been implicated in the increased risk of maladaptive remodeling that leads to CHF in diabetes. For example, silent myocardial infarctions, evidence of which is found in up to 40% of patients with diabetes presenting with a clinically apparent myocardial infarction, may lead to unrecognized regional and global ventricular dysfunction. As many as 50% of patients with diabetes and coronary artery disease have cardiac autonomic neuropathy, which is known to contribute to both systolic and diastolic dysfunction. Like hypertension, diabetes can cause fibrosis of the myocardium and increased collagen deposition. These effects are even more pronounced in patients with coexisting hypertension and diabetes. Enhanced endothelial dysfunction in diabetes has also been described as a pathophysiologic pathway to impaired microvascular perfusion and ischemia.

On a cellular level, both hyperglycemia and insulin resistance have direct negative effects on myocardial metabolism. Depression of myocardial GLUT 4 transporter protein levels in the setting of diabetes and ischemia inhibits glucose entry and glycolysis into the heart. As a result, intracellular metabolism shifts from glycolysis to FFA oxidation, thereby suppressing glycolytic ATP generation, a major source of energy under anaerobic (i.e., ischemic) conditions. The production of oxygen free radicals may also be enhanced in this situation, further depressing myocardial contractile function.

Collectively, these various abnormalities potentiate the characteristic left ventricular remodeling of diabetes, clinically manifested as serial wall motion changes, reduced regional ejection fraction, and increases in end-diastolic and end-systolic volumes.
THE DIABETIC FOOT

Of all the late complications of diabetes, foot problems are probably the most preventable. Thus, Joslin, who wrote in 1934 that “diabetic gangrene is not heaven-sent, but earthborn,” was correct: The development of foot ulceration mostly results from the way we care for our patients or the way in which patients care for themselves.

Increasing interest in the diabetic foot in recent years has resulted in a better understanding of the factors that interact to cause ulceration and amputation. Thus, the neuropathic foot does not spontaneously ulcerate; the combination of insensitivity with other factors, such as deformity and unperceived trauma (e.g., inappropriate footwear), leads to skin breakdown. This increase in the knowledge of pathogenesis should permit the design of appropriate screening programs for risk and preventive education. However, although much progress has been made, this has not yet resulted in a decrease in amputation rates. Indeed, in the VA system, for example, amputation rates are declining, except in diabetic patients. Thus, much research is still needed to implement strategies to reduce ulceration and amputation, and this is particularly required in the fields of behavioral and psychosocial aspects of the diabetic foot.

The year 2000 witnessed the publication of two major texts on the diabetic foot, to which, together with other major review articles, the reader is referred for detailed discussion of this topic.

EPIDEMIOLOGY AND PATHOGENESIS OF DIABETIC FOOT ULCERATION

Foot ulceration is common and occurs in both type 1 and type 2 diabetes. Approximately between 5% and 10% of diabetic patients have had past or present foot ulceration, and 1% have undergone amputation. Diabetes is the most common cause of nontraumatic lower limb amputation in the United States, with rates of 15 times those in the nondiabetic population. A recent study in the United States showed a cumulative incidence of ulceration of 5.8% in 3 years within a large health maintenance organization population of diabetic patients; 15% of ulcer patients required amputation, and, in common with other studies, foot ulcers were associated with increased mortality in the 3-year period.

Pathway to Ulceration

As stated earlier, foot ulceration results from an interaction of a number of component causes, none of which alone is sufficient to cause ulceration but, when combined, complete the causal pathway to skin breakdown. Knowledge of these component causes and their potential to interact facilitates the design of preventive foot care programs.

Diabetic Neuropathy

The three components of neuropathysensory, motor, and autonomicmay contribute to ulceration in the foot. Chronic sensorimotor neuropathy is common, affecting at least a third of older patients in western countries. Its onset is gradual and insidious, and symptoms may be so minimal that they go unnoticed by some patients. Although in some patients uncomfortable, painful, and paresthetic symptoms predominate, other patients may never experience symptoms. Clinical examination usually reveals a sensory deficit in a glove-and-stocking distribution, with signs of motor dysfunction, such as small muscle wasting in the feet and absent ankle reflexes. Thus, although a history of typical symptoms strongly suggests a diagnosis of neuropathy, absence of symptoms does not exclude the diagnosis and must NEVER be equated with a lack of foot ulcer risk. Therefore, assessment of the foot ulcer risk must always include a careful foot examination, whatever the history.

Sympathetic autonomic neuropathy affecting the lower limbs results in reduced sweating, dry skin, and development of cracks and fissures, and, in the absence of large-vessel arterial disease, there may be increased blood flow to the foot with arteriovenous shunting leading to the warm but at-risk foot.

The importance of neuropathy as a contributory cause to foot ulceration was confirmed in a large multicenter study from Europe and North America that reported a 7% annual risk of ulceration in neuropathic patients. Previous evidence had suggested that the risk in non-neuropathic patients is less than 1%.

Peripheral Vascular Disease

In the pathogenesis of ulceration, peripheral vascular disease itself in isolation rarely causes ulceration. However, the common combination of vascular disease with minor trauma may lead to ulceration. Thus, minor injury and subsequent infection increase the demand for blood supply beyond the circulatory capacity, and ischemic ulceration and risk of amputation develop. Early identification of those at risk for peripheral vascular disease is essential, and appropriate investigation involving noninvasive studies, together with arteriography, often leads to bypass surgery to improve blood flow to the extremities. Distal bypass surgery is frequently performed, with good short-term and long-term results in limb salvage. Doppler-derived ankle pressure can be misleadingly high in long-standing diabetes, but the presence or absence of a dorsalis pedis or posterior tibial pulse is the simplest and most reliable indicator of significant ischemia. Long-term results with bypass grafts to the dorsalis pedis artery are excellent. In one large study, graft patency of 82% and limb salvage of 87% were observed at 5 years.
Minor Foot Ulceration

Foot ulceration is most common in those patients with a history of similar problems, and even in experienced diabetic foot clinics, more than 50% of patients with new foot ulcers give a past ulcer history.

Other Diabetic Complications

Patients with retinopathy and renal dysfunction are at increased risk for foot ulceration.

Callus, Deformity, and High Foot Pressures

Motor neuropathy, with imbalance of the flexor and extensor muscles in the foot, frequently results in foot deformity, with prominent metatarsal heads and clawing of the toes [Fig. 31-43]. In turn, the combination of the proprioceptive loss due to neuropathy and the prominence of metatarsal heads leads to increases in the pressures and loads under the diabetic foot. High pressures, together with dry skin, frequently result in the formation of callus under weight-bearing areas of the metatarsal heads. The presence of such plantar callus has been shown in cross-sectional and prospective studies to be a highly significant marker of foot ulcer risk. Conversely, removal of plantar callus is associated with a reduction in foot pressures and thus a reduction in foot ulcer risk. [809]

It is the combination of two or more of the earlier described risk factors that ultimately results in diabetic foot ulceration. In 1999, a North American/United Kingdom collaborative study[814] assessed the risk factors that resulted in ulceration in more than 150 consecutive foot ulcer cases. From this study, a number of causal pathways were identified but the most common triad of component causes was present in 63% of incident ulcers and comprised neuropathy, deformity, and trauma. Edema and ischemia were also common component causes.
PREVENTION OF FOOT ULCERATION AND AMPUTATION

That diabetic foot ulceration is largely preventable is not disputed; small, mostly single-center studies have shown that relatively simple interventions can reduce amputations by up to 80%. Thus, strategies for the earlier identification of those at potential risk of ulceration are required, and education programs that can be adapted for widespread application need to be developed. Because foot ulcers precede more than 80% of amputations, are among the most common causes of hospital admission for patients with diabetes, and account for much morbidity and even mortality, the widespread application of preventive foot care strategies is urgently required.

**TABLE 31-12**  — Wagner Diabetic Foot Ulcer Classification System

<table>
<thead>
<tr>
<th>Grade</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No ulcer, but high-risk foot (e.g., deformities, callus, insensitivity)</td>
</tr>
<tr>
<td>1</td>
<td>Superficial, full-thickness ulcer</td>
</tr>
<tr>
<td>2</td>
<td>Deeper ulcer, penetrating tendons, without bone involvement</td>
</tr>
<tr>
<td>3</td>
<td>Deeper ulcer with bone involvement, osteitis</td>
</tr>
<tr>
<td>4</td>
<td>Partial gangrene (e.g., toes, forefoot)</td>
</tr>
<tr>
<td>5</td>
<td>Gangrene of whole foot</td>
</tr>
</tbody>
</table>


All patients with diabetes, whichever type and irrespective of duration, require regular review and screening of the feet for evidence of risk factors for foot ulceration. At a minimum, such screening should be carried out annually. Of all the long-term complications of diabetes, foot problems and their risk factors are probably the easiest to detect. No expensive equipment is required, and the feet can be examined for evidence of neuropathic and vascular deficits in the office setting using simple equipment. It must be remembered that neuropathy, vascular disease, and even foot ulceration may be the presenting feature of type 2 diabetes, and thus there can be no exception to the rule of screening.

A simple algorithm for the diabetic foot screen is provided in **Figure 31-44**. The most important message to practitioners is to have the patient remove the shoes and socks and to look at the feet for risk factors (e.g., presence of callus, deformity, muscle wasting, and dry skin), all of which are clearly visible on clinical inspection. A simple neurologic examination that might include a neuropathy disability score is recommended. Absence of the ability to perceive pressure from a 10-g monofilament, inability to perceive a vibrating 128-Hz tuning fork over the hallux, and absent ankle reflexes all have been shown to be predictors of foot ulceration.

**The Diabetic Foot Care Team**

Patients identified as being at high risk for foot ulceration should be managed by a team of specialists with interest and expertise in the diabetic foot. The podiatrist generally takes responsibility for follow-up and care of the skin and nails and, together with the specialist nurse or diabetes educator, provides foot care education. The orthotist, or shoe fitter, is invaluable for advising about and sometimes designing footwear to protect the high-risk feet, and these members of the team should work closely with the diabetologist and the vascular and orthopedic surgeons. The importance of wearing special shoes was demonstrated in a study in which 83% of patients wearing regular shoes experienced a recurrence of foot ulcers, whereas only 17% of those wearing special shoes did.  

**TABLE 31-13**  — University of Texas Wound Classification System

<table>
<thead>
<tr>
<th>Grade</th>
<th>Grade 0</th>
<th>Grade 1</th>
<th>Grade 2</th>
<th>Grade 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage A</td>
<td>Preulcer or postulcer lesion</td>
<td>Superficial ulcer</td>
<td>Deep ulcer to tendon/capsule</td>
<td>Wound penetrating bone/joint</td>
</tr>
<tr>
<td></td>
<td>No skin break</td>
<td>+ Infection</td>
<td>+ Infection</td>
<td>+ Infection</td>
</tr>
<tr>
<td>Stage B</td>
<td>+ Infection</td>
<td>+ Infection</td>
<td>+ Infection</td>
<td>+ Infection</td>
</tr>
<tr>
<td>Stage C</td>
<td>+ Ischemia</td>
<td>+ Ischemia</td>
<td>+ Ischemia</td>
<td>+ Ischemia</td>
</tr>
<tr>
<td>Stage D</td>
<td>+ Infection and Ischemia</td>
<td>+ Infection and Ischemia</td>
<td>+ Infection and Ischemia</td>
<td>+ Infection and Ischemia</td>
</tr>
</tbody>
</table>

Patients with risk factors for ulceration require preventive foot care education and frequent review.
CLASSIFICATION OF FOOT ULCERS

Many different classification systems have been reported in the literature, but the one developed by Wagner for grading diabetic foot ulcers has been widely used and accepted. More recently, the University of Texas (UT) group has developed an alternative classification system that, in addition to depth (as in the Wagner system), takes into account the presence or absence of infection and ischemia. A prospective study from 2001 has assessed and compared these two wound classification systems and concludes that the University of Texas scheme is a better predictor of outcome than the older Wagner system.
MANAGEMENT OF DIABETIC FOOT ULCERS

Basic principles of wound healing apply equally to diabetic foot ulcers as to wounds in any other site or condition. Basically, a diabetic foot ulcer will heal if the following three conditions are satisfied:

1. Arterial inflow is adequate.
2. Infection is treated appropriately.
3. Pressure is removed from the wound and the immediate surrounding area.

Although this approach may seem simplistic, failure of healing of diabetic foot ulcers is normally due to neglect of one or more of the following considerations:

First, the most common cause of nonhealing of neuropathic foot ulcers is the failure to address the third principle (off-loading the area). Medical practitioners forget that patients advised not to put pressure over an ulcer find it difficult to adhere to such advice if peripheral sensation is lost or reduced. Pain results in protection of an injured area; the lack of pain permits pressure to be put directly onto the ulcer and results in nonhealing.

Second, the next most frequent error is inappropriate management of infection. Topical applications are generally unhelpful, and if clinical infection is present, it must be treated appropriately (see later).

Third, the presence of peripheral neuropathy, ischemic symptoms may not be typical because of altered pain sensation. The most difficult ulcer to heal is the neuroischemic ulcer, and symptoms and even signs of ischemia may be altered in the diabetic state. Thus, appropriate noninvasive investigation and arteriography are indicated in the nonhealing diabetic foot ulcer where there is any question as to the vascular status.

Fourth, a final reason for slow or nonhealing of a diabetic foot ulcer, recognized more recently, relates to inappropriate débridement of the wound. The ability of patients with neuropathy to walk putting pressure on active ulcer areas leads to an often extensive buildup of callus tissue. Appropriate débridement and removal of all dead and macerated tissues are essential in the local treatment of a diabetic foot ulcer. Indeed, Steed and colleagues have demonstrated that aggressive débridement of neuropathic foot ulcers leads to a more rapid healing of those ulcers compared with those wounds that are inadequately débrided.

The principles of management of neuropathic and neuroischemic foot ulcers are considered next. Although it is not possible to give much detail on the individual stages and grades of ulcers, we refer to both the University of Texas and the Wagner grading systems.

Neuropathic Foot Ulcer without Osteomyelitis (Wagner Grades 1, 2; University of Texas Grades 1a, 2a)

The most important feature in the management of neuropathic foot ulcers that typically occur under weight-bearing areas such as the metatarsal heads and great toe is to provide adequate pressure relief. This is usually achieved by a cast such as a total contact cast (TCC) or a removable Scotch cast boot.

The total contact cast was recognized as the gold standard for off-loading a foot wound in the 1999 consensus statement on diabetic foot wounds by the American Diabetes Association. This endorsement has now been confirmed as correct in a randomized, controlled trial in which Armstrong and colleagues comparing three off-loading techniques and found that the total contact cast was associated with the shortest healing time. When a total contact cast or any other cast device is used, regular removal of the cast is essential because occasionally the cast itself might injure the insensitive skin. Regular débridement of the wound by a podiatrist is also essential.

Theoretically, complete healing of all superficial and neuropathic ulcers should be possible without the need for amputation. In neuropathic ulcers with a good peripheral circulation, there is no indication for antibiotics unless there are clear clinical signs of infection, including prominent discharge, local erythema, and cellulitis. The presence of any of these features in Wagner's grade 1 or 2 ulcers would warrant reclassification in the University of Texas system to 1b or 2b. In such cases, deep wound swabs should be taken and broad-spectrum oral antibiotic treatment started with agents such as amoxicillin-clavulanic acid combination (Augmentin) or clindamycin. Alterations may be required when sensitivity results become available.
Neuroischemic Ulcers (Wagner Grades 1, 2; University of Texas Grades 1C, 1D)

The principles of management of neuroischemic Wagner grades 1 and 2 ulcers are similar to those for neuropathic ulcers with the following important exceptions:

1. Total contact casts are not usually recommended for the management of neuroischemic ulcers, although removable casts and air cast boots may be used.
2. Antibiotic therapy is usually recommended for all neuroischemic ulcers.
3. Investigation of the circulation (including noninvasive assessment and, when required, arteriography with appropriate subsequent surgical management or angioplasty) is indicated.
Osteomyelitis (Wagner Grade 3; University of Texas Grade 3B)

Wagner or University of Texas grade 3 ulcers are deeper and involve underlying bone, often with abscess formation. Osteomyelitis is a serious complication of foot ulceration and may be present in as many as 50% of diabetic patients with moderate to severe foot infections. If the physician can probe to bone when probing a deep ulcer, the presence of osteomyelitis is strongly suggested. Plain radiographs are indicated in any nonhealing foot ulcer and are useful in the diagnosis of osteomyelitis in more than two thirds of patients, although it should be kept in mind that the radiologic changes may be delayed. In difficult cases, further investigation, such as magnetic resonance imaging, bone scans, or an indium (111 In)-labeled white blood cell scan can be useful in the diagnosis of bony infection.

Although the treatment of osteomyelitis is commonly surgical and involves resecting the infected bone, there have been reports of successful long-term treatment with antibiotics that treat the underlying bacterium, most commonly Staphylococcus aureus. Thus, agents such as clindamycin (which penetrates bone well) or flucloxacillin are often used.
Gangrene (Wagner Grades 4, 5)

The presence of gangrene or areas of tissue death is always a serious sign in the diabetic foot. However, localized areas of gangrene, especially in the toes, without cellulitis, spreading infection, or discharge, can occasionally be left to spontaneously "autoamputate." The presence of more extensive gangrene requires (1) urgent hospital admission; (2) treatment of infection, often with multiple antibiotics; (3) control of the diabetes, usually with intravenous insulin; and (4) detailed vascular assessment. It is in this area that the team approach is most important, with close collaboration between the diabetes specialist, the vascular surgeon, and the radiologist.
NEW TREATMENTS FOR FOOT ULCERS

Platelet-Derived Growth Factors

Several controlled trials of becaplermin, a recombinant human PDGF beta chain homodimer, have confirmed the efficacy of this topically applied agent in promoting healing of neuropathic foot ulcers. In a combined analysis of randomized, controlled studies, Smiell and co-workers showed that active treatment was associated with a significant increase in the probability of complete healing compared to placebo gel.

The use of such an agent should probably be reserved for the management of difficult-to-heal neuropathic ulcers that do not respond to standard treatment, such as off-loading or regular débridement. Its widespread use is somewhat limited by its cost; when used selectively, however, this agent can accelerate the wound healing of neuropathic foot ulcers.
Living Human Skin Equivalents

The recent development of living human skin equivalents produced by tissue engineering techniques has produced new possibilities for wound healing therapies and chronic ulcers, such as those caused by venous disease or diabetic neuropathy. However, both living skin equivalents and topically applied growth factors are expensive treatments that must be seen not as a replacement but as an addition to good wound care, which must always comprise adequate off-loading and regular débridement.
Charcot's Neuroarthropathy

Charcot's neuroarthropathy is a rare and disabling condition affecting the joints and bones of the feet. Permissive features for the development of this condition include the presence of severe peripheral neuropathy, together with autonomic dysfunction, with increased blood flow to the foot; the peripheral circulation is usually intact. In the Western world, diabetes is the most common cause of a Charcot's foot and increased awareness of this condition may enable earlier diagnosis and treatment to prevent severe deformity and disability.

The actual pathogenesis of the Charcot process is poorly understood; however, the patient with peripheral insensitivity and autonomic dysfunction with increased blood flow to the foot is vulnerable to frequently unrecognized trauma that may be so trivial that the patient cannot recall the event. Repetitive trauma results in increased blood flow through the bone, increased osteoclastic activity, and remodeling of bone. In certain cases, patients may walk on a fracture that leads to continuing destruction of bones and joints in that area.

Although sometimes difficult to distinguish from osteomyelitis or an inflammatory arthropathy, the patient with neuropathy and a unilateral swollen, hot foot must be considered to have Charcot's foot until proven otherwise.

Charcot's arthropathy can be diagnosed in most patients by plain radiograph and a high index of suspicion. Radiographs may reveal bone and joint destruction, fragmentation, and remodeling, although in the early stages, the radiographic finding may be normal. In such cases, the three-phase bisphosphonate bone scan shows increased bone uptake, although the \(^{111}\)In-labeled bone scan will be negative in the absence of infection.

After diagnosis, management of the acute phase involves immobilization, usually in a total contact cast, and recent evidence suggests that treatment with bisphosphonates, which reduce osteoclastic activity, may reduce swelling, discomfort, and bone turnover markers. \[^8\]

In summary, although rare, Charcot's neuroarthropathy should be suspected in any patient with unexplained swelling and heat in a neuropathic foot, and early intervention with immobilization and possibly bisphosphonate treatment may halt progression that in the untreated state may lead to marked foot deformity and the necessity for local or major amputations.

[^8]: Additional references would be beneficial to support the statements made in this section.
ACKNOWLEDGMENT

For the section "Diabetic Nephropathy," we acknowledge Anthony J. Joseph, M.D., SUNY Downstate Medical Center, New York. M.B. would like to thank his coauthors for their gracious collaboration and for allowing editorial changes to be made in the interest of a more unified chapter.
References


106. Giardino I, Edelstein D, Brownlee M. Nonenzymatic glycosylation in vitro and in bovine endothelial cells alters basic fibroblast growth factor activity: a model for intracellular glycosylation in


References


156. Ganz MB, Seftell A. Glucose-induced changes in protein kinase C and nitric oxide are prevented by vitamin E. Am J Physiol 2000; 278:E146E152.


157.


158.


159.


160.


161.


162.


163.


164.


165.


166.


167.


168.

Yakes FM, Van Houten B. Mitochondrial DNA damage is more extensive and persists longer than nuclear DNA damage in human cells following oxidative stress. Proc Natl Acad Sci USA 1997; 94:514519.

169.


170.


171.


280. The relationship of glycemic exposure (HbA1c) to the risk of development and progression of retinopathy in the diabetes control and complications trial. Diabetes 1996; 45:968983.


References


1574


1575


References


References


The effects of autonomic neuropathy associated with diabetes mellitus on cardiovascular function. 


References


1582


782. Lakshman MR, Reda DJ, Materson BJ, et al. Diuretics and beta-blockers do not have adverse effects at 1 year on plasma lipid and lipoprotein profiles in men with hypertension. Department of
Glucose is an obligate metabolic fuel for the brain under physiologic conditions. In contrast, other organs oxidize fatty acids as well as glucose. Because of this unique dependence on glucose and because it cannot synthesize glucose or store more than a few minutes' supply as glycogen, the brain requires a continuous supply of glucose from the circulation. Facilitated diffusion of glucose from the blood to the brain is a direct function of the arterial plasma glucose concentration. At normal plasma glucose concentrations, the rate of blood-to-brain glucose transport exceeds the rate of brain glucose metabolism. However, as the plasma glucose concentration falls below the physiologic range, blood-to-brain glucose transport becomes limiting to brain energy metabolism and, thus, to survival. Given the immediate survival value of maintenance of the plasma glucose concentration, it is not surprising that physiologic mechanisms that prevent or rapidly correct hypoglycemia have evolved. Indeed, these mechanisms are so effective that hypoglycemia is an uncommon clinical event except in people who use drugs that lower glucose levels (e.g., insulin, sulfonylureas, or alcohol).

Insight into the physiology of glucose counterregulation—the mechanisms that normally prevent or rapidly correct hypoglycemia and its pathophysiology in the context of clinical hypoglycemia, which has been reviewed in detail, has improved the management of hypoglycemia. Nonetheless, major gaps in understanding remain. Both are discussed here.
PHYSIOLOGY OF SYSTEMIC GLUCOREGULATION

Cellular and molecular glucoregulation is discussed in Chapter 29. Glucose metabolism and systemic glucose balance and their regulation are summarized here, with emphasis on the aspects relevant to glucose counterregulation and the prevention of hypoglycemia.

Glucose Metabolism

**Origins and Fates of Glucose**

Glucose is derived from three sources: intestinal absorption that follows digestion of dietary carbohydrates; glycogenolysis, the breakdown of glycogen, which is the polymerized storage form of glucose; and gluconeogenesis, the formation of glucose from precursors including lactate (and pyruvate), amino acids (especially alanine and glutamine), and, to a lesser extent, glycerol (Fig. 32-1).

Although most tissues express the enzyme systems required to synthesize (glycogen synthase) and hydrolyze (phosphorylase) glycogen, only the liver and kidneys express glucose-6-phosphatase, the enzyme necessary for the release of glucose into the circulation, at levels sufficient to permit these organs to contribute to the systemic glucose pool. The liver and kidneys also express the enzymes necessary for gluconeogenesis (including the critical gluconeogenic enzymes pyruvate carboxylase, phosphoenolpyruvate carboxykinase, and fructose-1,6-bisphosphatase).

There are multiple potential metabolic fates for glucose that is transported into cells (external losses are normally negligible) (see Fig. 32-1). It may be stored as glycogen, or may

![Figure 32-1 Schematic representation of glucose metabolism. TCA, tricarboxylic acid.](image)

undergo glycolysis to pyruvate, which can be reduced to lactate, transaminated to form alanine, or converted to acetyl coenzyme A (CoA), which in turn can be oxidized to carbon dioxide and water through the tricarboxylic acid cycle, converted to fatty acids (and stored as triglycerides), or utilized for ketone body (acetocacetate, -hydroxybutyrate) or cholesterol synthesis. Finally, glucose may be released into the circulation. As summarized in the following paragraphs, these outcomes differ in different organs.

**Hepatic (and Renal) Glucose Metabolism**

The liver is remarkably flexible in its role in glucose homeostasis and is the major source of net endogenous glucose production (through glycogenolysis and gluconeogenesis). Under conditions of high glucose output (e.g., fasting), the energy needs of the liver are largely provided by the beta oxidation of fatty acids. Conversely, the liver can also be an organ of net glucose uptake, with glucose stored as glycogen, oxidized for energy, or converted to fat, which can either remain in the liver or be transported to other tissues as very-low-density lipoproteins. The kidneys also produce (through gluconeogenesis) and utilize glucose.

**Glucose Utilization**

Muscle can store glucose as glycogen or metabolize glucose through glycolysis to pyruvate, which either is reduced to lactate or transaminated to form alanine or is oxidized. Lactate (and pyruvate) released from muscle is transported to the liver, where it serves as a gluconeogenic precursor (the Cori or glucose-lactate cycle). It may be stored as glycogen, or may undergo glycolysis to pyruvate, which can be reduced to lactate, transaminated to form alanine, or converted to acetyl coenzyme A (CoA), which in turn can be oxidized to carbon dioxide and water through the tricarboxylic acid cycle, converted to fatty acids (and stored as triglycerides), or utilized for ketone body (acetocacetate, -hydroxybutyrate) or cholesterol synthesis. Finally, glucose may be released into the circulation. As summarized in the following paragraphs, these outcomes differ in different organs.

Although quantitatively less important than muscle, adipose tissue can also use glucose for fatty acid synthesis or formation of glyceral-3-phosphate, which can then esterify fatty acids (derived largely from circulating very-low-density lipoproteins) to form triglycerides. During a fast, adipocytes decrease their glucose utilization and satisfy energy needs from the beta oxidation of fatty acids. Other tissues, such as the formed elements of the blood and the renal medullae, do not have the capacity to decrease glucose utilization during fasting and therefore produce lactate at relatively fixed rates.

As mentioned earlier, glucose is the predominant metabolic fuel used by the brain under most conditions. Glucose undergoes terminal oxidation to carbon dioxide and water in the brain. The brain respiratory quotient is approximately 1.0. Although the adult brain constitutes only about 2.5% of body weight its oxidative metabolism accounts for approximately 25% of the basal metabolic rate under physiologic conditions. However, when ketones are plentiful in the circulation, as during prolonged fasting, they can support the majority of the energy needs of the brain and thus reduce its glucose utilization.

* The author’s work cited has been supported by National Institutes of Health grants R37 DK27085, M01 RR00036, P01 NS06833, P60 DK20579, and T32 DK07120; grants and a fellowship award from the American Diabetes Association; and grants from the Juvenile Diabetes Foundation International.
Systemic Glucose Balance

Normally, rates of endogenous glucose influx into the circulation and those of glucose efflux out of the circulation into tissues other than the brain are coordinately regulated largely by the plasma glucoselowering (regulatory) hormone insulin and the plasma glucosesparing (counterregulatory) hormones glucagon and epinephrine such that systemic glucose balance is maintained, hypoglycemia (as well as hyperglycemia) is prevented, and a continuous supply of glucose to the brain is assured. This is accomplished despite wide variations in exogenous glucose influx (e.g., after feeding versus during fasting) and in glucose efflux (e.g., during exercise versus during rest). Hypoglycemia occurs when rates of glucose appearance in the circulation (the sum of endogenous glucose production, from the liver through both glycogenolysis and gluconeogenesis) (see Fig. 32-1) and, to a lesser extent, from the kidneys through gluconeogenesis, and of exogenous glucose delivery from ingested carbohydrates fail to keep pace with rates of glucose disappearance from the circulation (the sum of ongoing brain glucose metabolism and of variable glucose utilization by tissues such as muscle and fat as well as the liver and kidneys, among others).

Fasting

The postabsorptive state is the interdigestive period that begins approximately 5 to 6 hours after a meal. However, the term is most commonly used to refer to data obtained after a 10- to 14-hour overnight fast. In healthy adults, the physiologic postabsorptive (fasting) plasma glucose concentration is approximately 4.0 to 6.0 mmol/L (72 to 108 mg/dL) with a mean of approximately 5.0 mmol/L (90 mg/dL). In the postabsorptive steady state, rates of glucose production and utilization are equal. They average 12 μmol/kg/minute (2.2 mg/kg/minute) and range from about 10 to 14 μmol/kg/minute (1.8 to 2.6 mg/kg/minute) in healthy adults after an overnight fast. These rates are as much as threefold higher in infants, at least in part because of their greater brain mass relative to their body weight.

Approximately 60% of basal glucose utilization is accounted for by the brain. The remainder is used by glycolyzing tissues, such as the formed elements of the blood and the renal medullae and to some extent muscle and fat. Hepatic glucose production results from both glycogenolysis and gluconeogenesis even after an overnight fast. The liver is the predominant source of net endogenous glucose production after an overnight fast. The kidneys, which both use and produce glucose, contribute little to net glucose production. However, renal, like hepatic, glucose production is regulated. It is suppressed by insulin and stimulated by epinephrine (but not by glucagon). Thus, the common practice of equating endogenous glucose production with hepatic glucose production is not precise.

The importance of gluconeogenesis in providing new glucose and supporting hepatic glycogen stores after an overnight fast becomes apparent when one considers the limited availability of preformed glucose. The glucose pool, namely free glucose in the extracellular fluid and in the cells of certain tissues (primarily in the liver but also small amounts in the kidneys, intestinal mucosa, pancreatic islet cells, brain, and blood cells), is about 83 to 111 mmol (15 to 20 g) in the normal adult. Glycogen that can be mobilized to provide circulating glucose (e.g., hepatic glycogen) contains approximately 390 mmol glucose (70 g), with a range of about 135 to 722 mmol (25 to 130 g). Thus, in an adult of average size, preformed glucose can provide as little as a 3-hour supply of glucose and less than an 8-hour supply on average, even at the diminished rate of glucose utilization that occurs in the postabsorptive state. Clearly, therefore, gluconeogenesis is important for maintenance of the plasma glucose concentration even during an overnight fast.

If fasting is prolonged to 24 to 48 hours, the plasma glucose level declines and then stabilizes, hepatic glycogen content falls to less than 55 mmol (10 g), and gluconeogenesis becomes the sole source of glucose production. Because amino acids are the main gluconeogenic precursors that result in net glucose formation, muscle protein is degraded. Glucose utilization by muscle and fat virtually ceases. As lipolysis and ketogenesis accelerate and circulating ketone levels rise, ketones become a major source of fuel for the brain. Thus, glucose utilization by the brain declines by about half, resulting in a decrease in the rate of gluconeogenesis required to maintain the plasma glucose concentration and hence in diminished protein wasting. After prolonged fasting (40 days), ketones provide an estimated 80% to 90% of the energy used by the brain and renal gluconeogenesis provides up to half of the endogenous glucose production.

Feeding

After a meal, glucose absorption into the circulation is more than twice the rate of postabsorptive endogenous glucose production, depending on the carbohydrate content of the meal and the rate of its digestion and absorption. As glucose is absorbed, endogenous glucose production is suppressed and glucose utilization by liver, muscle, and fat accelerates. Thus, exogenous glucose is assimilated and the plasma glucose concentration returns to the postabsorptive level.

Exercise

Exercise increases glucose utilization (by muscle) to rates that can be severalfold greater than those of the postabsorptive state. Endogenous glucose production normally accelerates to match the utilization so that the plasma glucose concentration is maintained.

From these examples, it is clear that the plasma glucose concentration is normally maintained within a narrow range despite wide variations in glucose flux, a homeostatic feat accomplished by hormonal, neural, and substrate glucoregulatory factors. From a mechanistic perspective, hypoglycemia could result from decreased glucose production, increased glucose utilization, or both.
Glucoregulatory Factors

Hormonal Glucoregulatory Factors

Hormones are the most important glucoregulatory factors, and the regulation of their secretion is complex. Glucose, specifically the plasma glucose concentration, is the most important determinant of the secretion of glucoregulatory hormones, including insulin, glucagon, epinephrine, growth hormone, and cortisol.

Insulin, the dominant hormone in lowering glucose levels, controls the utilization of glucose. Hormones and neural signals are sensed directly by pancreatic beta cells, resulting in decreased insulin secretion. During hypoglycemia, activated sympathetic neural and adrenomedullary systems are also released. An array of neuropeptides (glucagon-like peptides and neuropeptide Y) from sympathetic, parasympathetic, and other neurons.

Substrate Glucoregulatory Factors

Glucose per se shifts hepatic metabolism in favor of glycogen storage. Hepatic glucose autoregulation (namely hepatic glucose production as an inverse function of the plasma glucose concentration) is an important determinant of the hyperglycemic response. In addition, lower plasma glucose levels may stimulate insulin secretion. Insulin secretion is under the control of substrate (glucose), neural, and hormonal signals. Insulin secretion is regulated by substrate, neural, and hormonal signals. (see Chapter 29). Decreasing plasma glucose concentrations are sensed directly by pancreatic beta cells, resulting in decreased insulin secretion. During hypoglycemia, activated sympathetic neural and adrenergic systems further limit insulin secretion (by -adrenergic mechanisms). The mechanisms of the increase in glucagon secretion during hypoglycemia in humans are


Conversely, decreased insulin secretion causes increased hepatic (and renal) glucose production and decreased glucose utilization by insulin-sensitive tissues such as muscle and thus tends to raise the plasma glucose concentration. Insulin is therefore both glucolowering (regulatory) and glucose-raising (counterregulatory) hormone. The rate of insulin secretion is regulated by a number of factors, the most important of which is glucose. A fall in the plasma glucose concentration has an immediate inhibitory effect on insulin secretion, thereby limiting a further fall in the plasma glucose level. Insulin is a potent and critical hormone. Either profound insulin deficiency or marked insulin excess can be lethal. But it is not the only glucoregulatory hormone.

The hyperglycemic effect of the hormone epinephrine (Fig. 32-2) is more complex. Epinephrine is secreted from chromaffin cells of the adrenal medulla in response to falling plasma glucose levels and both stimulates hepatic (and renal) glucose production and limits glucose utilization. The actions of epinephrine are both direct and indirect and are mediated through both -adrenergic and -adrenergic receptors. Adrenergic (1) limitation of insulin secretion is an important indirect hyperglycemic action of epinephrine. It allows the hyperglycemic response to occur. However, the increase in insulin secretion that occurs as plasma glucose rises limits the magnitude of the glycemic response. Adrenergic (1) stimulation of glucagon secretion also occurs, but its contribution to the hyperglycemic effect of epinephrine appears to be minor under physiologic conditions.

Epinephrine acts directly (i.e., independent of changes in other hormones or substrates) to increase hepatic glycogenolysis and gluconeogenesis. In humans the hepatic effect is mediated predominantly through -adrenergic mechanisms, although a small direct -adrenergic stimulation of hepatic glucose production has been reported. Epinephrine also mobilizes gluconeogenic precursors (e.g., lactate, alanine, and glycerol) and, like glucagon, acts within minutes to produce a transient increase in glucose production and support basal rates of glucose production thereafter. In contrast to glucagon, however, epinephrine also limits glucose utilization (i.e., it reduces glucose clearance) by insulin-sensitive tissues such as skeletal muscle, predominantly through direct -adrenergic mechanisms. Because of the persistent effect on glucose utilization, sustained hyperepinephrinemia results in persistent hyperglycemia.

Long-term elevations of growth hormone and of cortisol limit glucose utilization and stimulate glucose production. Initially, however, growth hormone has a plasma glucose-lowering (insulin-like) effect; its hyperglycemic effect does not appear for several hours. Similarly, cortisol causes an increase in the plasma glucose level after 2 to 3 hours. The hyperglycemic effect of the combination of glucagon, epinephrine, and cortisol is greater than the sum of the effects of the hormones individually.

Neural Glucoregulatory Factors

The sympathetic neurotransmitter norepinephrine exerts hyperglycemic actions by mechanisms assumed to be similar to those of epinephrine, except that norepinephrine is released primarily from terminals of sympathetic postganglionic neurons. These terminals are adjacent to adrenergic receptors on target cells within the innervated tissues. Electrical stimulation of hepatic sympathetic nerves decreases glycogen content, increases glucose release, and causes hyperglycemia in animals and in humans. Parasympathetic stimulation increases the hepatic glycogen content and decreases hepatic glucose release.

Substrate Glucoregulatory Factors

Glucose per se shifts hepatic metabolism in favor of glycogen storage. Hepatic glucose autoregulation (namely hepatic glucose production as an inverse function of the plasma glucose concentration) is an important determinant of the hyperglycemic response. In addition, lower plasma glucose levels may stimulate insulin secretion. Insulin secretion is under the control of substrate, neural, and hormonal signals. Insulin secretion and glucagon secretion are regulated by substrate, neural, and hormonal signals. (see Chapter 29). Decreasing plasma glucose concentrations are sensed directly by pancreatic beta cells, resulting in decreased insulin secretion. During hypoglycemia, activated sympathetic neural and adrenergic systems further limit insulin secretion (by -adrenergic mechanisms). The mechanisms of the increase in glucagon secretion during hypoglycemia in humans are
controversial. There is considerable evidence that autonomic nervous systems (sympathetic and parasympathetic) and adrenomedullary activation play an important role in experimental animals. However, the relative role of these central nervous system (CNS)-mediated mechanisms is less clear in humans. For example, the glucagon response to hypoglycemia is not reduced substantially in adrenalectomized, spinal conduction-blocked, or vagotomized humans. Furthermore, the glucagon response is not reduced by combined adrenergic and muscarinic cholinergic blockade; however, it has been reported to be reduced by ganglionic blockade with the nicotinic cholinergic antagonist trimethaphan.

Other mechanisms may include direct alpha cell sensing of falling plasma glucose concentrations, although isolated alpha cells do not appear to release glucagon when incubated in media with low glucose concentrations.

Finally, because alpha cells are downstream of beta cells in the pancreatic islet microcirculation and insulin is known to inhibit glucagon secretion, it is conceivable that basal insulin secretion tonically inhibits glucagon secretion and that a decrease in intraislet insulin triggers glucagon secretion during hypoglycemia. Indeed, evidence that prevention of a decrease in insulin secretion during hypoglycemia virtually eliminates the glucagon response to hypoglycemia in humans suggests that this is an important mechanism.

The autonomic (adrenomedullary, sympathetic, and parasympathetic), adrenocorticotropic hormone (ACTH)-mediated cortisol and growth hormone responses to hypoglycemia are mediated through the CNS. Although sympathetic reflexes at the spinal cord level can be elicited by various stimuli in patients with spinal cord transection, sympathochromaffin responses to hypoglycemia or to cellular glucopenia produced by 2-deoxyglucose do not occur in such individuals.

The ventromedial nucleus of the hypothalamus is an important site of glucose-sensing neurons that trigger CNS-mediated neuroendocrine responses to hypoglycemia. However, there is evidence that there are widespread glucose-sensing sites within the brain and in peripheral locations, including the portal vein.
Glycemic Thresholds for Responses to Hypoglycemia

Falling plasma glucose concentrations normally elicit a typical sequence of responses. 

Arterialized venous glycemic thresholds for several of these are listed in Table 32-1 and illustrated in Figure 32-3. Insulin secretion decreases (favoring increased glucose production as well as decreased glucose utilization by tissues other than the brain) as plasma glucose levels decline within the physiologic range. Secretion of counterregulatory hormones including glucagon (which stimulates hepatic glycogenolysis and favors gluconeogenesis) and epinephrine (which stimulates hepatic glycogenolysis and, by mobilizing precursors, hepatic and renal gluconeogenesis and which limits glucose utilization by insulin-sensitive tissues) increases as plasma glucose levels fall just below the physiologic range. Lower plasma glucose concentrations cause symptoms and signs of hypoglycemia and, ultimately, brain dysfunction.

When the same methods are used, the glycemic thresholds for the various responses to falling plasma glucose concentrations in healthy subjects are quite reproducible from laboratory to laboratory (see Fig. 32-3). Nonetheless, these thresholds are dynamic rather than static. As discussed later in this chapter, they shift to higher plasma glucose concentrations in people with poorly controlled diabetes (who often have symptoms of hypoglycemia at higher than normal glucose levels) and to lower plasma glucose concentrations in people who suffer recurrent hypoglycemia, such as those with well-controlled diabetes or with an insulinoma (who often tolerate subnormal glucose levels without symptoms).
Glucose Counterregulation

The physiology of glucose counterregulation—the mechanisms that normally prevent or rapidly correct hypoglycemia—has been reviewed in detail. Early studies of the mechanisms of the correction of short-term insulin-induced hypoglycemia and of more prolonged insulin-induced hypoglycemia are summarized in Figure 32-4 and Figure 32-5, respectively. These and studies of the prevention of hypoglycemia are detailed elsewhere.

The principles of glucose counterregulation are three. First, the prevention and the correction of hypoglycemia involve both waning of insulin and activation of glucose counterregulatory factors. These are not due solely to waning of insulin. Second, although insulin is the dominant plasma glucoselowering factor, there are redundant glucose counterregulatory factors including a decrease in insulin and increases in glucagon and epinephrine. Thus, there is a fail-safe system that prevents failure of the counterregulatory process even when one, or perhaps more, of the components of the system fails. Third, there is a hierarchy among the counterregulatory factors. Some are more important than others.

The physiology of glucose counterregulation is also summarized in Table 32-1. The first defense against falling plasma glucose concentrations is decreased insulin secretion. Among the counterregulatory factors, increased glucagon secretion plays a primary role. Glucose recovery from hypoglycemia is impaired and postabsorptive plasma glucose concentrations decline but then level off when glucagon secretion is deficient. Glucagon is the second defense against falling plasma glucose concentrations. Albeit demonstrably involved, increased epinephrine secretion is not normally critical. It becomes critical when glucagon is deficient. Epinephrine is the third defense against falling plasma glucose concentrations. Hypoglycemia develops or progresses when both glucagon and epinephrine are deficient and insulin is present despite the actions of the other glucose counterregulatory factors. Thus, insulin, glucagon, and epinephrine stand high in the hierarchy of redundant glucose counterregulatory factors.

Growth hormone and cortisol, both of which tend to increase plasma glucose concentrations after several hours, are involved in defense against prolonged hypoglycemia. However, neither is critical to recovery from even prolonged hypoglycemia or, at least in adults, to prevention of hypoglycemia after an overnight fast.

There is some evidence that glucose autoregulation is involved although only during severe hypoglycemia. Other hormones, neurotransmitters, and substrates other than glucose (and fatty acids that may mediate part of the effect of epinephrine) may also be involved. If so, they play relatively minor roles.
PATHOPHYSIOLOGY OF HYPOGLYCEMIA

Clinical Manifestations of Hypoglycemia

Whipple’s triad: symptoms consistent with hypoglycemia, a low plasma glucose concentration, and relief of those symptoms when the plasma glucose concentration is raised provides compelling evidence of hypoglycemia.

Symptoms of hypoglycemia can be divided into two categories, neuroglycopenic and neurogenic (autonomic) symptoms.[1] Neuroglycopenic symptoms are the direct result of CNS neuronal glucose deprivation. They include behavioral changes, confusion, fatigue or weakness, warmth, visual changes, seizure, loss of consciousness, and, if hypoglycemia is severe and prolonged, death. Neurogenic symptoms are the result of the perception of physiologic changes caused by the autonomic nervous system discharge triggered by hypoglycemia. They include adrenergic symptoms such as palpitations, tremor, and anxiety and cholinergic symptoms such as sweating, hunger, and paresthesias. Adrenergic symptoms are mediated by norepinephrine released from sympathetic postganglionic neurons, the adrenal medulla, or both and epinephrine released from the adrenal medulla. The relative contributions of these are not known, but palpitations have been attributed to circulating epinephrine and tremor to sympathetic neural norepinephrine. Cholinergic symptoms, at least sweating, are thought to be mediated by acetylcholine released from sympathetic postganglionic neurons.

Representative neurogenic and neuroglycopenic symptoms of hypoglycemia are listed in Figure 32-6, which also illustrates that awareness of hypoglycemia is the extent to which individuals perceive that their blood sugar is low. The result

---


---

of the perception of neurogenic symptoms. The autonomic, including the sympatho-chromaffin, response to hypoglycemia is initiated by glucose-sensitive neurons, including those in the ventromedial hypothalamus, that increase firing as extracellular glucose levels fall. Thus, although both neurogenic and neuroglycopenic symptoms could be viewed as fundamentally neuroglycopenic in origin, their mechanisms are different. Some prefer the term autonomic rather than neurogenic symptoms because they may be mediated by a circulating hormone (epinephrine).

---

*Figure 32-7 Neurogenic (autonomic) and neuroglycopenic symptoms of hypoglycemia in normal humans. Among the neurogenic symptoms, "sweaty," "hungry," and "tingling" are cholinergic and "shaky/tremulous," "heart pounding," and "nervous/anxious" are adrenergic. See text for discussion. Mean (±SE) subject scores for awareness of hypoglycemia (blood sugar low) symptoms consistent with hypoglycemia, a low plasma glucose concentration, and relief of those symptoms when the plasma glucose concentration is raised provides compelling evidence of hypoglycemia.

---

as well as by a neurotransmitter. Nonetheless, the autonomic response is CNS-mediated and, therefore, the resulting symptoms are fundamentally neurogenic.

Common signs of hypoglycemia include pallor and diaphoresis. Heart rate and systolic blood pressure are typically increased, but these findings may not be prominent. To the extent that they are observable, the neuroglycopenic manifestations are often valuable, albeit nonspecific, signs. Transient neurologic defects occur occasionally. Permanent neurologic damage is rare.

The magnitude of the responses to hypoglycemia is an inverse function of the nadir plasma glucose concentration rather than the rate of decrease in plasma glucose.[2] It was once thought that neurogenic symptoms are less prominent when hypoglycemia develops gradually. However, the relative paucity of symptoms at a given low plasma glucose concentration in individuals with recurrent hypoglycemia, such as those with tightly controlled diabetes[3] or with an insulinoma[4] is attributable to a shift in glycemic thresholds for responses to lower plasma glucose concentrations. Conversely, the threshold shifts to higher plasma glucose concentrations in patients with chronic hyperglycemia result in symptoms of hypoglycemia at relatively high glucose levels.[5] The mechanism of these shifts in thresholds is unknown. Potential mechanisms are discussed later in this chapter.

---

*Figure 32-8 Representative neurogenic and neuroglycopenic symptoms of hypoglycemia are listed in Figure 32-6, which also illustrates that awareness of hypoglycemia is the extent to which individuals perceive that their blood sugar is low. The result
Diagnosis of Hypoglycemia

The manifestations of hypoglycemia are nonspecific, vary among individuals, and may change from time to time in the same individual. They are also typically episodic. Thus, although the history is of fundamental importance in suggesting the possibility of hypoglycemia, the diagnosis cannot be made solely on the basis of symptoms and signs.

The diagnosis of hypoglycemia should also not be made solely on the basis of plasma glucose measurements unless they are unequivocally subnormal. It is not possible to define a plasma glucose concentration below which neuroglycopenia invariably occurs and above which neuroglycopenia never occurs. Although symptoms commonly occur with plasma glucose levels less than 3.0 mmol/L (54 mg/dL), they can occur at higher plasma glucose levels in poorly controlled diabetes and only at lower glucose levels in well-controlled diabetes or in other conditions that result in recurrent hypoglycemia such as insulinoma. In addition, venous plasma glucose concentrations substantially less than 3.0 mmol/L (54 mg/dL) may occur in normal individuals late after glucose ingestion (arterial glucose levels are higher) and in some women and children during fasting without producing recognizable symptoms. This is not to say that distinctly low plasma glucose measurements should be ignored. Some patients with endogenous hyperinsulinism or intensively treated diabetes tolerate glucose levels that are unequivocally subnormal, as mentioned earlier. Because these patients can have hypoglycemic symptoms at other times (presumably when glucose levels are even lower), it would be inappropriate to deny that they have hypoglycemia.

In general, venous plasma glucose concentrations greater than 3.9 mmol/L (70 mg/dL) after an overnight fast are normal, those between 2.8 and 3.9 mmol/L (50 and 70 mg/dL) are suggestive of hypoglycemia, and those less than 2.8 mmol/L (50 mg/dL) indicate postabsorptive hypoglycemia. Because substantial glucose extraction occurs across the forearm under hyperinsulinemic conditions, arterial glucose concentrations (those relevant to brain function) are as much as 30% higher than venous glucose concentrations after an oral glucose load. Artifactualy low measured glucose levels can result from glycolysis in vitro (pseudohypoglycemia), particularly in the presence of leukocytosis or polycythemia, or both, or if separation of plasma from the formed elements of the blood is delayed. The diagnosis of hypoglycemia is most convincingly established when it is based on Whipple’s triad: symptoms consistent with hypoglycemia, a low plasma glucose concentration, and relief of those symptoms when the plasma glucose concentration is increased to normal levels.
Postabsorptive versus Postprandial Hypoglycemia

Reproducible hypoglycemia in the postabsorptive state implies the presence of disease and requires diagnostic explanation and therapy. This condition is commonly referred to as postabsorptive, or fasting, hypoglycemia. However, it need not be apparent initially or exclusively during prolonged fasting or after an overnight fast; it may become symptomatic during the latter portion of any interdigestive period especially with exercise. In contrast, postprandial (reactive, stimulative) hypoglycemia usually does not imply a serious underlying disorder. Thus, the distinction between postabsorptive and postprandial hypoglycemia is useful.
Mechanisms of Hypoglycemia

Hypoglycemia indicates that the rate of glucose efflux from the circulation exceeds that of glucose influx into the circulation. It can result from excessive glucose efflux (excessive utilization, external losses) or deficient glucose influx (deficient endogenous production in the absence of exogenous glucose delivery), or both. Conditions in which glucose utilization is increased include exercise, pregnancy, and sepsis; renal losses can occur at physiologic plasma glucose concentrations (e.g., renal glycosuria, pregnancy). However, because of the capacity of the normal liver (and kidneys) to increase glucose production severalfold, as discussed earlier, clinical hypoglycemia rarely results solely from excessive glucose efflux. Rather it is commonly the result of inappropriately low glucose production relative to the rate of glucose utilization.

Hypoglycemia can be caused by regulatory, enzymatic, or substrate defects. Gluoregulatory defects include excessive secretion of insulin or deficient secretion of glucose counterregulatory hormones. Enzymatic defects in glucose production may be primary or may result from hepatic disease. Substrate defects include failure to mobilize or utilize gluconeogenic substrates.
Clinical Classification of Hypoglycemia

Hypoglycemia can be classified on the basis of glucose kinetic patterns, pathogenic mechanisms, or disease groups. The last approach is used in this chapter (Table 32-2). Postabsorptive, or fasting, hypoglycemia can be the result of drugs, critical illnesses including hepatic or renal failure, hormonal deficiencies, non-beta cell tumors, endogenous hyperinsulinism (including that caused by pancreatic beta cell tumors), or metabolic disorders of infancy and childhood. Postprandial, or reactive, hypoglycemia is rarely caused by congenital enzyme defects but can follow gastric surgery and perhaps occurs rarely as an idiopathic disorder.

Most episodes of hypoglycemia result from drugs, particularly insulin, sulfonylureas, or alcohol. In one series of patients treated in an emergency room for hypoglycemia, two thirds had diabetes mellitus and two thirds had been drinking alcohol. Clearly, the combination of drug-treated diabetes and alcohol ingestion can be devastating. Nearly one fourth of the patients were septic, but diabetes or alcohol ingestion was common even in those patients. Drugs are also a common cause of hypoglycemia in inpatients. In this case, however, critical illnesses such as renal or hepatic failure, sepsis, and inanition are common. Hypoglycemia resulting from hormonal

<table>
<thead>
<tr>
<th>TABLE 32-2 -- Clinical Classification of Hypoglycemia</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Postabsorptive (Fasting) Hypoglycemia</strong></td>
</tr>
<tr>
<td>Drugs</td>
</tr>
<tr>
<td>Especially insulin, sulfonylureas, alcohol</td>
</tr>
<tr>
<td>Also pentamidine, quinine</td>
</tr>
<tr>
<td>Rarely, salicylates, sulfonamides</td>
</tr>
<tr>
<td>Others</td>
</tr>
<tr>
<td>Critical illnesses</td>
</tr>
<tr>
<td>Hepatic failure</td>
</tr>
<tr>
<td>Cardiac failure</td>
</tr>
<tr>
<td>Renal failure</td>
</tr>
<tr>
<td>Sepsis</td>
</tr>
<tr>
<td>Inanition</td>
</tr>
<tr>
<td>Hormonal deficiencies</td>
</tr>
<tr>
<td>Cortisol or growth hormone, or both</td>
</tr>
<tr>
<td>Glucagon and epinephrine</td>
</tr>
<tr>
<td>Non-beta cell tumors</td>
</tr>
<tr>
<td>Endogenous hyperinsulinism</td>
</tr>
<tr>
<td>Pancreatic beta cell disorders</td>
</tr>
<tr>
<td>Tumor (insulinoma)</td>
</tr>
<tr>
<td>Nontumor</td>
</tr>
<tr>
<td>Beta cell secretagogue (e.g., sulfonylureas)</td>
</tr>
<tr>
<td>Autoimmune hypoglycemia</td>
</tr>
<tr>
<td>Insulin antibodies</td>
</tr>
<tr>
<td>Insulin receptor antibodies</td>
</tr>
<tr>
<td>? Beta cell antibodies</td>
</tr>
<tr>
<td>? Ectopic insulin secretion</td>
</tr>
<tr>
<td>Hypoglycemias of infancy and childhood</td>
</tr>
</tbody>
</table>

| **Postprandial (Reactive) Hypoglycemia**             |
| Congenital deficiencies of enzymes of carbohydrate metabolism |
| Hereditary fructose intolerance                      |
| Galactosemia                                          |
| Alimentary hypoglycemia                              |
| Idiopathic (functional) postprandial hypoglycemia     |
HYPOGLYCEMIA IN DIABETES MELLITUS

Clinical Context

It is now well established that comprehensive care makes a difference for people with diabetes. A fundamentally important component of comprehensive care of diabetes is glycemic control because it prevents or delays the long-term specific complications of diabetes (retinopathy, nephropathy, and neuropathy) and may reduce its macrovascular complications. However, iatrogenic hypoglycemia is the limiting factor in the glycemic management of both type 1 diabetes mellitus (T1DM) and type 2 diabetes mellitus (T2DM) both conceptually and in practice.

Conceptually, were it not for the potentially devastating effects of hypoglycemia on the brain, hyperglycemia would be rather easy to treat. Administration of enough insulin (or any effective medication) to lower plasma glucose levels to or below the normal range would eliminate symptoms of hyperglycemia; prevent diabetic ketoacidosis and hyperosmolar coma; almost assuredly prevent diabetic retinopathy, nephropathy, and neuropathy; and probably reduce atherosclerotic risk. The devastating effects of hypoglycemia are real, however, and the glycemic management of diabetes is therefore complex.

In practice, euglycemia, even near euglycemia, cannot be achieved and maintained safely in most patients with T1DM and many patients with T2DM because of the barrier of iatrogenic hypoglycemia. Because of that barrier, retinopathy, nephropathy, and neuropathy develop or progress in some patients with T1DM or T2DM despite aggressive attempts to achieve glycemic control, albeit at lower rates than during less aggressive therapy. Indeed, the inability to maintain euglycemia over time, because of the barrier of hypoglycemia, may explain the limited impact of aggressive glycemic therapy on the atherosclerotic complications of diabetes.

In T1DM, aggressive attempts to achieve glycemic control increase the risk of severe, at least temporarily disabling, iatrogenic hypoglycemia (i.e., that requiring the assistance of another individual) more than threefold. Documented in both of the controlled clinical trials with sample sizes large enough to demonstrate beneficial effects of intensive therapy on the long-term complications of diabetes, the Diabetes Control and Complications Trial (DCCT) and the Stockholm Diabetes Intervention Study, that fact was confirmed in a meta-analysis that also included 12 smaller controlled clinical trials of intensive therapy. However, it is possible to reduce the risk of hypoglycemia during aggressive therapy of T1DM. For example, the sixfold increased risk of severe hypoglycemia during intensive therapy in the feasibility phase of the DCCT was reduced by half in the full-scale trial.

Because of the interplay of therapeutic insulin excess and compromised physiologic and behavioral defenses against falling plasma glucose concentrations, as discussed later in this chapter, people with T1DM are at ongoing risk for episodes of hypoglycemia. Those attempting to achieve glycemic control suffer untold numbers of episodes of asymptomatic hypoglycemiaplasma glucose levels may be lower than 2.8 mmol/L (50 mg/dL) as much as 10% of the time and an average of two episodes of symptomatic hypoglycemia per week. They suffer an episode of severe, at least temporarily disabling hypoglycemia every year or two on average. Although seemingly complete recovery from even severe hypoglycemia is the rule, permanent neurologic deficits can result. It has been estimated that 2% to 4% of deaths of people with T1DM are caused by hypoglycemia. In addition, hypoglycemia can cause recurrent or even persistent psychosocial morbidity. The reality of hypoglycemia, the rational fear of hypoglycemia, or both can be a barrier to glycemic control.

Iatrogenic hypoglycemia is generally less frequent in T2DM. However, it occurs during treatment with sulfonylureas or other insulin secretagogues (and has been reported in patients treated with metformin) or with insulin (Table 32-4; see Table 32-3). The frequency of hypoglycemia approaches that in T1DM in those who reach the insulin-deficient end of the spectrum of T2DM. Indeed, in one series, the frequency of severe hypoglycemia was similar in patients with T2DM and T1DM matched for duration of insulin therapy.

### Table 32-3 -- Severe Hypoglycemia during Aggressive Glycemic Therapy of Diabetes

<table>
<thead>
<tr>
<th>Type 1 Diabetes</th>
<th>Episodes Per 100 Patient-Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edinburgh (Diabet Med 1993; 10:238)</td>
<td>170</td>
</tr>
<tr>
<td>Utrecht (Diabetes Care 2000; 23:1467)</td>
<td>150</td>
</tr>
<tr>
<td>Stockholm Diabetes Intervention Study (Diabetes 1994; 43:313)</td>
<td>110</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type 2 Diabetes</th>
<th>Episodes Per 100 Patient-Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edinburgh (Diabet Med 1993; 10:238)</td>
<td>73</td>
</tr>
<tr>
<td>Veterans Affairs Pump Study (JAMA 1996; 276:1322)</td>
<td>10</td>
</tr>
<tr>
<td>Veterans Affairs Cooperative Study (Diabetes Care 1993; 8:1113)</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Therapy</th>
<th>N</th>
<th>Hemoglobin A1c, HbA1c (%)</th>
<th>Percent with Hypoglycemia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diet</td>
<td>379</td>
<td>8.0</td>
<td>Any 3.0, Major 0.2</td>
</tr>
<tr>
<td>Sulfonylurea</td>
<td>922</td>
<td>7.1</td>
<td>45.0, 3.3</td>
</tr>
<tr>
<td>Insulin</td>
<td>689</td>
<td>7.1</td>
<td>76.0, 11.2</td>
</tr>
<tr>
<td>Diet</td>
<td>297</td>
<td>8.2</td>
<td>Any 2.8, Major 0.4</td>
</tr>
<tr>
<td>Metformin</td>
<td>251</td>
<td>7.4</td>
<td>17.6, 2.4</td>
</tr>
</tbody>
</table>


*Taking assigned medication.*

*Comparing severe hypoglycemia (that requiring the assistance of another individual) in 65% of intensively treated patients over 6.5 years in the Diabetes Control and Complications Trial.*
Study investigators concluded that over time hypoglycemia becomes limiting in the treatment of T2DM just as it is in the treatment of T1DM. Given the now well-established long-term benefits of glycemic control and the short-term potentially devastating effects of iatrogenic hypoglycemia, it is clear that the goals of both reducing mean glycemia and minimizing hypoglycemia are important for people with diabetes. Minimizing the risk of hypoglycemia in T1DM involves both application of the principles of aggressive therapy—education and empowerment of patients, frequent self-monitoring of blood glucose, flexible insulin regimens, individualized glycemic goals, and ongoing professional guidance and support—and implementation of hypoglycemia risk reduction. As discussed later in this chapter, hypoglycemia risk reduction requires consideration of the roles of both therapeutic insulin excess and compromised physiologic and behavioral defenses against developing hypoglycemia.
Risk Factors

Insulin Excess

The conventional risk factors for iatrogenic hypoglycemia in T1DM (Table 32-5) are based on the premise that relative or absolute therapeutic insulin excess, which must occur from time to time because of the gross pharmacokinetic imperfections of all current insulin replacement regimens, is the sole determinant of risk.

Relative or absolute therapeutic insulin excess occurs when:

1. Insulin doses are excessive, ill-timed, or of the wrong type.
2. The influx of exogenous glucose is decreased (as during the overnight fast or after missed meals or snacks).
3. Insulin-independent glucose utilization is increased (as during exercise).
4. Endogenous glucose production is decreased (as after alcohol ingestion or administration or other drugs and with loss of renal parenchyma).
5. Sensitivity to insulin is increased (as after exercise; in the middle of the night; with glycemic control; with increased fitness, weight loss, or both; or with administration of certain drugs).
6. Insulin clearance is decreased (as in renal failure).

### TABLE 32-5 -- Comprehensive Risk Factors for Hypoglycemia in Diabetes

Premise: Iatrogenic hypoglycemia in type 1 diabetes is the result of the interplay of therapeutic insulin excess and compromised glucose counterregulation.

1. Absolute or relative therapeutic insulin excess (the conventional risk factors)
   a. Insulin doses excessive, ill-timed, wrong type
   b. Decreased food intake
   c. Increased glucose utilization (e.g., exercise)
   d. Decreased glucose production (e.g., alcohol)
   e. Increased sensitivity to insulin (e.g., after exercise, during the night, glycemic control, weight loss)
   f. Decreased insulin clearance (e.g., renal failure)
2. Compromised glucose counterregulation
   a. Absolute insulin deficiency (C-peptide negativity)
      - Cell destruction: No insulin in response to glucose
      - Unknown: No in glucagon in response to glucose
   b. History of severe hypoglycemia or aggressive therapy per se (lower glucose goals, lower hemoglobin A1c)

Episodes of hypoglycemia:
- Attenuated autonomic (including epinephrine) activation and symptoms in response to glucose (defective glucose counterregulation and hypoglycemia unawareness)

These are the issues with which people with diabetes and their health care providers deal routinely as they attempt to minimize iatrogenic hypoglycemia. However, it became clear early in the DCCT that these conventional risk factors explain only a minority of episodes of severe iatrogenic hypoglycemia. Indeed, in a multivariate model none was found to be statistically significant. Clearly, we must look beyond these risk factors if we are to understand the majority of episodes of severe hypoglycemia in T1DM.

Interplay of Insulin Excess and Compromised Glucose Counterregulation

Iatrogenic hypoglycemia in T1DM is more appropriately viewed as the result of the interplay of relative or absolute therapeutic insulin excess (the conventional risk factors) and compromised glucose counterregulation (see Table 32-5). Three clinically well-documented risk factors for iatrogenic hypoglycemia in T1DM are (1) absolute insulin deficiency (i.e., C-peptide negativity), (2) a history of severe hypoglycemia, and (3) aggressive glycemic therapy per se as evidenced by lower glycosylated goals or lower hemoglobin A1c levels. (Obviously, iatrogenic hypoglycemia occurs in people with diabetes who are not C-peptidenegative, have no history of severe hypoglycemia, and are not practicing aggressive glycemic therapy. Nonetheless, these are associated with a substantially increased risk of hypoglycemia.) These three risk factors are clinical surrogates of compromised physiologic and behavioral defenses against falling plasma glucose concentrations: the clinical syndromes of defective glucose counterregulation and of hypoglycemia unawareness and the pathophysiologic concept of hypoglycemia-associated autonomic failure.
Pathophysiology of Glucose Counterregulation in Diabetes

As the person with T1DM becomes absolutely insulin-deficient over the first few months or years of clinical T1DM, circulating insulin levels then simply the passive result of absorption of exogenous insulin not fall as plasma glucose levels decline. The first defense against hypoglycemia is lost.

Over the same time frame, the glucagon response to hypoglycemia is lost in T1DM. This is a selective defect; the glucagon responses to other stimuli are largely, if not entirely, intact. The mechanism of the defect is unknown, but it is tightly linked to absolute insulin deficiency. Given that and the finding that a decrease in intracellular secretion is normally a potent stimulus to the glucagon secretory response to hypoglycemia, the absent glucagon response may be a direct result of absent insulin secretion. Thus, the clinical hypoglycemia risk factor of C-peptide negativity indicates that the first defense against hypoglycemia (decreased insulin secretion) is lost and predicts accurately that the second defense against hypoglycemia (increased glucagon secretion) is lost. Therefore, patients with established (i.e., C-peptide negative) T1DM are largely dependent on the third defense against hypoglycemia, increased epinephrine secretion.

The epinephrine response to hypoglycemia is attenuated in many patients with T1DM, particularly those with the other clinical risk factors for hypoglycemia such as a history of severe hypoglycemia or aggressive glycemic therapy per se as evidenced by lower glycemic goals, lower hemoglobin A1C levels, or both. The former indicates and the latter implies recurrent episodes of prior hypoglycemia. In contrast to the absent glucagon response, the attenuated epinephrine response represents a threshold shift; an epinephrine response can be elicited, but lower plasma glucose concentrations are required. This threshold shift to lower plasma glucose concentrations is largely the result of recent antecedent iatrogenic hypoglycemia.

Recent antecedent hypoglycemia reduces autonomic (including adrenomedullary epinephrine) and symptomatic, among other, responses to a given level of subsequent hypoglycemia in nondiabetic individuals and in patients with T1DM. It shifts the glycemic thresholds for these responses to lower plasma glucose concentrations. As a result, it also impairs glycemic defense against hyperinsulinemia and developing hypoglycemia in T1DM and in patients with T1DM. In addition to this functional threshold shift, there may be an anatomic component of adrenomedullary chromaffin cells-the reduced epinephrine response in patients with classical diabetic autonomic neuropathy. Nonetheless, the epinephrine response is typically reduced in patients with no clinical evidence of classical diabetic autonomic neuropathy.

The development of an attenuated epinephrine response to falling glucose levels loss of the third defense against hypoglycemia a critical pathophysiologic event. Patients with T1DM who have combined deficiencies of glucagon and epinephrine responses have been shown in prospective studies to suffer severe hypoglycemia at rates 25-fold or more higher than those of patients with absent glucagon but intact epinephrine responses during aggressive glycemic therapy. They have the clinical syndrome of defective glucose counterregulation. The mechanisms of altered and defective glucose counterregulation are illustrated in Figure 32-12.

By reducing the autonomic, specifically the sympathochromaffin, responses to subsequent hypoglycemia, recent antecedent iatrogenic hypoglycemia also causes loss of the warning, largely if not exclusively neurogenic, symptoms of developing hypoglycemia that previously allowed the patient to recognize that glucose levels were falling and prompted the appropriate behavioral response (e.g., ingestion of food) to abort the episode. Thus, the first clinical manifestation of a hypoglycemic episode is neuroglycopenia, and it is often too late for the patient to recognize and self-treat the episode. This is the clinical syndrome of hypoglycemia unawareness. It, too, has been shown in a prospective study to be associated with a high frequency of severe iatrogenic hypoglycemia.

The concept of hypoglycemia-associated autonomic failure in T1DM, a functional disorder distinct from the fixed autonomic failure of classical diabetic autonomic neuropathy, was formulated and then verified experimentally to unify the pathogenesis of the clinical syndromes of defective glucose counterregulation and hypoglycemia unawareness.

The concept of hypoglycemia-associated autonomic failure in T1DM posits that:

1. Periods of relative or absolute therapeutic insulin excess in the setting of absent glucagon responses lead to episodes of hypoglycemia.
2. These episodes, in turn, cause reduced autonomic (including adrenomedullary epinephrine) responses to falling glucose concentrations on subsequent occasions.
3. These reduced autonomic responses result in reduced symptoms of, and therefore behavioral responses to, developing hypoglycemia (i.e., hypoglycemia unawareness) and because epinephrine responses are reduced in the setting of absent glucagon response impaired physiologic defenses against developing hypoglycemia (i.e., defective glucose counterregulation).

Thus, a vicious circle of recurrent hypoglycemia is created and perpetuated.

Perhaps the most compelling support for the concept of hypoglycemia-associated autonomic failure in T1DM is the finding, in three independent laboratories, that hypoglycemia unawareness (Fig. 32-14) and, at least in part, the reduced epinephrine component of defective glucose counterregulation are reversible after as little as 2 weeks of scrupulous avoidance of iatrogenic hypoglycemia in most affected patients. This involves a shift of glycemic thresholds for autonomic and
symptomatic responses back toward higher plasma glucose concentrations.

The basic mechanism of hypoglycemia-associated autonomic failure remains to be determined. There is evidence that it is mediated by the cortisol response to previous hypoglycemia\textsuperscript{112}, although that remains to be confirmed independently. Evidence, obtained with the Kety-Schmidt technique, that it involves increased brain glucose uptake during hypoglycemia has been reported\textsuperscript{113,114}. However, evidence that recent antecedent hypoglycemia does not increase blood-to-brain glucose transport or cerebral glucose metabolism, measured with \textsuperscript{11\textsubscript{C}}glucose and positron emission tomography, or cerebral blood flow, measured with \textsuperscript{15\textsubscript{O}}water, has been presented\textsuperscript{115}. The latter data do not exclude regional increments in blood-to-brain glucose transport. Alternatively, the alteration may lie beyond the blood-brain barrier.

Consistent with the concept of hypoglycemia-associated autonomic failure in T1DM, recent antecedent hypoglycemia also shifts glycemic thresholds for hypoglycemic cognitive dysfunction to lower plasma glucose concentrations\textsuperscript{116,117} and impairs detection of hypoglycemia in the clinical setting in patients with T1DM.\textsuperscript{118} In addition to shifting the thresholds for the adrenomedullary (plasma epinephrine) and parasympathetic (plasma pancreatic polypeptide) response to lower plasma glucose concentrations, recent antecedent hypoglycemia has been reported to reduce the sympathetic neural response to subsequent hypoglycemia,\textsuperscript{119-121} although the latter has been questioned.\textsuperscript{122}

There is also evidence that reduced sensitivity to catecholamines, measured as a reduced heart rate response to the \textsuperscript{-adrenergic agonist isoproterenol, contributes to the pathogenesis of hypoglycemia unawareness in T1DM.\textsuperscript{123} Hypoglycemia has been reported to reduce \textsuperscript{-adrenergic sensitivity, tested about 10 hours later, in T1DM (but not in nondiabetic individuals).\textsuperscript{124} Thus, it is conceivable that both reduced activation of the sympathochromaffin system and reduced sensitivity to released catecholamines might play a role in the pathogenesis of hypoglycemia unawareness and defective glucose counterregulation induced by recent antecedent iatrogenic hypoglycemia.

The extent to which these pathophysiologic concepts, developed in T1DM, apply to patients with T2DM remains to be assessed in detail. They may well apply to those approaching the insulin-deficient end of the spectrum of T2DM because hypoglycemia becomes limiting to glycemic control in such patients.\textsuperscript{126} Indeed, in one series the frequency of severe hypoglycemia was found to be similar in patients with T2DM and T1DM matched for duration of insulin therapy.\textsuperscript{127} This issue is complicated by the fact that some patients with apparent T2DM may actually have late-onset T1DM.\textsuperscript{128} Nonetheless, patients with advanced T2DM have been reported to have reduced glucagon responses to hypoglycemia\textsuperscript{129} (Table 32-6), a key feature of defective glucose counterregulation in T1DM. Furthermore, patients with T2DM have reduced epinephrine and neurogenic symptom responses to hypoglycemia after episodes of hypoglycemia, key features of defective glucose counterregulation, hypoglycemia unawareness, and hypoglycemia-associated autonomic failure in T1DM.\textsuperscript{130}

The concept of hypoglycemia-associated autonomic failure in T1DM is illustrated in Figure 32-11, and the comprehensive risk factors for iatrogenic hypoglycemia in T1DM, viewed in the context of the interplay of therapeutic insulin excess and compromised glucose counterregulation, are outlined in the Table 32-5.
Hypoglycemia Risk Reduction in Diabetes

Clearly, every effort must be made to minimize the risk of iatrogenic hypoglycemia and eliminate the risk of severe hypoglycemia while pursuing the greatest degree of glycemic control that can be achieved safely in an individual person with diabetes. Hypoglycemia risk reduction involves (1) addressing the issue of hypoglycemia in every contact with the patient, (2) applying the principles of aggressive glycemic therapy, and (3) considering each of the comprehensive risk factors for hypoglycemia.

In addition to questioning the patient about episodes of symptomatic and biochemical hypoglycemia and looking for

![Figure 32-12 Schematic representation of the pathophysiology of glucose counterregulation in people with type 1 diabetes mellitus (T1DM). See text for discussion.](image)

low values in the self-monitoring of the blood glucose (SMBG) log, it is important to assess the patient's awareness of hypoglycemia. A history of hypoglycemia unawareness identifies that clinical syndrome (and also implies defective glucose counterregulation). It is also important to determine the extent to which the patient is concerned about the reality or the possibility of hypoglycemia. Fear of hypoglycemia can be a barrier to glycemic control. If episodes of hypoglycemia are identified, their frequency, severity, timing, and clinical contexts need to be determined.

Once the problem of iatrogenic hypoglycemia is recognized, it is appropriate to review the treatment plan with respect to the principles of aggressive glycemic therapy. These include (1) education and empowerment of the patient; (2) frequent SMBG; (3) flexible insulin (or other drug) regimens; (4) rational, individualized glycemic goals; and (5) ongoing professional guidance and support. Particularly in T1DM, but also in advanced T2DM, glycemic control is achieved safely by a well-informed, thoughtful person with diabetes who must make judgments about the management of his or her diabetes several times each day. The patient must be given the resources to make those judgments.

In the context of these therapeutic principles, hypoglycemia risk reduction requires consideration of both the conventional risk factors that lead to episodes of absolute or relative insulin excess insulinsulin (or other drug) dose, timing, and type; patterns of food ingestion and of exercise; interactions with alcohol or other drugs; and altered sensitivity to or clearance of insulin and the risk factors for compromised glucose counterregulation that impair physiologic and behavioral defenses against developing hypoglycemia (see Table 32-5). The underlying principle is that iatrogenic hypoglycemia is the result of the interplay of insulin excess and compromised glucose counterregulation rather than insulin excess alone.

As discussed earlier, the clinical surrogates of risk attributable to compromised glucose counterregulation include absolute insulin deficiency, which may be apparent from a history of ketosis-prone diabetes requiring insulin therapy from diagnosis, although it is now recognized that absolute insulin deficiency can sometimes develop more gradually in late-onset T1DM or advanced T2DM and a history of recurrent hypoglycemia or, absent that, aggressive glycemic therapy per se as evidenced by lower glycosylated hemoglobin levels, lower hemoglobin A1c levels, or both. It is possible to test for defective glucose counterregulation with an insulin infusion test, but that is generally neither practical nor useful given the now recognized dynamic nature of hypoglycemia unawareness and the reduced epinephrine component of defective glucose counterregulation discussed earlier. On the other hand, a diagnosis of partial or complete hypoglycemia unawareness can be made if the patient is unaware about the requirement for therapy with insulin for an average of 5 years and reduced plasma C-peptide levels.

### TABLE 32-6 — Glucagon and Epinephrine Responses to Hypoglycemia in Type 2 Diabetes

<table>
<thead>
<tr>
<th>Reference</th>
<th>Glucagon</th>
<th>Epinephrine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boden et al (Diabetes 1983; 32:1055)</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td>Heller et al (Diabetologia 1987; 30:924)</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td>Menelis et al (Diabetes 1994; 43:403)</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td>Shamoon et al (J Clin Invest 1994; 93:2562)</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td>Peacey et al (Diabetes Care 2000; 23:1023)</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td>Segel et al (Diabetes 2000; 49:A131)</td>
<td>Normal</td>
<td>Normal</td>
</tr>
</tbody>
</table>

*In advanced type 2 diabetes, as evidenced by the requirement for therapy with insulin for an average of 5 years and reduced plasma C-peptide levels.*
complete hypoglycemia unawareness can be made from the history.

Clinical hypoglycemia unawareness (which also suggests defective glucose counterregulation) implies recurrent antecedent iatrogenic hypoglycemia whether that has or has not been documented. If such hypoglycemia is not apparent to the patient or to his or her family or in the SMBG log, it is probably occurring during the night.

Indeed, hypoglycemia occurring severe hypoglycemia occurs most commonly during the night in people with T1DM. That is typically the longest interdigestive period and time between SMBG and the time of maximal sensitivity to insulin. In addition, sleep limits the recognition of warning symptoms of developing hypoglycemia, and thus the appropriate behavioral response, and has been found to further reduce the epinephrine response to hypoglycemia and thus to further compromise the physiologic defense against developing hypoglycemia.

In addition to regimen adjustments, approaches to the problem of nocturnal hypoglycemia include use of newer insulin analogues and bedtime treatments. Substitution of a preprandial rapid-acting insulin analogue (e.g., lispro or aspart) for short-acting (regular) insulin during the day reduces the frequency of nocturnal hypoglycemia. Substitution of a long-acting insulin analogue (e.g., glargine) for neutral protamine Hagedorn (NPH) or Ultralente insulin at bedtime may also reduce the frequency of nocturnal hypoglycemia. Bedtime treatments intended to reduce nocturnal hypoglycemia include bedtime snacks, although their efficacy is largely limited to the first half of the night. Experimental approaches include bedtime administration of uncooked cornstarch, of the glucagon-releasing amino acid alanine or of the epinephrine-simulating \textsuperscript{2}adrenergic agonist terbutaline. The efficacy of uncooked cornstarch in the 5.0-g dose recommended remains to be established. Although bedtime alanine and bedtime terbutaline have been shown to prevent nocturnal hypoglycemia more effectively than a conventional bedtime snack, alanine is probably impractical and terbutaline has not been studied further in a large clinical trial.

Obviously, with a history of recurrent hypoglycemia, one should determine when it occurs and adjust the treatment regimen appropriately. With a basal-bolus insulin regimen, morning fasting hypoglycemia implicates the long-acting or immediate-acting basal insulin, daytime hypoglycemia implicates the rapid-acting or short-acting insulin, and nighttime hypoglycemia may implicate either, all in the context of the other risk factors for insulin excess. A history of severe iatrogenic hypoglycemia that requiring the assistance of another individual is a clinical red flag. Unless it was the result of an easily remediable factor, such as a missed meal after insulin administration or vigorous exercise without the appropriate regimen adjustment, a substantive change in the regimen must be made. If it is not, the risk of recurrent severe hypoglycemia is unacceptably high.

In a patient with hypoglycemia unawareness, a 2- to 3-week period of scrupulous avoidance of iatrogenic hypoglycemia is advisable and can be assessed by return of awareness of hypoglycemia. This return of awareness has been accomplished without or with minimal compromise of glycemic control, but that required substantial involvement of health professionals. In practice, it can involve acceptance of somewhat higher glucose levels over the short term. However, with the return of symptoms of developing hypoglycemia, empirical approaches to better glycemic control can be tried.

Hypoglycemia is a fact of life for people with T1DM (and some with T2DM) who attempt to achieve near-euglycemia. Because of the pharmacokinetic imperfections of all current insulin replacement regimens, it is not practical to maintain euglycemia while eliminating episodes of asymptomatic and even symptomatic hypoglycemia in T1DM. That awaits the ultimate goal of the prevention and cure of diabetes or, in the shorter term, development of clinical strategies for perfect insulin replacement (e.g., transplantation of insulin-secreting cells or development of a closed-loop insulin replacement system) or for near-perfect insulin replacement coupled with measures that prevent, correct, or compensate for compromised glucose counterregulation.
Treatment of Hypoglycemia in Diabetes

Most episodes of asymptomatic hypoglycemia (detected by SMBG) and mild to moderate symptomatic hypoglycemia are effectively self-treated by ingestion of glucose tablets or carbohydrate in the form of juices, soft drinks, milk, crackers, candy, or a meal. A commonly recommended dose of glucose is 20 g (0.3 g/kg in children). However, the glycemic response to oral glucose is transient, usually less than 2 hours in insulin-induced hypoglycemia in T1DM (Fig. 32-15). Thus, ingestion of a more substantial mixed snack or meal shortly after the plasma glucose level is raised is generally advisable.

Parenteral treatment is necessary when a hypoglycemic patient is unable or unwilling (because of neuroglycopenia) to take carbohydrate orally. Glucagon is commonly injected subcutaneously or intramuscularly by a spouse or family member. The standard dose, 1 mg (15 µg/kg in children), can cause substantial but transient hyperglycemia (see Fig. 32-15). Intranasal administration of glucagon causes a glycemic response similar to that to injected glucagon. Although glucagon can be administered intravenously by medical personnel, intravenous glucose, 25 g initially, is the standard intravenous therapy. Because the glycemic response is transient, a subsequent glucose infusion is often needed and food should be provided orally as soon as the patient is able to take it safely.
HYPOGLYCEMIC DISORDERS

Hypoglycemia is most often caused by drugs, including those used to treat diabetes, just discussed, and alcohol. Among these, insulin, sulfonylureas, and perhaps metformin, used to treat diabetes, are the common offenders as discussed earlier in this chapter. Insulin and particularly sulfonylureas are possible causative agents even when there is no history of diabetes because these are sometimes taken surreptitiously, administered with criminal intent, or taken as the result of a pharmacy or other error. Established and putative hypoglycemia-causing drugs are listed in Table 32-7.

Ethanol inhibits gluconeogenesis, possibly because its metabolism to acetaldehyde and then acetate (by alcohol dehydrogenase

<table>
<thead>
<tr>
<th>Disorder Treated</th>
<th>Drug</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diabetes mellitus</td>
<td>Insulin, sulfonylureas and other insulin secretagogues, metformin</td>
</tr>
<tr>
<td>Infusions</td>
<td>Pentamidine, quinine, sulfonamides</td>
</tr>
<tr>
<td>Arrhythmia</td>
<td>Quinidine, disopyramide, cibenzoline</td>
</tr>
<tr>
<td>Pain</td>
<td>Acetylsalicylic acid</td>
</tr>
<tr>
<td></td>
<td><strong>Putative</strong></td>
</tr>
<tr>
<td>Infection</td>
<td>Ciprofloxacin, chloramphenicol, ketoconazole, oxeflecycline, ethionamide, isoniazid, p-aminosalicylic acid, p-aminobenzoate</td>
</tr>
<tr>
<td>Pain</td>
<td>Acetaminophen, indomethacin, propanolol, phenylbutazone</td>
</tr>
<tr>
<td>Hypertension, heart disease</td>
<td>-Adrenergic antagonists (nonselective &gt;-selective), angiotensin-converting enzyme inhibitors</td>
</tr>
<tr>
<td>Edema</td>
<td>Furosemide, acetazolamide</td>
</tr>
<tr>
<td>Depression</td>
<td>Monoamine oxidase inhibitors, fluoxetine, imipramine</td>
</tr>
<tr>
<td>Psychoses</td>
<td>Haloperidol, chlorpromazine, perhexiline</td>
</tr>
<tr>
<td>Hyperlipidemia</td>
<td>Clofibrate, bezafibrate</td>
</tr>
<tr>
<td>Allergies</td>
<td>Orphenadrine, diphenhydramine</td>
</tr>
<tr>
<td>Gastric hyperacidity</td>
<td>Cimetidine, ranitidine</td>
</tr>
<tr>
<td>Gout</td>
<td>Colchicine, sulfinpyrazone</td>
</tr>
<tr>
<td>Seizures</td>
<td>Phenytoin</td>
</tr>
<tr>
<td>(Anesthetics)</td>
<td>Enflurane, halothane</td>
</tr>
</tbody>
</table>
7% experienced hypoglycemia. 14% experienced-induced hypoglycemia include therapy of longer duration and increased doses, previous pentamidine therapy, and renal insufficiency. 126,127

Hypoglycemia occurs commonly in severe malaria. Although associated with relative hyperinsulinemia attributed to quinine-induced insulin release in some patients, hypoglycemia can occur in the absence of quinine therapy 128 and in the absence of hyperinsulinemia in quinine-treated patients. 129 Nonetheless, quinine has been reported to cause hypoglycemia in patients not afflicted by malaria. 130 and treatment of malaria with quinine, compared with the artemisinin-derivative artesunate, has been associated with higher plasma insulin/glucose ratios and lower glucose turnover rates and reported to produce higher rates of postadministration hypoglycemia. 131 Among anilinthrugs, quinidine, disopyramide, 132 and cibenzoline 133 have been reported to cause hypoglycemia.

Hypoglycemia has been attributed to many other drugs (see Table 32-7). In many of the cases, other potential causes of hypoglycemia have been present. For example, although hypoglycemia attributed to propranolol has been reported in healthy children, 134,135 most of the reported incidents occurred in insulin-treated diabetes. Although nonselective -adrenergic antagonists such as propranolol would be expected to reduce symptoms of developing hypoglycemia and impair epinephrine-mediated glucose counterregulation, 136 compelling evidence that these drugs increase the frequency of clinical hypoglycemia in insulin-treated diabetes has not been forthcoming. Nonetheless, it would be reasonable to use a relatively selective -adrenergic receptor antagonist (e.g., metoprolol or atenolol) in such patients.

Critical Illnesses

Among hospitalized patients, drugs, particularly insulin, are still the most common cause of hypoglycemia. 137 However, serious diseases, particularly renal failure but also hepatic or cardiac failure, sepsis, or inanition are second only to drugs.

In addition to appropriate glucoregulatory signals and a sufficient supply of gluconeogenic precursors, maintenance of the postabsorptive plasma glucose concentration requires a structurally and functionally intact liver. Renal glucose production notwithstanding, total hepatectomy results in profound hypoglycemia. 138 Extensive liver disease is required to produce hypoglycemia. Hepatogenous hypoglycemia occurs most commonly when destruction of the liver is rapid and massive (e.g., toxic hepatitis). It has been reported in fulminant viral hepatitis, 139 in fatty liver attributed to alcohol repose and in cholangitis and biliary obstruction. It is unusual in common forms of cirrhosis and hepatitis, although glucose metabolism is altered demonstrably (with lower postabsorptive plasma glucose concentrations, diminished glycemic responses to glucagon, and reduced hepatic glycogen contents) in uncomplicated viral hepatitis. 140 It is also unusual in metastatic liver disease despite extensive hepatic replacement. 141 Hypoglycemia can be caused by primary malignant tumors but is the result of a glucoregulatory abnormality, insulin-like growth factor II overproduction (see "NonBeta Cell Tumors").

The pathogenesis of hypoglycemia in occasional patients with severe cardiac failure is unknown. Possibilities include hepatic congestion and hypoxia, inanition, and gluconeogenic precursor limitation. The finding of elevated blood lactate levels associated with hypoglycemia 142 raises the possibility of inhibited gluconegogenesis. Postabsorptive hypoglycemia occurs in some patients with renal failure, 143,144,145,146 and the finding of a high frequency of renal insufficiency among patients with low plasma glucose levels 147 suggests that compromised glucose counterregulation may be a feature of renal failure. However, the pathogenesis of hypoglycemia in such patients is not known; it may involve multiple mechanisms. Most patients with hypoglycemia attributed to renal failure are cachectic. One such patient had reduced glucose turnover, diminished gluconeogenesis from alanine, and reduced alanine turnover. 148 During fasting, plasma glucose levels fell, blood lactate levels did not increase, and blood alanine levels fell. Hypoglycemia was attributed to substrate limitation of gluconeogenesis. However, at least one patient did not respond to substrate (glycerol, alanine) administration. 149 Glycemic responses to glucagon have been found to be reduced, suggesting impaired glycogenolysis, in some 150,151 but not all 152 studies.

Some patients with hypoglycemia attributed to renal failure have had diabetes, 153,154 but hypoglycemia persisted or recurred after insulin or oral hypoglycemic agents were withdrawn. The kidneys are a major site of insulin clearance, and decreasing insulin requirements parallel decreasing renal function in patients with insulin-treated diabetes. In the absence of insulin or insulin secretagogue therapy, however, endogenous insulin secretion should decrease and hypoglycemia would, therefore, not be expected. The mechanism by which reduced renal glucose production contributes to hypoglycemia in end-stage renal disease is unknown. However, the capacity of the normal liver to produce substantial amounts of glucose, loss of glucose-producing renal parenchyma would not be expected to contribute to hypoglycemia in patients with glucose-6-phosphatase deficiency. 155 Thus, one functioning kidney is not sufficient to provide normal endogenous glucose production in the virtual absence of hepatic glucose production.

Sepsis is a relatively common cause of hypoglycemia. 156,157 Increased glucose utilization (by skeletal muscle 158 and by macrophage-rich tissues such as liver, spleen, and leptom 159 ) and, at least initially, glucose production characterize experimental sepsis. 160,161 Hypoglycemia develops when hepatic glucose production decreases. 162,163 The factors responsible for the increased glucose turnover and the ultimate failure of glucose production to keep pace in sepsis are not entirely clear. Cytokines such as tumor necrosis factor (TNF) and interleukin-6, among others, are thought to increase glucose utilization. 164,165,166,167 The initial increase in glucose production is at least in part mediated by increased gluconeogenesis, 168 and catecholamines 169 and catecholaminergic responses to accelerated gluconeogenesis. These may also be stimulated by cytokines. 170,171 For example, TNF-infusion increases glucose production, an effect attributable to TNF-stimulated glucagon secretion, in dogs. 172 The later decline in glucose production, which results in hypoglycemia, is not the result of glucose counterregulatory failure. Rather, it appears to be the result of decreased hepatic responsiveness to appropriate glucoregulatory stimuli, that is, low insulin and high glucagon and epinephrine levels. 173 Hepatic hyperfusion is a plausible mechanism. It has been suggested that inactivation (oxidation) of catecholamines by superoxide anions plays a role in the pathogenesis of septic shock. 174,175

Hypoglycemia caused by inanition 176 is thought to be rare in developed countries but has been reported in the United States. 177 Because hypoglycemia can persist despite high rates of glucose infusion, such patients must have high rates of glucose utilization. Beyond this, the pathogenesis of hypoglycemia is unknown. An entirely speculative suggestion is that glucose becomes the sole oxidative fuel in the setting of total body fat depletion and that high rates of glucose utilization exceed the capacity to produce glucose because of limitation of substrate (e.g., amino acids). Postabsorptive hypoglycemia (with low blood alanine levels) has been reported to produce higher rates of postadmission hypoglycemia.

Hypoglycemia in children with deficient secretion of cortisol, growth hormone, or both is generally preceded by a period of caloric deprivation. That is consistent with the observation that hypoglycemia can sometimes be provoked by 24 to 30 hours of fasting in children with hypopituitarism who do not exhibit hypoglycemia after an overnight fast. 178,179 This intolerance of fasting is largely corrected by glucocorticoids.
TABLE 32-8  — Biochemical Patterns in Patients with Various Causes of Hyperinsulinemic Hypoglycemia

<table>
<thead>
<tr>
<th>Insulin</th>
<th>C Peptide</th>
<th>Proinsulin</th>
<th>Sulfonylurea</th>
<th>Insulin Antibody</th>
<th>Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Exogenous insulin</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Insulinoma, CHI</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+</td>
<td></td>
<td></td>
<td>Sulfonylurea</td>
<td></td>
</tr>
<tr>
<td>±</td>
<td></td>
<td>±</td>
<td></td>
<td>Insulin autoimmune</td>
<td></td>
</tr>
<tr>
<td>±</td>
<td>-</td>
<td></td>
<td></td>
<td>Insulin receptor autoimmune</td>
<td></td>
</tr>
</tbody>
</table>

* 20% of insulin value. Congenital hyperinsulinism. Free C peptide and proinsulin. § Insulin receptor antibody +.

replacement, whereas growth hormone replacement has a lesser effect. The findings suggest that a defect in gluconeogenesis causes hypoglycemia when hepatic glycogen stores are depleted.

Cortisol supports gluconeogenesis both by increasing gluconeogenic enzyme activities and by mobilizing gluconeogenic precursors to the liver (and the kidneys). However, oral alanine administration only partially reverses hypoglycemia. Finally, because cortisol deficiency causes reduced epinephrine secretion presumably because of reduced induction of adrenomedullary phenylethanolamine N-methyltransferase by adrenocortical cortisolepinephrine deficiency might contribute to the pathogenesis of hypoglycemia in this setting. Glucagon secretion is not reduced in such patients. Thus, given the key role of glucagon in glucose counterregulation, it is not surprising that glucose recovery, at least from short-term hypoglycemia, is generally normal in children with deficient secretion of cortisol, growth hormone, or both. Adults with hypopituitarism occasionally suffer postabsorptive hypoglycemia, particularly when glucose utilization or loss is increased, as during exercise or in pregnancy, respectively, or when glucogenic acidosis is impaired, as after alcohol ingestion. Again, these observations suggest that impaired gluconeogenesis becomes limiting to glucose production in the setting of glycogen depletion resulting from caloric deprivation.

As discussed earlier (see "Glucose Counterregulation"), hypoglycemia develops or progresses when both glucagon and epinephrine are deficient and insulin is present. This TABLE 32-9 — Plasma Glucose, Insulin, C-Peptide, and Proinsulin Concentrations used by Service (1999) and by Marks and Teale (1996) to Diagnose Fasting Hypoglycemia Caused by Endogenous Hyperinsulinism (e.g., in a Patient with an Insulinoma)

<table>
<thead>
<tr>
<th>Service</th>
<th>Marks and Teale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucose</td>
<td>2.5</td>
</tr>
<tr>
<td>mmol/L</td>
<td>45</td>
</tr>
<tr>
<td>mg/dL</td>
<td>6</td>
</tr>
<tr>
<td>Insulin</td>
<td>0.2</td>
</tr>
<tr>
<td>pmol/L</td>
<td>0.6</td>
</tr>
<tr>
<td>C peptide</td>
<td>5</td>
</tr>
</tbody>
</table>

combination occurs in patients with established T1DM. They must be treated with insulin, have no glucagon response to hypoglycemia, and typically have a reduced epinephrine response to hypoglycemia and are, as a result, at high risk for iatrogenic hypoglycemia as discussed earlier (see "Hypoglycemia in Diabetes Mellitus"). However, systematic studies of an unusual patient suggest that clinical postabsorptive hypoglycemia may not develop when both glucagon and epinephrine are deficient but insulin secretion decreases normally as plasma glucose levels decline. That patient, with hypopituitarism attributed to sarcoidosis involving the hypothalamus, was shown to have no glucagon or epinephrine response to insulin-induced hypoglycemia; glucose recovery from that short-term experimental hypoglycemia was impaired markedly, as expected. Plasma glucose concentrations were relatively low after overnight fasts (3.2 ± 0.5 mmol/L, 58 ± 8 mg/dL). However, during a 72-hour fast plasma insulin levels declined to about 30 pmol/L (5 µU/mL); glucose levels declined to only 2.8 mmol/L (50 mg/dL), a value indistinguishable from that of similarly fasted healthy subjects; and symptoms of hypoglycemia did not occur.

Hypoglycemia is not a feature of the epinephrine-deficient state that results from bilateral adrenalectomy if glucocorticoid replacement is appropriate. and hypoglycemia does not occur during pharmacologic blockade of catecholamine actions when other glucose counterregulatory systems are intact. Hypoglycemia has been attributed to epinephrine deficiency. For example, urinary and plasma epinephrine responses to hypoglycemia are reduced in ketotic hypoglycemia of childhood, and therapeutic responses to ephedrine, a catecholamine-releasing drug, have been reported in uncontrolled studies of such patients. Also, some patients have been reported to have diminished glycemic responses to glucagon during fasting.

Postabsorptive hypoglycemia has also been attributed to epinephrine deficiency in one member of each of three sets of twins. Compared with their unaffected twins, the hypoglycemic children had reduced, but not absent, urinary epinephrine responses to infused 2-deoxyglucose and to hypoglycemia induced by fasting. The glucagon secretory responses were not evaluated, and the affected infants had inappropriately high insulin levels while they were hypoglycemic. Finally, reduced epinephrine excretion in infants of diabetic mothers has been associated with the occurrence of neonatal hypoglycemia.

Isolated glucagon deficiency would be expected to result in lowered postabsorptive plasma glucose concentrations but not frank hypoglycemia if insulin secretion was suppressed appropriately and epinephrine secretion were intact. A seemingly convincing example of hypoglycemia attributable to isolated glucagon deficiency has been described in an abstract but the case was never published. Postabsorptive hypoglycemia has been reported in another glucagon-deficient adult, but cortisol secretion and growth hormone secretion were also deficient. Neonatal hypoglycemia has also been attributed to glucagon deficiency. However, plasma insulin levels were inappropriate high during hypoglycemia.

Elevated plasma levels of glucose counterregulatory hormones during hypoglycemia exclude deficiencies of these. Levels that are not elevated during a spontaneous episode of hypoglycemia provide a diagnostic clue that requires definitive testing. In a patient with postabsorptive hypoglycemia that is not readily explained, it is my practice to seek clinical clues of hypopituitarism or primary adrenocortical insufficiency and often assess the plasma cortisol response to cosyntropin (synthetic ACTH) and to pursue such clues with definitive testing (e.g., the responses to insulin-induced hypoglycemia). Given the evidence, just summarized, that isolated
Postabsorptive hypoglycemia is occasionally caused by nonbeta cell tumors (nonslet cell tumor hypoglycemia). The majority are large retropitoneal, intrathoracic, or extrathoracic tumors that are typically slow-growing and malignant. Epithelial tumors that can cause hypoglycemia include hepatomas, adrenocortical carcinomas, and carcinoid tumors. More common carcinomas and hemolologic or lymphoid malignancies rarely cause hypoglycemia. Affected patients often have relatively high rates of glucose utilization, a pattern resembling that of hyperinsulinemia. Reports of a few patients with hypoglycemia attributed to ectopic insulin secretion have been published, but plasma insulin and C-peptide levels are suppressed appropriately during hypoglycemia in the vast majority of patients with nonbeta cell tumor hypoglycemia.

Similarly, insulin-like growth factor I (IGF-I) levels are typically suppressed. Overproduction of insulin-like growth factor II (IGF-II), specifically an incompletely processed form ("big IGF-II") that does not complex normally with circulating binding proteins and thus more readily gains access to target tissues, is the cause of hypoglycemia in most patients. The diagnosis is usually not difficult. The tumors are often apparent clinically and plasma insulin, C-peptide, and proinsulin levels are low during hypoglycemia. Free IGF-II levels (and levels of proIGF-II [E212]) are elevated. It should be noted, however, that both of these are often elevated in patients with renal failure. Presumably because of negative feedback mediated by IGF-II, growth hormone secretion is suppressed. Thus, serum IGF-I levels are low and the ratio of IGF-II to IGF-I is distinctly elevated.

Hypoglycemia related to excessive endogenous insulin secretion can be caused by:

1. A primary pancreatic islet beta cell disorder, typically a beta cell tumor (insulinoma), sometimes multiple insulinomas, or, especially in infants or young children, a functional beta cell disorder with beta cell hyperplasia or without an anatomic correlate.
2. A beta cell secretagogue, often a sulfonylurea, theoretically a beta cell stimulating autoantibody.
3. An antibody to insulin.

None of these is common. Endogenous hyperinsulinism is more likely in an overtly well individual with postabsorptive hypoglycemia, that is, a person with no relevant drug history or critical illness and no clinical clues to hormone deficiencies or a nonbeta cell tumor. In such an individual, accidental, surreptitious, or even malicious administration of a sulfonylurea, another insulin-releasing drug, or insulin should also be considered.

The critical pathophysiological feature of endogenous hyperinsulinism is failure of insulin secretion, assessed by plasma insulin and C-peptide levels, to fall to very low rates during hypoglycemia. The plasma insulin, C-peptide, proinsulin, sulfonylurea, and insulin antibody patterns in the various diagnostic categories (including exogenous as well as endogenous hyperinsulinism) are shown in Table 32-8.

The plasma glucose, insulin, C-peptide, and proinsulin levels advocated by Service and by Marks and Teale are summarized in Table 32-9. Insulin and C-peptide levels need not be high in the absolute (i.e., relative to euglycemic postabsorptive norms) but only inappropriately high during postabsorptive hypoglycemia. The diagnostic concept of relative hyperinsulinemia is fundamentally important. Measurements of insulin and C-peptide levels when the patient is not hypoglycemic are not useful diagnostically. Plasma insulin, C-peptide, and proinsulin (and sulfonylurea) levels need to be determined when the patient is clearly hypoglycemicplasma glucose at least less than 2.8 mmol/L (50 mg/dL), preferably with symptoms in the postabsorptive state.

This determination accomplishes two diagnostic goals:

1. It generally establishes that the patient does, in fact, have postabsorptive hypoglycemia. Even during a prolonged (e.g., 48 to 72 hours) fast, plasma glucose levels rarely fall to less than 2.8 mmol/L in healthy adult men although lower levels sometimes occur in healthy adult women and children. However, healthy subjects should not have symptoms of hypoglycemia and their insulin, C-peptide, and proinsulin levels should be low.
2. It determines whether the patient has hyperinsulinemic or hypoinsulinemic hypoglycemia, the latter excluding hyperinsulinism as a diagnostic consideration.

Plasma C-peptide levels are low in endogenous hyperinsulinism, whether that is therapeutic, surreptitious, or malicious. Plasma proinsulin concentrations, like insulin and C-peptide levels, are disproportionately high in patients with insulinomas and related disorders. Sulfonylureas produce glucose, insulin, and C-peptide patterns indistinguishable from those produced by a primary beta cell disorder, but a sulfonylurea is measurable. (This potential diagnostic problem could be confounded by the availability of new nonsulfonylurea insulin-releasing drugs such as repaglinide and nateglinide.) Antibodies to insulin, plasma proinsulin, and insulin antibody patterns in the various diagnostic categories are shown in Table 32-8.

Sulfonylureas produce glucose, insulin, and C-peptide patterns indistinguishable from those produced by a primary beta cell disorder, but a sulfonylurea is measurable. (This potential diagnostic problem could be confounded by the availability of new nonsulfonylurea insulin-releasing drugs such as repaglinide and nateglinide.) Antibodies to insulin, plasma proinsulin, and insulin antibody patterns in the various diagnostic categories are shown in Table 32-8.

Figure 32-16 Diagnostic algorithm for suspected hypoglycemia. GI, gastrointestinal.

Hypoglycemia can also be caused, rarely, by insulin receptor-stimulating autoantibodies. Plasma glucose and C-peptide levels are low but insulin levels tend to be high, presumably because receptor-bound antibodies impair the clearance of insulin. Antibodies that stimulate beta-cell insulin secretion in vitro have been described in sera from patients with hypoglycemia, but a corresponding clinical syndrome has not been defined. Finally, as noted earlier, ectopic insulin secretion has been described. Nonetheless, insulin arteriovenous differences across such tumors remain to be documented.

In summary, the diagnostic strategy is to measure plasma glucose, insulin, C-peptide, proinsulin, and sulfonylureas when the plasma glucose concentration is distinctly low and the patient has symptoms in the postabsorptive state. If hypoglycemia is not demonstrable after short-term (e.g., overnight) fasting, it may be necessary to extend the fast to 48 to 72 hours. Antibodies to insulin need not be measured during hypoglycemia. The need to measure these is sometimes debated because autocrine hyperglycemia appears to be rare in the United States. But, as discussed subsequently, it occurs around the world, the presence of insulin antibodies is key to the diagnosis, and insulin antibodies cause artifactual insulin elevations in the conventional double-antibody insulin radioimmunoassays. If the plasma glucose concentration after overnight fasting is unequivocally normal (e.g., 4.0 to 6.0 mmol/L [72 to 108 mg/dL]) on several occasions, a judgment has to be made on the basis of the degree of clinical suspicion of postabsorptive hypoglycemia. It is decided that the diagnosis should be pursued or if the plasma glucose level after overnight fasting is equivocally low (e.g., 3.0 to 4.0 mmol/L [54 to 72 mg/dL]), a prolonged fast is indicated. Intermediate attempts can be made by extending the fast on an outpatient basis, but a full 48- to 72-hour fast requires hospitalization. Although there are rare exceptions, the absence of symptomatic hypoglycemia after a 72-hour fast excludes a diagnosis of endogenous hyperinsulinism caused by a primary beta-cell disorder. Indeed, some submit that a negative 48-hour fast virtually excludes that diagnosis. Other diagnostic tests, such as the tolbutamide tolerance test and the C-peptide suppression test, are not recommended. An algorithm for the approach to a patient with suspected hypoglycemia is shown in Figure 32-16.

Insulinomas, the most common cause of hypoglycemia related to endogenous hyperinsulinism in adults, are rare. The estimated incidence is one case per 250,000 patient-years. However, because approximately 90% of insulinomas are benign, they are generally a treatable cause of potentially fatal hypoglycemia.
The fetus relies on a continuous supply of glucose from the maternal circulation. After birth, the neonate must make the transition to endogenous glucose production.

In addition to changes in the level of consciousness (irritability, lethargy, stupor), tremor, seizures, and coma, signs consistent with hypoglycemia in infants include hypoglycemia levels of 2.2 mmol/L (40 mg/dL) to 1.8 mmol/L (33 mg/dL). As noted earlier, unusually low plasma glucose concentrations are required to produce symptoms of hypoglycemia in patients with an insulinoma because of the shift of glycemic thresholds to lower plasma glucose concentrations caused by recurrent hypoglycemia. Although symptomatic hypoglycemia can occur after an overnight fast, it often follows exercise. Rarely, symptomatic hypoglycemia follows meals but postabsorptive hypoglycemia is typically also demonstrable in such patients. The common symptoms in patients with an insulinoma are listed in Table 32-10.

Given convincing clinical and biochemical evidence of insulinoma, it is useful to localize the tumor if that can be done noninvasively. Percutaneous imaging-ultrasoundography, computed tomography, magnetic resonance imaging-may or may not localize an insulinoma but usually demonstrates metastases in the minority of patients with a malignant insulinoma. Thus, one of these is advisable prior to surgery. Octreotide scans localize about half of insulinomas. Arteriography has been used extensively in the past, but false-positives as well as false-negatives occur. A variation is selective regional arterial calcium injections with the end point of a sharp increase in hepatic venous insulin concentrations, but this is seldom necessary in the clinical setting. The same is true of transhepatic portal venous sampling. The other hand, intraoperative pancreatic ultrasonography almost invariably localizes insulinomas that are not readily palpable by the surgeon. Surgical resection of solitary insulinomas is generally curative. Medical therapy of unresectable insulinomas includes administration of diazoxide; the somatostatin analogue octreotide is sometimes effective. The majority of infants and young children with hyperinsulinemic hypoglycemia do not have discrete insulinomas. Other causes of hypoglycemia related to endogenous hyperinsulinism are discussed later in this chapter (see "Hypoglycemia in Infancy and Childhood").

Autoimmune hypoglycemia are thought to be quite rare. Approximately 90% of the reported cases of hypoglycemia attributed to autoantibodies to insulin have been from Japan. A history of other autoimmune disorders, particularly Graves' disease, and of treatment with sulfhydryl medications, particularly methimazole, is common. Whereas most autoantibodies against insulin receptor are antagonists and cause (type B) insulin resistance, some are agonists and cause hypoglycemia. Again, clinical or biochemical evidence of other autoimmune disorders and the finding of acanthosis nigricans are common. Finally, the finding of beta cell stimulating antibodies in the sera of patients with hypoglycemia raised the possibility of a new form of hypoglycemia. Nonetheless, a corresponding clinical syndrome has not been defined.

Causes of hypoglycemia unique to, or typically with Onset in, Infancy and Childhood

Transient Intolerance of Fasting
- Premature or small-for-gestational age infants
- Hypopituitarism, adrenal hypoplasia, congenital adrenal hyperplasia
- Ketotic hypoglycemia of childhood

Hyperinsulinism
- Infant of a diabetic mother
- Maternal drugs (sulfonylurea, 2-adrenergic agonist)
- Congenital hyperinsulinism, insulinoma
- Miscellaneous: Rh incompatibility, Beckwith-Wiedemann syndrome, exchange transfusions

Enzyme Defects
- Carbohydrate metabolism: glycogen storage disease types I, III, and VI; glycogen synthase deficiency; fructose-1,6-bisphosphatase deficiency; fructose-1-phosphate aldolase deficiency; galactose-1-phosphate uridyltransferase deficiency
- Protein metabolism: branched-chain -keto acid dehydrogenase complex deficiency
- Fat metabolism: fatty acid oxidation defects including deficiencies in the carnitine cycle, the beta oxidation spiral, the electron transport system, and the ketogenesis sequence

A corresponding clinical syndrome has not been defined.

Causes of hypoglycemia unique to, or typically with their clinical onset in, infancy and childhood include (1) transient intolerance of fasting, (2) hyperinsulinism, and (3) enzyme defects in carbohydrate, protein, or fat metabolism. Hypoglycemia in children can also be caused by the mechanisms discussed earlier. These include drugs and critical illnesses. For example, 9 (18%) of 49 children receiving resuscitative care for altered consciousness, status epilepticus, respiratory failure, cardiac failure, or cardiopulmonary arrest were hypoglycemic with plasma glucose levels ranging from 0.1 mmol/L (2 mg/dL) to 1.8 mmol/L (33 mg/dL). Four of the nine were septic. Of the 10 who died, 5 were hypoglycemic.

In addition to changes in the level of consciousness (irritability, lethargy, stupor), tremor, seizures, and coma, signs consistent with hypoglycemia in infants include apnea, cyanotic spells, hyperthermia, hypotonia, and poor feeding (especially after feeding well). The definition of neonatal hypoglycemia is controversial. Comnath and co-authors suggested the following operational plasma glucose thresholds for glucose administration: (1) less than 2.5 mmol/L (45 mg/dL) in infants with clinical manifestations compatible with hypoglycemia and (2) less than 2.0 mmol/L (36 mg/dL) in infants at risk for hypoglycemia.

The fetus relies on a continuous supply of glucose from the maternal circulation. After birth, the neonate must make the transition to endogenous glucose production with intermittent exogenous glucose delivery. Probably because of their large brains relative to their body weights, infants have rates of glucose utilization approximately threefold higher than those of adults when expressed per unit of body weight. Correspondingly high rates of endogenous glucose production are required to maintain systemic glucose balance. Because mobilizable glycogen stores are limited and feeding is intermittent, the newborn is largely dependent on

### TABLE 32-10 -- Symptoms of Hypoglycemia in Patients with Insulinomas

<table>
<thead>
<tr>
<th>Symptom</th>
<th>Incidence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Various combinations of diplopia, blurred vision, sweating, palpitations, or weakness</td>
<td>85</td>
</tr>
<tr>
<td>Confusion or abnormal behavior</td>
<td>80</td>
</tr>
<tr>
<td>Unconsciousness or amnesia</td>
<td>53</td>
</tr>
<tr>
<td>Grand mal seizures</td>
<td>12</td>
</tr>
</tbody>
</table>


### TABLE 32-11 -- Causes of Hypoglycemia Unique to, or Typically with Onset in, Infancy and Childhood

- Transient Intolerance of Fasting
  - Premature or small-for-gestational age infants
  - Hypopituitarism, adrenal hypoplasia, congenital adrenal hyperplasia
  - Ketotic hypoglycemia of childhood

- Hyperinsulinism
  - Infant of a diabetic mother
  - Maternal drugs (sulfonylurea, 2-adrenergic agonist)
  - Congenital hyperinsulinism, insulinoma
  - Miscellaneous: Rh incompatibility, Beckwith-Wiedemann syndrome, exchange transfusions

- Enzyme Defects
  - Carbohydrate metabolism: glycogen storage disease types I, III, and VI; glycogen synthase deficiency; fructose-1,6-bisphosphatase deficiency; fructose-1-phosphate aldolase deficiency; galactose-1-phosphate uridyltransferase deficiency
  - Protein metabolism: branched-chain -keto acid dehydrogenase complex deficiency
  - Fat metabolism: fatty acid oxidation defects including deficiencies in the carnitine cycle, the beta oxidation spiral, the electron transport system, and the ketogenesis sequence

A corresponding clinical syndrome has not been defined.

Causes of hypoglycemia unique to, or typically with their clinical onset in, infancy and childhood include (1) transient intolerance of fasting, (2) hyperinsulinism, and (3) enzyme defects in carbohydrate, protein, or fat metabolism. Hypoglycemia in children can also be caused by the mechanisms discussed earlier. These include drugs and critical illnesses. For example, 9 (18%) of 49 children receiving resuscitative care for altered consciousness, status epilepticus, respiratory failure, cardiac failure, or cardiopulmonary arrest were hypoglycemic with plasma glucose levels ranging from 0.1 mmol/L (2 mg/dL) to 1.8 mmol/L (33 mg/dL). Four of the nine were septic. Of the 10 who died, 5 were hypoglycemic.

In addition to changes in the level of consciousness (irritability, lethargy, stupor), tremor, seizures, and coma, signs consistent with hypoglycemia in infants include apnea, cyanotic spells, hyperthermia, hypotonia, and poor feeding (especially after feeding well). The definition of neonatal hypoglycemia is controversial. Comnath and co-authors suggested the following operational plasma glucose thresholds for glucose administration: (1) less than 2.5 mmol/L (45 mg/dL) in infants with clinical manifestations compatible with hypoglycemia and (2) less than 2.0 mmol/L (36 mg/dL) in infants at risk for hypoglycemia.

The fetus relies on a continuous supply of glucose from the maternal circulation. After birth, the neonate must make the transition to endogenous glucose production with intermittent exogenous glucose delivery. Probably because of their large brains relative to their body weights, infants have rates of glucose utilization approximately threefold higher than those of adults when expressed per unit of body weight. Correspondingly high rates of endogenous glucose production are required to maintain systemic glucose balance. Because mobilizable glycogen stores are limited and feeding is intermittent, the newborn is largely dependent on
levels), structural and enzymatic integrity of the liver (and kidneys), and availability of sufficient gluconeogenic precursors are essential. In the setting of relatively low plasma glucose concentrations, the combination of hypoinsulinemia and activated glucose counterregulatory systems also favors lipolysis. High nonesterified fatty acid levels provide an alternative fuel for tissues other than the brain and limit glucose utilization by muscle and fat. They also drive ketogenesis, thus providing an alternative fuel for tissues other than the brain during gluconeogenesis. Impairment of any of these adaptations to extrauterine life can cause transient neonatal hypoglycemia. Persistent defects cause recurrent or persistent hypoglycemia.

Transient intolerance of fasting occurs in preterm or small-for-gestational age infants; in hypopituitarism, adrenal hypoplasia, or congenital adrenal hyperplasia; or, later, in ketotic hypoglycemia of childhood. At least in the absence of seizure or coma, neonatal hypoglycemia (that developing in the first 72 hours after birth) is usually transient. It is particularly common in preterm or small-for-gestational age infants and is thought to result from incomplete development of gluconeogenic mechanisms although glucose counterregulatory signals may be impaired. Deficiencies of cortisol, growth hormone, or both can be congenital and cause hypoglycemia through mechanisms discussed earlier.

In general, children tolerate fasting less well than adults. Indeed, hypoglycemia is the rule after 24 to 48 hours of fasting in normal children. The syndrome of ketotic hypoglycemia of childhood, which typically has its onset between ages 2 and 5 years and remits spontaneously before age 10 years, may account for the fraction of children who develop fasting hypoglycemia. Hypoglycemia can usually be elicited by fasting during an oral glucose tolerance test. It appears to involve diminished mobilization of gluconeogenic precursors including alanine. Blood alanine levels are low during hypoglycemia, and alanine infusion increases plasma glucose levels. Glycolytic and gluconeogenic mechanisms appear to be intact and, aside from low epinephrine levels, gluconeogenic signals are appropriate. Deficient epinephrine secretion might impair alanine mobilization because epinephrine stimulates alanine turnover in humans. Nonetheless, epinephrine deficiency per se does not cause hypoglycemia, as discussed earlier.

The most common cause of hyperinsulinemic neonatal hypoglycemia is maternal diabetes. Infants of diabetic mothers are hyperglycemic (in proportion to the mother's hyperglycemia) and correspondingly hyperinsulinemic. Presumably reflecting chronic stimulation of fetal insulin secretion in utero and its failure to become suppressed normally as glucose levels fall shortly after birth, transient neonatal hypoglycemia occurs. Transient hyperinsulinism also underlies neonatal hypoglycemia in infants with RH factor incompatibility or with the Beck-with-Wiedemann syndrome (macroglomia, ophthalmocele, and visceromegaly). Hypoglycemia, resulting from hyperinsulinism stimulated by glucose flush during the procedure, can also follow exchange transfusion. Neonatal hypoglycemia can be caused by drug exposure, the mother, including agents that stimulate fetal insulin secretion (e.g., a sulfonylurea) or that produce maternal and fetal hyperglycemia and thus fetal hyperinsulinism (e.g., α₂-adrenergic agonist used to delay labor). Accidental or malicious administration of a sulfonylurea or insulin is a rare cause of hyperinsulinemic hypoglycemia in children.

In contrast to these causes of transient hypoglycemia, congenital hyperinsulinism (or persistent hyperinsulinemic hypoglycemia of infancy) may persist from the neonatal period or become apparent clinically in the first year of life. Patients in that age range rarely have a discrete insulinoma, although insulinomas are found in children who develop hyperinsulinemic hypoglycemia after the first year. Although partial pancreatectomy may become necessary, patients with congenital hyperinsulinism are treated medically initially with glucose administration for stabilization; frequent feedings; and diazoxide (often with a thiazide), octreotide, and glucagon, often tried in that sequence in the anticipation of amelioration of hypoglycemia over time because there is a high frequency of diabetes late after partial pancreatectomy.

Congenital hyperinsulinism is the most common cause of nontransient neonatal hypoglycemia. It is often inherited as an autosomal recessive trait and the result of mutations of the genes that encode adenosine triphosphatase-sensitive potassium (KATP) channels, specifically the sulfonylurea receptor (SUR1) or the channel itself (Kir6.2). Homozygous mutations result in diffuse beta-cell hypersecretion; focal beta-cell hypersecretion has been attributed to loss of maternal heterozygosity and expression of the paternal KATP mutation. In general, patients with KATP channel mutations suffer from severe neonatal hypoglycemia that is unresponsive to diazoxide. Successful treatment with the calcium channel antagonist nifedipine has been reported. Other causes of congenital hyperinsulinism which typically cause less marked hypoglycemia and are more likely to be responsive to diazoxide include autosomal dominant activating mutations of the glutamate dehydrogenase gene (the hyperinsulinism-hyperammonemia syndrome) and of the glucokinase gene. Hyperinsulinemic hypoglycemia has also been reported in an infant with phosphomannose isomerase deficiency and was found to be responsive to mannose administration.

Hypoglycemia that develops in infancy or childhood persists and can be treated with effective therapy can also be caused by enzymatic defects in carbohydrate metabolism (e.g., glycogen storage disease types I, III, and IV; glycogen synthase deficiency; fructose-1,6-bisphosphatase, phosphoenolpyruvate carboxykinase, or pyruvate dehydrogenase deficiencies; or in protein metabolism (e.g., branched-chain -keto acid dehydrogenase complex deficiency), in protein metabolism (e.g., branched-chain -keto acid dehydrogenase complex deficiency), or in fat metabolism (e.g., various defects in fatty acid oxidation). All of these disorders cause postabsorptive hypoglycemia except for hereditary fructose intolerance caused by fructose-1-phosphate aldolase deficiency and galactosaemia caused by galactose-1-phosphate uridyltransferase deficiency, which cause postprandial hypoglycemia. Although clinical features and biochemical patterns suggest a subset of diagnostic possibilities and in some instances provide a specific diagnosis, definitive diagnosis often requires either documentation of deficient enzyme activity in affected tissues or, increasingly, identification of a mutation in the relevant gene.

As first documented by Cori and Cori in 1952, deficient glucose-6-phosphatase activity causes glycogen storage disease type I (von Gierke's disease). Type I glycogen storage disease is the prototype glycogen storage disease. Because hydrolysis of glucose-6-phosphate to glucose is the common pathway for synthesis of glucose from both hepatic and nonhepatic sources, and gluconeogenesis and glycogenolysis are linked processes, they cause (1) profound post-absorptive hypoglycemia with hyperinsulinism; (2) activated glucose counterregulatory systems with elevated lactate, alanine, nonesterified fatty acid, ketone body, and triglyceride levels; and (3) metabolic acidosis with hyperuricemia. Hepatomegaly (caused by hepatocyte accumulation of fat as well as glycogen) is a universal finding. With the exception of hepatomegaly, the abnormalities can be reversed by the prevention of hypoglycemia with frequent feedings during waking hours and continuous intragastric glucose infusion during sleep or with

bedtime administration of large doses of uncooked cornstarch. Liver transplantation corrects hypoglycemia and the associated metabolic abnormalities. Adults with (presumably inadequately treated) type I glycogen storage disease have more frequency of hepatic adenomas and renal disease. Interestingly, renal cell carcinoma is a rare complication in patients with type I glycogen storage disease. The mechanism by which these patients maintain some level of endogenous glucose production is unclear.

The glucose-6-phosphatase system is complex. Most patients with type I glycogen storage disease have mutations of the gene encoding the catalytic subunit (type la). Others do not have such mutations; the defect has been attributed to mutations in the glucose-6-phosphatase translocase gene (type Ib or nontype la). Type Ib or nontype la glycogen storage disease is usually also caused by glycogen synthase deficiency, which, unlike the glycogen storage diseases, does not cause hepatomegaly. Because it blocks gluconeogenesis, fructose-1,6-phosphatase deficiency causes profound postabsorptive hypoglycemia with lactic acidosis, ketosis, and elevated alanine levels. Hyperlipidemia, hyperuricemia, and hepatomegaly (related to fat accumulation) occur as in type I glycogen storage disease. Hypoglycemia has also been attributed to phosphoenolpyruvate carboxykinase and pyruvate carboxylase deficiencies.

Finally, with respect to postabsorptive hypoglycemia and defects in glucose metabolism, CNS glucopenia occurs in patients with mutations of the GLUT-1 glucose transporter gene. Plasma glucose levels are normal but cerebrospinal glucose levels are low because of reduced GLUT-1 mediated glucose transport across the blood-brain barrier. Treatment includes the ketogenic (low-carbohydrate) diet designed to raise ketone levels and thus provide an alternative fuel to the brain. Hypoglycemia has also been attributed to GLUT-2 deficiency in the Fanconi-Bickel syndrome.

Postprandial, rather than postabsorptive, hypoglycemia can be a feature of hereditary fructose intolerance and, rarely, galactosemia. Fructose-1-phosphate aldolase deficiency, the enzymatic defect in hereditary fructose intolerance, causes vomiting and severe hypoglycemia after fructose ingestion. Fructose-1-phosphate...
accumulates and inhibits gluconeogenesis (at the phosphorylase level) and gluconeogenesis (at the mutant aldolase level). The patients are well when fructose is omitted from the diet. Galactose uridyltransferase deficiency, one of the causes of galactosemia, can also cause postprandial hypoglycemia, which has been attributed to inhibition of gluconeogenesis.  

Deficiencies of enzymes involved in protein metabolism that can cause postabsorptive hypoglycemia include that of the branched-chain keto acid dehydrogenase complex, the basis of branched-chain ketoaciduria (maple syrup urine disease).  The levels of leucine, isoleucine, and valine particularly leucine in plasma and urine are elevated. The pathogenesis of hypoglycemia is not entirely clear, although it results from defective gluconeogenesis. Hypoglycemia related to impaired gluconeogenesis also occurs in methylmalonic aciduria.  

Several defects that ultimately impair fatty acid oxidation result in postabsorptive hypoglycemia with hypoketonaemia. Normally, low-insulin, high-glucagon (and catecholamine) states such as fasting favor the mobilization of fatty acids from fat (lipolysis) and their transport to other tissues including the liver and skeletal and cardiac muscle. These regulatory conditions also favor fatty acid oxidation (with ATP formation) and ketogenesis over triglyceride, phospholipid, and cholesterol ester synthesis and peroxisomal oxidation. Mitochondrial fatty acid oxidation and ketogenesis require transport of fatty acids across the plasma membrane, formation of fatty acyl-CoA derivatives, and transport of the derivatives into mitochondria. Because the inner mitochondrial membranes are not permeable to long-chain (as opposed to medium-chain and short-chain) fatty acyl-CoA esters, the long-chain fatty acyl-CoA esters are transferred to fatty acylcarnitines at the outer surface of the membranes by carnitine palmitoyltransferase II, CPT-II, and converted to the fatty acyl-CoA esters by carnitine palmitoyltransferase I, CPT-I, at the inner surface of the membranes. Then they can be oxidized or converted to ketones.

Insulin decreases fat oxidation and ketogenesis by decreasing lipolysis and by increasing lipogenesis and the formation of malonyl-CoA, which inhibits CPT-I. Conversely, low insulin levels favor fatty acid oxidation and ketogenesis. High glucagon levels do so by decreasing malonyl-CoA. Catecholamines do so largely by stimulating lipolysis. Any defect in the sequence of events in the carnitine cycle (carnitine transport defect, CPT-I deficiency, carnitine-acylcarnitine translocase deficiency, defects in the beta oxidation spiral (long-chain acyl-CoA dehydrogenase [LCAD] deficiency, long-chain 3-hydroxyacyl-CoA dehydrogenase [LCHAD] deficiency, short-chain 3-hydroxyacyl-CoA dehydrogenase [SCHAD] deficiency, 2,4-dienoyl-CoA reductase deficiency, medium-chain acyl-CoA dehydrogenase [MCAD] deficiency, short-chain acyl-CoA dehydrogenase [SCAD] deficiency), several defects of electron transfer or defects in ketogenesis (hydroxymethylglutaryl [HMG]-CoA lyase deficiency, HMG-CoA synthetase deficiency) decreases fatty acid oxidation (and ketogenesis) and reciprocally increases glucose oxidation, resulting in hypoketonaemic postabsorptive hypoglycemia. Reduced plasma carnitine levels (20% to 50% of normal) are the rule in these disorders, but extremely low carnitine levels characterize the carnitine transport defect, a true carnitine deficiency state that is responsive to carnitine supplementation. CPT-I deficiency, a rare disorder, is treatable by administration of medium-chain triglycerides, which do not require the CPT system for oxidation, or by a high-carbohydrate, low-fat diet. CPT-II deficiency, which is typically seen with episodes of muscle pain and myoglobinuria but can also cause hypoglycemia, is more common.  

The child affected with a disorder of fatty acid oxidation typically presents with hypoketonicemia hypoglycemia; intravenous glucose causes prompt improvement. Some have presented with Reye's syndrome. All are at risk for sudden death, presumably from cardiac causes. Treatment includes provision of an adequate caloric intake, avoidance of fasting, and support of the plasma glucose concentration during intercurrent illnesses. The diagnosis of specific fatty acid oxidation defects is typically accomplished by blood acylcarnitine profiling, although molecular diagnosis is increasingly possible. Interestingly, the presence of a defect in fatty acid oxidation in a fetus may have implications for the mother. While carrying a fetus with a specific mutation (Glu474Gln) causing long-chain 3-hydroxyacyl-CoA dehydrogenase deficiency, 15 (79%) of 19 mothers suffered fatty liver of pregnancy or the HELLP (hemolysis, elevated liver enzyme levels, and low platelet count) syndrome. Given the array of causes of hypoglycemia in infancy and childhood just summarized, it is reasonable to suggest an extensive biochemical assessment during a hypoglycemic episode when the hypoglycemic mechanism is obscure. In addition to the concurrent plasma glucose concentration, this might include (1) plasma insulin, C peptide, sulfonlureas, growth hormone, and cortisol; (2) plasma or blood lactate, amino acids (including alanine), nonesterified fatty acids, and -hydroxybutyrate; (3) serum liver enzymes; (4) plasma acylcarnitine profile; and (5) urine ketones and organic acid profile.
The Postprandial (Reactive) Hypoglycemias

Postprandial (reactive, stimulative) hypoglycemia occurs exclusively after meals, typically within 4 hours after food ingestion. All disorders that cause postabsorptive hypoglycemia can also result in hypoglycemia detected after a meal. However, the diagnostic and therapeutic approach is that of postabsorptive hypoglycemia in such a patient.

Congenital deficiencies of enzymes of carbohydrate metabolism, such as those that cause hereditary fructose intolerance and galactosemia discussed earlier, are rare causes of postprandial hypoglycemia that becomes apparent early in life. Postprandial hypoglycemia can occur in individuals who have undergone gastric surgery that results in rapid movement of ingested food into the small intestine. Termed alimentary hypoglycemia, this is thought to be the result of marked early hyperinsulinemia caused by rapid increments in plasma glucose, enhanced secretion of incretins (gut factors that enhance glucose-stimulated insulin secretion), or both. Hypoglycemia occurs 1.5 to 3.0 hours after food ingestion. Symptoms of hypoglycemia must be distinguished from those of the dumping syndrome (abdominal fullness, nausea, and weakness) which occur less than an hour after ingestion. Administration of an α-glucosidase inhibitor (e.g., acarbose or miglitol) is a conceptually attractive treatment for alimentary hypoglycemia, although controlled trials indicating its efficacy are lacking.

The frequency, and even the existence, of clinically relevant idiopathic (functional) postprandial hypoglycemia is a matter of debate. Idiopathic postprandial hypoglycemia is often erroneously diagnosed by patients and by physicians. For example, only 16 of 118 patients suspected of having postprandial hypoglycemia in one series had both a plasma glucose concentration lower than the 10th percentile of asymptomatic control subjects and typical symptoms after an oral glucose load; only 5 of those 16 patients had similar symptoms after their regular meals. Furthermore, most patients thought to have hypoglycemic symptoms and low glucose levels after glucose ingestion have normal glucose levels after a mixed meal. In one series in which blood glucose was measured during symptomatic episodes, only 5% of 132 episodes were associated with blood glucose levels of 2.8 mmol/L (50 mg/dL) or less. Service and colleagues reported on five adults judged to have hyperinsulinemic ("pancreatogenous") postprandial hypoglycemia. The patients were assessed thoroughly, and insulinomas were not found. Insulin levels were judged to be inappropriately high during hypoglycemia (but norms from prolonged fasts were used); antibodies to insulin were not reported. Mutations of the Kir6.2 and SUR1 genes were not found. An alternative theoretical possibility would be an attenuated resumption of glucagon secretion during the transition from the postprandial to the postabsorptive state. That would plausibly explain the pathogenesis of the postprandial syndrome including compensatory enhancement of epinephrine secretion, the production of symptoms attributable to the enhanced sympa-thochromaffin response, and the prevention of severe hypoglycemia and restoration of euglycemia.

Lower plasma glucagon levels in persons with plasma glucose nadirs less than 2.8 mmol/L (50 mg/dL) than in those with higher nadir glucose levels after an oral glucose load have been reported. However, glucagon levels were also lower at baseline and were not discernibly lower after glucose ingestion in the two individuals with the lowest plasma glucose nadirs (1.5 mmol/L [27 mg/dL] and 1.3 mmol/L [2.3 mg/dL]). On the other hand, in another report, similarly selected individuals (nadir glucose less than 2.8 mmol/L [50 mg/dL]) had normal pancreatic glucagon levels after glucose ingestion. Nonetheless, enhanced, presumably compensatory, plasma epinephrine responses have been reported in individuals with sweating, tremor, and greater heart rates temporally related to the glucose nadir late after glucose ingestion. In such individuals, the postprandial syndrome may be the result of an appropriately enhanced sympathochromaffin response to falling plasma glucose concentrations rather than hypoglycemia per se.

A diagnosis of postprandial hypoglycemia should not be made on the basis of seemingly low plasma glucose concentrations during an oral glucose tolerance test. The lower limits of normal for venous plasma glucose concentrations late after glucose ingestion can be defined statistically. For example, in 650 individuals who remained asymptomatic after ingestion of 100 g of glucose, nadir glucose concentrations were lower 5th percentile, 2.4 mmol/L (43 mg/dL); lower 10th percentile, 2.6 mmol/L (47 mg/dL); and lower 25th percentile, 3.0 mmol/L (54 mg/dL). The absence of symptoms in response to such seemingly low plasma glucose concentrations is most plausibly explained by venous sampling. Although glucose arteriovenous differences are negligible in the postabsorptive state, there is substantial glucose extraction across the forearm under hyper-insulinemic conditions. Thus, arterial glucose levels—those relevant to glucose delivery to the brain—must be lower than those measured by venous sampling. The diagnosis requires documentation of appropriate symptoms temporally related to a low plasma glucose concentration after a mixed meal and relief of those symptoms as the plasma glucose concentration rises (Whipple's triad). The diagnosis cannot be made on the basis of an oral glucose tolerance test.

Diet low in carbohydrate and high in protein are commonly recommended to patients thought to have postprandial hypoglycemia. Their efficacy has not been established by controlled clinical trials. Frequent feedings and avoidance of simple sugars are also advised. Anticholinergic drugs have been reported to be beneficial in uncontrolled studies of patients with the postprandial syndrome hypoglycemia but may cause undesirable side effects. The α-adrenergic antagonist propranolol reduces symptoms (except diaphoresis) in patients with postgastrectomy postprandial hypoglycemia, and administration of pectin is said to decrease symptoms in such patients. As mentioned earlier, to the extent that an excessive initial increase in plasma glucose plays a role in the pathogenesis of alimentary hypoglycemia, administration of an α-glucosidase inhibitor to delay carbohydrate digestion is a conceptually attractive treatment. Finally, such patients have been treated surgically with reversal of a segment of proximal jejunum.
TREATMENT OF POSTABSORPTIVE HYPOGLYCEMIA

In view of the vulnerability of the brain to prolonged hypoglycemia, the plasma glucose concentration must be raised at least to normal levels as rapidly as possible and recurrence of hypoglycemia must be prevented. Because it is self-limited, postprandial hypoglycemia rarely requires urgent treatment. In contrast, postabsorptive hypoglycemias are typically persistent or progressive and require short-term and long-term therapy.

The urgent treatment of iatrogenic hypoglycemia in individuals with diabetes with oral carbohydrate or glucose per se or with parenteral glucagon or glucose was discussed earlier under "Hypoglycemia in Diabetes Mellitus." Clinical improvement should occur within about 15 to 20 minutes after the plasma glucose level is increased and maintained provided that brain damage has not occurred. Whenever possible, the presence of hypoglycemia should be documented before therapy and the response to therapy should be followed by frequent measurements of the plasma glucose level. If these are not available and there is no clinical response within 15 minutes, the initial therapy should be repeated and access to plasma glucose monitoring and intravenous glucose infusion should be obtained as soon as possible. Even if there is a response to initial therapy, glucose monitoring is essential to ensure maintenance of the plasma glucose concentration.

Although CNS function usually recovers promptly after restoration of the plasma glucose level, recovery may be delayed, perhaps because of cerebral edema. Unconsciousness lasting more than 30 minutes after the plasma glucose concentration has been raised to normal and maintained is referred to as posthypoglycemic coma. It has been treated with intravenous mannitol (40 g as a 20% solution over 20 minutes) or glucocorticoids (e.g., dexamethasone, 10 mg), or both, along with maintenance of normal plasma glucose levels.

Definitive treatment of the postabsorptive hypoglycemia requires correction of the underlying defect whenever possible. When that is not possible, attempts must be made to increase exogenous delivery or endogenous glucose production and to limit glucose utilization by tissues other than the brain. Although the judicious use of snacks is a useful component of therapy for individuals with diabetes, frequent feedings are less than ideal for the long-term treatment of chronic hypoglycemia. One problem is weight gain. However, frequent feedings, even overnight gastric infusions, are sometimes necessary when other measures are inadequate.

Hypoglycemia caused by drugs is limited to the duration of action of the offending drug. The management is straightforward: discontinuation of the drug (at least temporarily), maintenance of the plasma glucose level while drug action continues, and adjustment of subsequent drug regimens to avoid recurrent hypoglycemia if the causative drug is known. Therapy is more difficult if the drug is used surreptitiously or given accidentally or maliciously.

As discussed earlier, postabsorptive hypoglycemia related to endogenous hyperinsulinism is often curable by the surgical removal of an insulinoma. If this is not possible because of multiple or metastatic tumors or the absence of a definable lesion, diazoxide is sometimes effective. Diazoxide (100 to 800 mg/day in adults and 5 to 30 mg/kg/day in infants) raises the plasma glucose concentration by suppressing insulin secretion. Diazoxide is bound tightly to albumin and has a plasma half-time of 20 to 30 hours. When given by rapid intravenous injection, it is a potent hypotensive drug, but when given orally or by slow intravenous infusion, it has little hypotensive action; indeed, hypertensive responses may occur. Although chemically related to the thiazide diuretics, diazoxide causes sodium retention. Coadministration of a thiazide diuretic both limits sodium retention and potentiates the hyperglycemic action of diazoxide. Both edema formation and gastrointestinal side effects (anorexia, nausea, sometimes vomiting) are dose-related. Generalized growth of lanugo hair (hypertrichosis lanuginosa) may occur during prolonged therapy. Allergic reactions, including skin rashes and agranulocytosis, are rare. Other treatments include octreotide and calcium channel antagonists.

The treatment of hypoglycemia associated with nonbeta cell tumors involves short-term measures pending effective medical, surgical, or radiotherapeutic treatment of the tumor. Administration of a glucocorticoid or growth hormone sometimes alleviates hypoglycemia. The former, but not the latter, has been reported to reduce IGF-II levels. Hypoglycemia resulting from glucocorticoid deficiency is corrected by replacement therapy. Hypoglycemia is rarely an indication for growth hormone replacement. Remissions of autoimmune hypoglycemias have been associated with immunosuppressive therapy, including glucocorticoids, but controlled trials are lacking. The treatment of hypoglycemia related to infection, hepatic or renal disease, cardiac failure, or sepsis includes short-term measures and, when possible, treatment or management of the underlying disease process. The treatment of the hypoglycemias of infancy and childhood and that of postprandial hypoglycemia were discussed earlier.
APPROACH TO THE PATIENT WITH HYPOGLYCEMIA

In addition to recognition and documentation of hypoglycemia and often urgent treatment, management of hypoglycemia requires diagnosis of the hypoglycemic mechanism leading to treatment that prevents, or at least minimizes, recurrent hypoglycemia. The differential diagnosis of hypoglycemia, discussed earlier, is summarized in Table 32-2. A diagnostic algorithm is shown in Figure 32-16. The thought process is summarized in Table 32-12.

Recognition and Documentation of Hypoglycemia

Hypoglycemia is sometimes detected serendipitously. However, a report of a distinctly low plasma glucose measurement in a person who does not have a history of corresponding symptoms raises the possibility of pseudohypoglycemia—a measured low glucose level resulting from ongoing metabolism of glucose by the formed elements of the blood after the sample is drawn. Pseudohypoglycemia is particularly common when leukocyte, thrombocyte, or erythrocyte counts are abnormally high, but it can occur in the absence of these if separation of the plasma or serum from the formed elements is delayed. Nonetheless, a report of a distinctly low plasma glucose concentration measured in a reliable laboratory cannot be ignored.

Hypoglycemia is often sought because of a history of suggestive symptoms. But the symptoms are not specific for hypoglycemia, and a normal plasma glucose concentration measured when the patient is free of those symptoms does not exclude the possibility of hypoglycemia at the time of those earlier symptoms. Convincing documentation of hypoglycemia requires demonstration of Whipple’s triads—symptoms consistent with hypoglycemia, a low plasma glucose concentration, and relief of those symptoms after the plasma glucose concentration is raised. This can be accomplished easily, by measuring the plasma glucose concentration and then administering glucose, if the patient is seen while symptomatic. When the patient is not symptomatic when seen but has a history of a previous low measured plasma glucose concentration, of previous symptoms suggestive of hypoglycemia, or both, the initial diagnostic strategy is to obtain samples from the patient under conditions in which Whipple’s triad would be expected to be demonstrable if a hypoglycemic disorder exists. In most instances, that condition would be the postabsorptive state, initially after an overnight fast but after a longer fast if necessary. If the history suggests only postprandial hypoglycemia, that condition would be after a mixed meal.

<table>
<thead>
<tr>
<th>TABLE 32-12 – Diagnostic Approach to an Adult with Documented Fasting Hypoglycemia</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Think of drugs, critical illness, endocrine deficiency, nonbeta cell tumor, and hyperinsulinism while supporting the plasma glucose concentration if necessary.</td>
</tr>
<tr>
<td>2. Search the history, physical examination, and available laboratory data for clinical clues to the hypoglycemic mechanism and pursue the plausible mechanism or mechanisms:</td>
</tr>
<tr>
<td>a. Insulin-treated or sulfonylurea-treated diabetes</td>
</tr>
<tr>
<td>b. Use of other drugs known or suspected to cause hypoglycemia</td>
</tr>
<tr>
<td>c. Hepatic, renal, or cardiac failure; sepsis; or inanition</td>
</tr>
<tr>
<td>d. Anorexia, weight loss, change in skin pigmentation, known pituitary or adrenocortical disease, hypotension, hypoatremia, hyperkalemia</td>
</tr>
<tr>
<td>e. Known nonbeta cell tumor, mass on examination or imaging studies</td>
</tr>
<tr>
<td>3. In the absence of clinical clues, consider medication error, endogenous hyperinsulinism, and surreptitious or malicious sulfonylurea or insulin administration.</td>
</tr>
<tr>
<td>4. A metabolic enzyme deficiency is rarely first detected in an adult.</td>
</tr>
</tbody>
</table>

IGF, insulin-like growth factor.
Urgent Treatment

If the patient is hypoglycemic when seen, urgent treatment is often necessary. When possible, a sample for documentation of the plasma glucose concentration by a quantitative analytical method (not a blood glucose monitor) should be obtained prior to treatment. Obviously, glucose administration based on clinical suspicion of hypoglycemia, a low monitor-measured glucose level, or both need not be delayed until the result for the initial sample is reported. The potential detrimental effects of delayed treatment of hypoglycemia far outweigh any ill effect of unnecessary treatment. In addition, if the hypoglycemic mechanism is obscure, plasma samples for insulin, C peptide, sulfonylureas, and ethanol, at a minimum, should be obtained before glucose administration.

Oral treatment, with glucose tablets or glucose-containing fluids, candy, or food, is appropriate if the patient is able and willing to take these. A reasonable initial dose is 20 g of glucose (see Fig. 32-15). If the patient is unable or unwilling (because of neuroglycopenia) to take oral feedings, parenteral therapy is necessary. Intravenous glucose, 25 g initially, is preferable. If intravenous therapy is not practical, subcutaneous, intramuscular, or even intranasal glucagon can be used.

All of these urgent treatments raise plasma glucose concentrations only transiently (see Fig. 33-15). The plasma glucose concentration, as well as the patient’s clinical status, should be monitored after treatment. Intravenous glucose infusion is often necessary, and the patient should eat as soon as that is practicable.
Diagnosis of the Hypoglycemic Mechanism

In a patient with documented hypoglycemia, a plausible hypoglycemic mechanism (see Table 32-2) is usually apparent clinically from the history, physical examination, and available laboratory data. Iatrogenic hypoglycemia is reasonably assumed in the vast majority of instances in a patient treated with insulin, a sulfonylurea, or another insulin secretagogue or metformin for diabetes. In an adult who does not have diabetes, the use of a relevant drug, including alcohol; the presence of a relevant critical illness (hepatic, renal, or cardiac failure, sepsis, or inanition); clues to deficient secretion of cortisol, growth hormone, or both; or evidence of a nonbeta cell tumor leads to a presumptive mechanistic diagnosis and guides further diagnostic evaluation. Absent such clues, one must consider medication error, endogenous hyperinsulinism, or surreptitious or malicious sulfonylurea or insulin administration (see Fig. 32-16). Congenital metabolic defects are occasionally first recognized in an adult. The same differential diagnosis should be considered for children and even infants, although the hypoglycemic disorders unique to infancy and childhood, discussed earlier, must also be considered.
Prevention of Recurrent Hypoglycemia

Prevention of recurrent hypoglycemia in the long term requires treatment that corrects or circumvents the hypoglycemic mechanism. Offending drugs can be discontinued or their doses reduced. Underlying critical illnesses can often be treated. Cortisol (and growth hormone) can be replaced. Surgical, radiotherapeutic, or chemotherapeutic reduction of a nonbeta cell tumor can alleviate hypoglycemia even if the tumor cannot be cured; glucocorticoid (or growth hormone) administration may alleviate hypoglycemia in such patients. Surgical resection of an insulinoma is often curative; medical therapy with diazoxide, octreotide, or both can be used if that is not possible and in patients with a nontumor primary beta cell disorder. The treatment of autoimmune hypoglycemia (e.g., with a glucocorticoid) is more problematic, but the disorder is typically self-limited. Failing these treatments, provision of exogenous glucose with frequent feedings, large doses of uncooked cornstarch at bedtime, or even overnight intragastric glucose infusion may be necessary.
Acknowledgments

The author acknowledges the substantive contributions of several colleagues and collaborators that shaped the views expressed in this chapter; the assistance of the nursing, dietary, laboratory, informatics, and biostatistical staffs of the Washington University General Clinical Research Center and the laboratory staff of the Washington University Diabetes Research and Training Center in the performance of the original work cited; and the help of Ms. Karen Muehlhauser in the preparation of the manuscript.
References


37. Havel P, Ahren B. Activation of autonomic nerves and the adrenal medulla contributes to increased glucagon secretion during moderate insulin-induced hypoglycemia in women. Diabetes 1997;
61. Boyle PJ, Cryer PE. Growth hormone, cortisol, or both are involved in defense against, but are not critical to recovery from, prolonged hypoglycemia in humans. Am J Physiol 1991; 260:E395E402.


130. Paramore DS, Fanelli CG, Shah SD, Cryer PE. Hypoglycemia per se stimulates sympathetic neural as well as adrenomedullary activity but, unlike the adrenomedullary response, the forearm sympathetic neural response is not reduced following recent hypoglycemia. Diabetes 1999; 48:1429-1436.


References


1615


References


Obesity is a chronic disease that is causally related to serious medical illnesses. In the United States alone, the consequences of obesity account for an estimated 300,000 deaths per year. The medical expenses and cost of lost productivity related to obesity are greater than $100 billion per year. This chapter addresses the important clinical and pathophysiologic issues in obesity.
DEFINITION

Body Mass Index

Body mass index (BMI) is calculated by dividing weight (in kilograms) by height (in meters squared) or by dividing weight (in pounds) multiplied by 704 by height (in inches squared). There is a strong curvilinear relation between BMI and relative body fat mass. However, the current practical definition of obesity is based on the relationship between BMI and health outcome rather than BMI and body composition.

Table 33-1 summarizes the guidelines for classifying weight status by BMI, proposed by the major national and international health organizations. Large epidemiologic studies have established that there is a strong inverse relationship between BMI and mortality. Men and women with a BMI of 25.0 to 29.9 kg/m^2 are considered overweight, and those with a BMI 30 kg/m^2 or greater are considered obese. Obese persons have higher risk for adverse health consequences than those who are overweight (Fig. 33-1). These criteria for overweight and obesity represent imposed cutoff values along a continuum between mortality rate and BMI. The prevalence of obesity-related diseases, such as diabetes, begins to increase at BMI values below 25 kg/m^2 (Fig. 33-2).
Factors Affecting Body Mass Index-Related Risk

As shown in Table 33-1, several factors influence BMI-related health risk. For example, obese persons with excess abdominal fat are at higher risk for diabetes, hypertension, dyslipidemia, and ischemic heart disease than obese persons whose fat is located predominantly in the lower body. Waist circumference is highly correlated with abdominal fat mass and is therefore often used as a surrogate marker for abdominal (upper body) obesity. Waist circumference values denoting increased risk for metabolic diseases have been proposed on the basis of epidemiologic data. For men, a waist circumference greater than 102 cm (40 inches) and, for women, a waist circumference greater than 88 cm (35 inches) have been proposed as cutoff values for increased risk. However, this proposal imposes arbitrary cutoff values on the continuous relationship between waist circumference and metabolic disease risk.

Another factor that modifies the risk of obesity-related complications is weight gain during adulthood. In both men and women, weight gain of 5 kg or more since age 18 to 20 years is associated with an increased risk of diabetes, hypertension, and coronary heart disease, and the risk of disease increases with the amount of weight gained. However, this proposal imposes arbitrary cutoff values on the continuous relationship between waist circumference and metabolic disease risk.

Risks of developing obesity-associated diabetes or cardiovascular disease can also be modified by aerobic fitness. Blair and colleagues monitored more than 8000 men for an average of 6 years. Across a range of body adiposity, incidences of diabetes and cardiovascular mortality were lower in those who were fit, as defined by maximal ability to consume oxygen during exercise, than in those who were unfit.

BMI-associated health risk is also influenced by ethnicity. For example, when the subjects are matched on BMI, the risk

<table>
<thead>
<tr>
<th>Status</th>
<th>Obesity Class</th>
<th>Body Mass Index (kg/m²)</th>
<th>Risk of Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Underweight</td>
<td></td>
<td>≤18.5</td>
<td>Increased</td>
</tr>
<tr>
<td>Normal</td>
<td></td>
<td>18.524.9</td>
<td>Normal</td>
</tr>
<tr>
<td>overweight</td>
<td></td>
<td>25.029.9</td>
<td>Increased</td>
</tr>
<tr>
<td>Obesity I</td>
<td></td>
<td>30.034.9</td>
<td>High</td>
</tr>
<tr>
<td>II</td>
<td></td>
<td>35.039.9</td>
<td>Very high</td>
</tr>
<tr>
<td>III</td>
<td></td>
<td>40.0</td>
<td>Extremely high</td>
</tr>
</tbody>
</table>

Additional adiposity-related risk factors are (1) waist circumference >40 inches in men and >35 inches in women and (2) weight gain of 5 kg since age 18-20 years.


of diabetes is higher in Southeast Asian populations than in whites.
PATHOGENESIS

Energy Balance

Obesity is caused by an excessive intake of calories in relation to energy expenditure over a long period of time. The

Figure 33-1 Relationship between body mass index and cardiovascular mortality in adult men and women in the United States who never smoked and had no preexisting illness. The vertical line separates underweight and lean subjects (left side) from overweight and obese subjects (right side). (Adapted from Calle EE, Thun MJ, Petrelli JM, et al. Body-mass index and mortality in a prospective cohort of U.S. adults. N Engl J Med 1999; 341:1097.)

Figure 33-2 Relationship between body mass index and type 2 diabetes in adult men and women in the United States. The vertical line separates underweight and lean subjects (left side) from overweight and obese subjects (right side). The data demonstrate that the risk of diabetes begins to increase at the upper end of the lean body mass index category. (Adapted from Colditz GA, Willett WC, Rotnitzky A, Manson JE. Weight gain as a risk factor for clinical diabetes mellitus in women. Ann Intern Med 1995; 122:481-486; Chan JM, Rimm EB, Colditz GA, et al. Obesity, fat distribution, and weight gain as risk factors for clinical diabetes in men. Diabetes Care 1994; 17:981969.)

gastrointestinal tract has the capacity to absorb large amounts of nutrients. Large increases in body fat can result from even minor but chronic differences between energy intake and energy expenditure. In 1 year, the ingestion of only 5% more calories than expended can promote the gain of approximately 5 kg in adipose tissue. Over 30 years, the ingestion of only 8 kcal/day more than expended can increase body weight by 10 kg. This increase represents the average amount of weight gained by Americans during the 30-year period between ages 25 and 55 years. [20]
Genes and Environment

Body size depends on the complex interaction between genetic background and environmental factors. In humans, genetic background explains only an estimated 40% of the variance in body mass. Moreover, the marked increase in the prevalence of obesity over the last 20 years must have resulted largely from alterations in environmental factors that increase energy intake and reduce physical activity. For example, more meals are now eaten outside the home, there is greater availability of convenience and snack foods, serving sizes are larger, and daily physical activity has decreased because of sedentary lifestyle and work activities.

Environmental Effects in High-Risk Populations

Dramatic examples of the influence of environment on body weight have been reported globally and illustrate that persons of certain genetic backgrounds are especially likely to gain weight and experience obesity-related diseases when exposed to a modern lifestyle. Over the past 50 years, the striking change in the lifestyle of Pima Indians living in Arizona has led to an epidemic of obesity and diabetes. The diet of these urbanized Pimas is much higher in fat (50% of energy as fat) than their traditional diet (15% of energy as fat). In addition, these urbanized Pimas are much more sedentary than the Pimas who remain in the Sierra Madre mountains of northern Mexico and are isolated from Western influences. The rural Pimas eat a traditional diet, are physically active as farmers and sawmill workers, and have much lower incidences of obesity and diabetes than their Arizona kindred.

The aborigines of northern Australia are another high-risk population whose weight and health status has been compromised by exposure to a modern environment. Urbanized aborigines are heavier than their usually lean (BMI < 20 kg/m²) hunter-gatherer kindred and have high prevalences of type 2 diabetes and hypertriglyceridemia. The traditional hunter-gatherer lifestyle of the Aborigines involves a low-fat, low-calorie diet of wild game, fish, and plants and a high level of physical activity. Short-term (7 weeks) reexposure to the traditional lifestyle resulted in weight loss and significant improvement or normalization of glucose tolerance and fasting blood glucose, insulin, and triglyceride concentrations in urbanized Aborigines with type 2 diabetes and hypertriglyceridemia.

Influences of Childhood and Parental Obesity

The risk of becoming an obese adult is influenced both by having been obese as a child and by having one or more obese parents. The risk of adult obesity increases with increasing age and severity of obesity in childhood. For example, the risk of being obese at 21 to 29 years of age ranged from 8% for persons who were obese at 1 to 2 years of age and had nonobese parents to 75% for persons who were obese at 10 to 14 years of age and had at least one obese parent. Whereas persons who were obese at 1 to 2 years of age and had lean parents did not have an increased risk of obesity in adulthood, persons who became obese after 6 years of age had more than a 50% chance of becoming obese adults.
Monogenic Causes of Obesity

Since the discovery of the adipose tissue protein leptin, much progress has been made in understanding the molecular basis of body fat regulation. It has nonetheless been disappointing that a genetic cause of obesity has been identified in only a few individuals. Several rare, monogenic causes of obesity have been described.

Leptin Gene Mutations

The pathophysiologic relevance of leptin was established in two extremely obese cousins with hyperphagia who belonged to a consanguineous family of Pakistani origin.[26] These cousins were homozygous for a single nucleotide deletion at position 398 of the leptin gene. This mutation resulted in a frameshift of the leptin-coding region and a premature termination of leptin synthesis. The parents of the cousins were heterozygous for this mutation. Another mutation, this time involving a homozygous single nucleotide transversion in the leptin gene that resulted in an Arg Trp substitution in the mature peptide and low serum leptin levels, was discovered in three extremely obese individuals. Two of these persons are adults. Both the adult man and woman were hyperinsulinemic.[27] The man exhibited hypothalamic hypogonadism and dysfunction of the sympathetic nervous system, and the woman had primary amenorrhea.

Leptin treatment has successfully reversed the obesity of leptin-deficient patients. Treatment with recombinant human leptin resulted in weight loss of 1 to 2 kg per month over a 12-month period. Loss of fat mass accounted for 95% of this weight loss.[28]

The possibility that leptin levels are reduced in obesity has been investigated in large groups of subjects. However, serum leptin levels increase exponentially with fat mass, suggesting that most obese persons are resistant or insensitive to body weight regulation by endogenous leptin.[29]

Leptin Receptor Mutation

Three extremely obese sisters from a consanguineous family were found to have markedly high serum leptin levels and were homozygous for a single nucleotide substitution at the splice site of exon 16 of the leptin receptor gene.[30] This mutation resulted in a truncated protein that lacked both the transmembrane and intracellular domains of the receptor. The sisters displayed hypogonadotropic hypogonadism, failure of pubertal development, growth delay, and secondary hypothyroidism. This finding confirms the endocrine abnormalities in leptin-deficient subjects and implies a role for the leptin/leptin receptor system in the central regulation of energy balance and hypothalamic endocrine functions in humans.

Prohormone Convertase 1 Gene Mutation

A mutation in the prohormone convertase 1 (PC1) gene was identified in a 43-year-old obese woman with a history of severe childhood obesity.[31] The woman had impaired glucose tolerance, postprandial hypoglycemia, low plasma cortisol levels, and hypogonadotropic hypogonadism. In addition, she had increased plasma concentrations of proinsulin and pro-opiomelanocortin (POMC) but low plasma insulin concentrations. She was a compound heterozygote for two mutations in the PC1 gene, which resulted in loss of the autocatalytic cleavage ability of PC1. Melanocortins, including -melanocortin-stimulating hormone, are formed through the processing of POMC by PC1. Therefore, reduced production of melanocortin may have been responsible for obesity in this patient.

Pro-opiomelanocortin Gene Mutation

A mutation in the POMC gene was described in two obese children with hyperphagia.[32] The children also had red hair pigmentation and were deficient in adrenocorticotropic hormone. The mutations resulted in complete loss of the ability to synthesize -melanocortin-stimulating hormone and adrenocorticotropic hormone. The red hair pigmentation and obesity are thought to be due to deficiency of -melanocortin-stimulating hormone.

Melanocortin 4 Receptor Mutation

Although rare, melanocortin 4 receptor mutations are the most common monogenic cause of obesity.[33] Moreover, melanocortin 4 receptor mutations are characterized by both dominant and recessive modes of inheritance, in contrast to the other monogenic forms of obesity, which involve recessive modes of inheritance. Persons with melanocortin 4 receptor mutations have no apparent phenotypic abnormalities other than obesity.

SIM1 Gene Mutation

A de novo balanced translocation between chromosomes 1 and 6 was found in a severely obese girl who weighed 47 kg at 67 months of age.[34] The mutation caused a disruption in the SIM1 gene, which encodes a transcription factor involved in the formation of the paraventricular and supraoptic nuclei. It is likely that this abnormality altered energy balance by stimulating food intake because measured resting energy expenditure was normal.
Polygenic Causes of Obesity

In contrast to the small number of single-gene mutations that clearly cause obesity, a large number of human genes have been identified that show variations in deoxyribonucleic acid (DNA) sequences and may also contribute to obesity. In large population surveys, more than 250 genes, markers, and chromosomal regions have been linked to human obesity. Some of these associations will undoubtedly prove to be more important than others. It remains a major challenge to identify genes that contribute to human obesity because of the potential interactions between multiple genes and the interaction between genes and environment that can lead to expression of an obesity phenotype.
ENERGY METABOLISM

The components of daily total energy expenditure (TEE) are:

1. Resting energy expenditure (REE), accounting for approximately 70% of TEE.
2. Energy expended in physical activity, accounting for approximately 20% of TEE.
3. The thermic effect of food (TEF), accounting for approximately 10% of TEE.

REE represents the energy expended for normal cellular and organ function under postabsorptive resting conditions. Energy expended in physical activity includes the energy costs of both volitional activity, such as exercise, and nonvolitional activity, such as spontaneous muscle contractions, maintaining posture, and fidgeting. The thermic effect of food represents the energy expended in digestion, absorption, and sympathetic nervous system activation after ingestion of a meal.

Cross-sectional studies have investigated whether alterations in energy metabolism are involved in obesity. Obese individuals usually have greater rates of REE than lean individuals of the same height because obese individuals have a greater amount of lean and adipose tissue cell mass. Defects in REE or TEE have not been detected in diet-resistant patients who maintain their weight despite the claim of strict adherence to a low-calorie diet. Instead, such patients appear to underestimate their food intake and actually consume twice as many calories as they record in food intake diaries.

It is not known whether obese individuals expend less total energy in daily physical activity because they are less active than lean individuals. During nonweight-bearing activity (e.g., cycling), obese individuals expend the same amount of energy as lean individuals to perform the same amount of work. During weight-bearing activities, however, obese individuals expend more energy than lean individuals because more work is required to carry their greater body weight. Evidence from studies of obese and lean subjects, matched for either fat mass or lean body mass, suggests that obese subjects have a small (75 kcal/day) but potentially important reduction in the thermic effect of food. This reduction in the thermic effect of food may arise from the insulin resistance and blunted sympathetic nervous system activity that occur in obesity.

Although extensive research has not revealed significant defects in the energy metabolism of individuals who are already obese, the possibility remains that inherent abnormalities in energy metabolism contribute to the development of obesity. However, today's research technology has only limited ability to detect small, but possibly clinically significant, chronic defects in energy metabolism. Moreover, it is difficult to establish a causal relationship between energy expenditure and the development of obesity because energy metabolism measurements capture only a brief point in time and may not reveal abnormalities that emerge during specific life stages.

Most studies do not support the involvement of a defect in metabolic rate in the development of obesity. In one longitudinal study, daily TEE at 3 months of age was 21% lower in infants who later became overweight than in those who maintained a normal weight. However, larger subsequent studies have not confirmed this finding. In a longitudinal study of 126 Pima Indians, those in the lowest tertile of REE at baseline had the highest cumulative incidence of a 10-kg weight gain 1 to 4 years later. In contrast, the Baltimore Longitudinal Study on Aging, which monitored 775 men for an average of 10 years, did not detect a relationship between initial REE and weight change.

When energy intake exceeds energy expenditure, weight gain usually occurs. However, genetic factors may influence the amount of weight gained with overfeeding. Bouchard and colleagues observed variable weight gain among 12 monozygotic twin pairs who were chronically overfed 1000 kcal/day. However, the members of each twin pair gained similar amounts of weight. In another study, the increase in body fat after 8 weeks of overfeeding was inversely related to changes in nonvolitional energy expenditure (e.g., fidgeting). Therefore, in some individuals, nonvolitional energy expenditure during periods of overingestion could be a mechanism that limits weight gain through the dissipation of excess ingested energy.

Diet-induced weight loss decreases REE, which promotes weight regain. This observation underlies the set-point theory, which posits that body weight is predetermined so that weight loss (or gain) promotes a decrease (or increase) in metabolic rate that acts to restore body weight to a preset level. In both lean and obese persons, hypocaloric feeding reduces REE by 15% to 30%. This reduction in REE cannot be completely accounted for by the accompanying decrease in body size or lean body mass and is considered a normal part of the physiologic adaptation to energy restriction. The reduction in REE that occurs with negative energy balance is transient and does not persist during maintenance of a lower body weight.

As reported in several studies, long-term maintenance of weight loss is not associated with an abnormal decrease in REE or TEE when adjustments are made for changes in body composition. In a meta-analysis of 15 studies, the REE of subjects who were formerly obese was found to be similar to that of subjects who were never obese. Although the decrease in energy metabolism with weight loss is largely appropriate for the concomitant changes in body composition, this decrease may nonetheless promote weight regain.
ADIPOSE TISSUE AND TRIGLYCERIDE METABOLISM

Triglycerides stored within adipose tissue constitute the body's major energy reserve. (Table 33-2) Triglycerides are a much more compact fuel than glycogen because of their energy density and hydrophobic nature. Triglycerides yield 9.3 kcal/g upon oxidation and are compactly stored as oil inside the fat cell, accounting for 85% of adipocyte weight. Glycogen, in contrast, yields only 4.1 kcal/g upon oxidation and is stored intracellularly as a gel containing approximately 2 g of water for every gram of glycogen.

Adipose tissue is an effective storage mechanism for transportable fuel that allows mobility and survival when food is scarce. During starvation, the duration of survival is determined by the size of the adipose tissue mass. Lean persons die after only approximately 60 days of starvation when more than 35% of body weight is lost. In contrast, obese persons, have tolerated therapeutic fasts for more than a year without adverse effects. In the longest reported fast, a 207-kg man ingested only noncaloric fluids, vitamins, and minerals for 382 days and lost 126 kg, or 61% of his initial weight.

Triglyceride Storage

The major function of adipocytes is the storage of triglycerides for future use as energy substrate. Lipogenesis from glucose makes only a limited contribution to triglyceride storage in the adipocyte. Most of the triglyceride in adipocytes is derived from chylomicrons and very-low-density lipoprotein (VLDL) triglycerides that originate, respectively, from dietary and hepatic sources. These plasma triglycerides are hydrolyzed by lipoprotein lipase (LPL), a key regulator of fat cell triglyceride uptake from circulating triglycerides. LPL is synthesized by adipocytes and transported to the endoluminal surface of endothelial cells. The interactions of LPL with chylomicrons and VLDL release fatty acids from plasma triglycerides, which are then taken up by local adipocytes. Plasma free fatty acids themselves can also be taken up by adipose tissue, independent of lipoprotein lipolysis.

Insulin and cortisol are the principal hormones involved in regulation of LPL activity and expression. The activity of LPL within individual tissues is a key factor in partitioning triglycerides among different body tissues. Insulin influences this partitioning through its stimulation of LPL activity in adipose tissue. Insulin also promotes triglyceride storage in adipocytes through other mechanisms, including inhibition of lipolysis, stimulation of adipocyte differentiation, and increasing glucose uptake. The importance of cortisol in fat distribution is supported by the clinical appearance of patients with Cushing’s syndrome. The obesity-promoting effect of cortisol may involve a synergistic effect of cortisol and insulin on the induction of LPL in adipose tissue, which has been demonstrated in vitro. Testosterone, growth hormone, catecholamines, and tumor necrosis factor (TNF) and other related cytokines inhibit LPL activity.
Lipolysis

The balance between triglyceride storage and lipolysis is regulated by complex hormonal and neuronal mechanisms. To become available as an energy substrate, triglycerides stored within adipocytes must be hydrolyzed by hormone-sensitive lipase into fatty acids. These fatty acids can be released from adipocytes into the circulation.

The circulating half-life of plasma fatty acids is only 3 to 4 minutes. The excess availability of fatty acids in plasma provides a ready supply of oxidizable substrate to respond to sudden changes in energy requirements, such as are induced by exercise. These fatty acids are the major precursors of hepatic VLDL triglyceride synthesis. In turn, VLDL triglycerides are secreted by the liver and redistributed throughout the body, depending on tissue-specific factors such as the activity of LPL. These observations imply that there is continuous redistribution of triglycerides between adipose tissue and the rest of the body.

There is considerable variation in the rate of lipolysis and, consequently, plasma fatty acid level both within and between subjects. Insulin and catecholamines are the major circulating hormones that influence lipolysis in adipocytes. Insulin inhibits lipolysis through its effect on hormone-sensitive lipase, whereas catecholamines stimulate lipolysis. Small changes in plasma concentrations of insulin and catecholamines may have major effects on lipolytic rate. Half-maximal suppression of lipolysis occurs at postabsorptive insulin levels, and maximal suppression of lipolysis occurs at insulin levels within the range observed after a regular meal. Only minor increases in resting catecholamine levels stimulate lipolysis. Other factors modulate the rate of lipolysis. For example, growth hormone and cortisol stimulate lipolysis. In general, the effects of these other factors are less potent than the effects of insulin and catecholamines.

In contrast to the tight feedback regulation of insulin secretion by glucose levels, insulin and catecholamine concentrations are not regulated by lipolysis or fatty acid levels. Although free fatty acid levels affect glucose-stimulated insulin release, there is no feedback between insulin release and rate of lipolysis. The wide physiologic variation in plasma free fatty acid concentrations between individuals can be explained, in part, by the finely tuned dose-response effects of insulin and catecholamines on lipolysis in combination with the absence of tight feedback regulation of insulin and catecholamine levels by free fatty acids.

Basal plasma fatty acid concentrations are often increased in obese persons, particularly those with abdominal (upper body) obesity. An increased rate of free fatty acid release into plasma because of an increased rate of lipolysis from upper body subcutaneous fat is responsible for the higher levels of circulating fatty acids. The excess free fatty acid availability in plasma may lead to increased hepatic free fatty acid uptake, VLDL triglyceride synthesis, intramuscular triglyceride formation, and insulin resistance.
ADIPOSE TISSUE AS AN ENDOCRINE ORGAN

Traditionally, adipocytes have been viewed as energy depots that store triglycerides during feeding and release fatty acids during fasting to provide fuel for other tissues. However, it has become evident that adipose tissue has major integrative physiologic functions and secretes numerous proteins. In part, these factors participate in autocrine and paracrine regulation within adipose tissue. In addition, these factors have profound effects on the function of distant organs, such as muscle, pancreas, liver, and brain. The realization that adipose tissue functions as an endocrine organ has important implications for understanding the pathophysiologic relationship between excess body fat and pathologic states, such as insulin resistance and type 2 diabetes mellitus.

### TABLE 33-3 -- Adipocyte-Secreted Proteins

<table>
<thead>
<tr>
<th>Category</th>
<th>Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>Potential hormone</td>
<td>Leptin, resistin, angiotensinogen, Adiponectin/ACRP 30, estrogens</td>
</tr>
<tr>
<td>Cytokine</td>
<td>Interleukin-6, tumor necrosis factor</td>
</tr>
<tr>
<td>Extracellular matrix protein</td>
<td>Type I, III, IV, and VI collagen, fibronectin, osteonectin, laminin, entactin, matrix metalloproteinases 2 and 9</td>
</tr>
<tr>
<td>Complement factor</td>
<td>Adipsin, complement C3, factor B</td>
</tr>
<tr>
<td>Enzyme</td>
<td>Cholesterol ester transfer protein, lipoprotein lipase</td>
</tr>
<tr>
<td>Acute phase response proteins</td>
<td>α-Acid glycoprotein, haptoglobin</td>
</tr>
<tr>
<td>Other</td>
<td>Fatty acids</td>
</tr>
<tr>
<td></td>
<td>Plasminogen activator inhibitor-1</td>
</tr>
<tr>
<td></td>
<td>Prostacyclin</td>
</tr>
</tbody>
</table>

Leptin

Adipocytes produce leptin and secrete it into the blood stream. Leptin has pleiotropic effects on food intake, hypothalamic neuroendocrine regulation, reproductive function, and energy expenditure. There is a direct relationship between plasma leptin concentrations and BMI or percent body fat. However, there can be considerable variability in leptin concentrations among persons with the same BMI, suggesting that leptin production is also regulated by factors other than adipose tissue mass per se.

Leptin levels decrease rapidly within 12 hours after the start of starvation; conversely, they increase in response to overfeeding. Therefore, plasma leptin concentrations reflect adipose tissue mass and are influenced by energy balance. In this perspective, leptin is a bidirectional signal that switches physiologic regulation between fed and starved states. Plasma leptin concentrations increase with increasing fat mass and decrease rapidly during early fasting. At present, the relative importance of the central versus peripheral effects of leptin remains to be elucidated.
Resistin

Resistin is another signaling polypeptide secreted by adipocytes. Resistin levels are increased in mice with diet-induced and genetic forms of obesity and insulin resistance. Administration of recombinant resistin to normal mice impaired glucose tolerance and insulin action. Neutralization of resistin reduced hyperglycemia in obese, insulin-resistant mice, in part by improving insulin sensitivity. It has therefore been proposed that resistin is a hormone that links obesity to diabetes by inducing insulin resistance.
Estrogens

Adipose tissue has aromatase activity. This enzyme is important for transforming androstenedione into estrone. Estrone is the second major circulating estrogen in premenopausal women and the most important estrogen in postmenopausal women. The rate of conversion of androstenedione to estrone increases with age and obesity and is higher in lower body than in upper body obesity. In addition to a role in endocrine regulation, the effects of P450 aromatase on estrogen metabolism may have a role in autocrine or paracrine action because estrogen receptors are present in adipose tissue.
Tumor Necrosis Factor

Adipocytes secrete TNF-α, and TNF-α expression is increased in the enlarged adipocytes of obese subjects. However, plasma TNF-α levels are generally at or below the detection limit of available assays, which suggests that the TNF-α produced within adipose tissue has paracrine rather than endocrine functions. The multiple effects of TNF-α on adipocytes include impairment of insulin signaling. Therefore, it has been proposed that TNF-α may partially contribute to insulin resistance in obesity.\ref{23}
ADIPOCYTE BIOLOGY

Obesity is associated with an increased number of adipocytes. A lean adult has about 35 billion adipocytes, each containing about 0.4 to 0.6 µg of triglyceride; an extremely obese adult can have four times as many adipocytes (125 billion), each containing twice as much lipid (0.8 to 1.2 µg of triglyceride).

Our understanding of adipocyte differentiation is largely derived from studies of preadipocytes in culture. The current concept is that adipocytes are derived from fibroblast precursor cells after the concerted actions of extracellular signals and intrinsic transcription factors and coactivators. Many extranuclear factors and intracellular transduction pathways influence the adipogenic potential of cells in vitro and in vivo.

Although in the future it may be possible to regulate adipogenesis in vivo, decreasing adipogenesis without altering energy balance may result in the deposition of triglycerides in other tissues. Excessive triglycerides in nonadipose tissues can have deleterious effects, as suggested by the liver steatosis, dyslipidemia, and diabetes observed when adipogenesis was prevented in mice.

The cornerstone of obesity therapy is to increase the utilization of endogenous fat stores as fuel by reducing energy intake below energy expenditure. With dieting, weight loss is composed of approximately 75% to 85% fat and 15% to 25% fat-free mass (FFM). An energy deficit of approximately 3500 kcal is required to oxidize 1 pound of adipose tissue. However, because of the oxidation of lean tissue and associated water losses, a 3500-kcal energy deficit reduces body weight by more than 1 pound. The distribution of fat loss is characterized by regional heterogeneity. Particularly in men and women with initially increased intra-abdominal fat, there are greater relative losses of intra-abdominal fat than of total body fat mass. A decrease in the size (triglyceride content) of existing adipocytes accounts for most, if not all, of the fat loss. In humans, there is also evidence that the number of adipocytes is reduced with large, long-term fat loss. However, it is possible that this perception of decreased fat cell number is false because standard cell-counting techniques may fail to detect adipocytes that have undergone marked shrinkage.

There are two possible mechanisms through which weight loss can eliminate fat cells: (1) dedifferentiation, the morphologic and biochemical reversion of mature adipocytes to preadipocytes, and (2) apoptosis. Adipocyte dedifferentiation has been observed in vitro, but there is no evidence that it occurs in vivo. Adipocyte apoptosis has been induced in vitro and has been shown to occur in vivo in some patients with cancer. To date, it is not known whether diet-mediated weight loss induces adipocyte apoptosis.
PREVALENCE OF OBESITY

The worldwide prevalence of obesity has increased dramatically over the last several decades. In the United States alone, an estimated 61% (110 million) of adults 20 to 74 years of age are now considered overweight or obese. According to national population surveys conducted since 1960, the prevalence of overweight in the United States increased only slightly from 30.5% to 34.0% but the prevalence of obesity (BMI >30 kg/m²) more than doubled, from 12.8% to 27%. In the United States, the prevalence of obesity increases progressively from 20 to 50 years of age but then declines after 60 to 70 years of age.

The prevalence of obesity has also risen in children and adolescents. As defined by a BMI greater than the 95th percentile for age and gender from the revised National Center for Health Statistics growth charts, 10% to 15% of 6- to 17-year-old children and adolescents in the United States are overweight. These data indicate that overweight prevalence rates for children and adolescents, reported by earlier surveys, have doubled. Diseases commonly associated with obesity in adults, such as type 2 diabetes mellitus, hypertension, hyperlipidemia, gallbladder disease, nonalcoholic steatohepatitis, sleep apnea, and orthopedic complications, are now increasingly observed in children.
CLINICAL FEATURES AND COMPLICATIONS OF OBESITY

Obesity is strongly associated with many serious medical complications that impair quality of life and lead to increased morbidity and premature death (Fig. 33-4).

The complications associated with obesity have been reviewed in detail previously. 9

Endocrine and Metabolic Disease

The Metabolic or Insulin Resistance Syndrome

In the metabolic or insulin resistance syndrome, also known as syndrome X, the specific phenotype of upper body or abdominal obesity is associated with a cluster of metabolic risk factors for coronary heart disease (CHD). Features of this syndrome include:

1. Insulin resistance, including hyperinsulinemia, impaired glucose tolerance, impaired insulin-mediated glucose disposal, and type 2 diabetes mellitus.
2. Dyslipidemia, characterized by hypertriglyceridemia and low serum high-density lipoprotein (HDL) cholesterol levels.
3. Hypertension.

Abdominal obesity has also associated with the metabolic risk factors of (1) increased serum levels of apolipoprotein B; (2) small, dense low-density lipoprotein (LDL) particles; and (3) plasminogen activator inhibitor 1 with impaired fibrinolysis. 14 15 16 The metabolic syndrome does not affect only those with frank obesity; it has also been reported in persons of normal weight, who presumably have an increased amount of abdominal fat. 17

The metabolic syndrome was originally identified and defined on the basis of epidemiologic associations. The underlying pathogenesis and the interrelationships between the individual features have not been completely elucidated. Insulin resistance has been hypothesized to be the common underlying pathogenic mechanism. 18 However, according to a factor analysis of data from nondiabetic subjects in the Framingham Offspring Study, insulin resistance may not be the only precedent condition and more than one independent physiologic process may be involved. 19 Abdominal obesity is clearly associated with insulin resistance. It is not clear whether the visceral (omental and mesenteric) or subcutaneous depots of abdominal fat are more closely related to insulin resistance because data from different studies are contradictory. In addition, it is difficult to define the relationship between each abdominal adipose tissue depot and insulin resistance because the size of the depots is closely correlated. Furthermore, it is not known whether visceral fat and abdominal fat actually participate in the pathogenesis of the metabolic syndrome or merely serve as markers of increased risk for the metabolic complications of obesity. 20

There is increasing evidence that the ectopic distribution of triglycerides in nonadipose tissue may be involved in the complications of obesity. In cross-sectional studies, insulin resistance was highly correlated with the intramyocellular concentration of triglyceride. 21 It is not known whether triglycerides per se interfere with insulin action or whether triglycerides serve as a surrogate marker for some other fatty acid-derived entity from plasma or intracellular sources that impairs insulin signaling. 22

Type 2 Diabetes Mellitus

The marked increase in the prevalence of obesity has played an important role in the 25% increase in the prevalence of diabetes that has occurred in the United States. 23 According to data from the Third National Health and Nutrition Examination Survey (NHANES III), two thirds of the adult men and women in the United States with a diagnosis of type 2 diabetes had a BMI of 27 kg/m² or greater. 24 The risk of diabetes increases linearly with BMI; the prevalence of diabetes increased from 2% in those with BMI 25 to 29.9 kg/m², to 6% in those with 30 to 34.9 kg/m², and finally to 13% in those with BMI greater than 35 kg/m². 25 In the Nurses Health Study, the risk of diabetes began to increase when BMI exceeded the "normal" value of 22 kg/m² (see Fig. 33-2). 26 27 In addition, the risk of diabetes increases with increments in abdominal fat mass, waist circumference, or ratio of waist to hip circumference at any given BMI value. 28 29 30

The risk of diabetes also increases with weight gain during adulthood. Among men and women 35 to 60 years of age, the risk of diabetes was three times greater in those who gained 5 to 10 kg after the age of 18 to 20 years than in those who maintained their weight within 2 kg. 31 32

Dyslipidemia

Obesity is associated with several serum lipid abnormalities, including hypertriglyceridemia, reduced HDL cholesterol levels, and an increased proportion of small, dense LDL particles. 33 34 35 This association is especially strong in persons with abdominal obesity. In addition, most studies suggest that serum concentrations of total cholesterol and LDL cholesterol are elevated in obesity. 36 Data from NHANES III showed that in men there was a progressive increase in the prevalence of hypercholesterolemia (total blood cholesterol >240 mg/dL or 6.21 mmol/L) with increasing BMI. 37 In women, by contrast, the prevalence of hypercholesterolemia was highest at BMI of 25 to 27 kg/m² and did not increase further at higher BMI values. The serum lipid abnormalities associated with obesity are important risk factors for CHD. 38 39 40
Cardiovascular Disease

Hypertension

There is a linear relationship between hypertension and BMI. In NHANES III, the age-adjusted prevalence of hypertension (defined as systolic blood pressure > 140 mm Hg, diastolic blood pressure > 90 mm Hg, or the need for antihypertensive medication) in obese men and women was 42% and 38%, respectively. These prevalence rates are more than twice as high as the prevalence rates of hypertension in lean men and women (15% prevalence rate in both men and women).

The risk of hypertension also increases with weight gain. Among subjects in the Framingham Study, there was an increase of 6.5 mm Hg in blood pressure with every 10% increase in body weight.

Coronary Heart Disease

The risk of CHD is increased in obese persons, particularly those with increased abdominal fat distribution and those who gained weight during young adulthood. Moreover, CHD risk starts to increase at the normal BMI levels of 23 kg/m$^2$ in men and 22 kg/m$^2$ in women. In the Nurses Health Study, the risk of fatal and nonfatal myocardial infarctions was greater in women with the lowest BMI but highest ratio of waist to hip circumference than in women with the highest BMI but lowest ratio of waist to hip circumference. At any BMI level, the risk of CHD increases with the presence of increased abdominal fat. The risk of fatal and nonfatal myocardial infarction also increases when 5 kg or more is gained after 18 years of age.

Obesity-related CHD risk factors, particularly hypertension, dyslipidemia, impaired glucose tolerance, and diabetes, are largely responsible for the increase in CHD. However, even after adjustments for other known risk factors, several long-term epidemiologic studies found that overweight and obesity increased the risk of CHD. As a result, the American Heart Association classified obesity as a major preventable risk factor for CHD.

Cerebrovascular and Thromboembolic Disease

The risk of fatal and nonfatal ischemic stroke is approximately twice as great in obese as in lean persons and increases progressively with increasing BMI. The risks of venous stasis, deep vein thrombosis, and pulmonary embolism are also increased in obesity, particularly in persons with abdominal obesity. Lower extremity venous disease may result from increased intra-abdominal pressure, impaired fibrinolysis, and the increase in inflammatory mediators.
Pulmonary Disease

Restrictive Lung Disease

Obesity increases the pressure placed on the chest wall and thoracic cage, which restricts pulmonary function by decreasing respiratory compliance, increasing the work of breathing, restricting ventilation (measured as decreased total lung capacity, forced vital capacity, and maximal ventilatory ventilation), and limiting ventilation of lung bases.\[116\]

Obesity-Hypoventilation Syndrome

In obesity-hypoventilation syndrome, the partial pressure of carbon dioxide (P$_{\text{CO}_2}$) is less than 50 mm Hg because of decreased ventilatory responsiveness to hypercapnia or hypoxia, or both, and inability of respiratory muscles to meet the increased ventilatory demand imposed by the mechanical effects of obesity. Alveolar ventilation is reduced because of shallow and inefficient ventilation related to decreased tidal volume, inadequate inspiratory strength, and elevation of the diaphragm. Symptoms increase when patients are lying down because of increased abdominal pressure on the diaphragm. The resulting increase in intrathoracic pressure further compromises lung function and respiratory capacity.

The pickwickian syndrome is a severe form of the obesity-hypoventilation syndrome. Named after an obese character in Charles Dickens' *Pickwick Papers*, this syndrome involves extreme obesity, irregular breathing, somnolence, cyanosis, secondary polycythemia, and right ventricular dysfunction.

Obstructive Sleep Apnea

In patients with obstructive sleep apnea (cessation of breathing), excessive episodes of apnea and hypopnea during sleep are caused by partial or complete upper airway obstruction despite persistent respiratory efforts. Daytime sleepiness and cardiopulmonary dysfunction result from the interruption in nighttime sleep and arterial hypoxemia. In general, patients with sleep apnea are characterized by a BMI greater than 30 kg/m$^2$, abdominal fat distribution, and a large neck girth (>17 inches in men and >16 inches in women).\[117\], \[118\], \[119\]
Musculoskeletal Disease

Gout

Both hyperuricemia and gout are associated with obesity.\textsuperscript{120, 121}

Osteoarthritis

The risk of osteoarthritis of weight-bearing joints is increased in overweight and obese persons.\textsuperscript{122} The knees are most often involved because much more body weight is exerted across the knees than across the hips during weight-bearing activity.\textsuperscript{122} There is a stronger relationship between body size and osteoarthritis in women than in men, and even small increases in body weight in women can promote osteoarthritis. In a study of twins, symptomatic or asymptomatic lower extremity osteoarthritis was found in individuals who were only 3 to 5 kg heavier than their twin sibling.\textsuperscript{123}
Cancer

Overweight and obesity increase the risk of cancer of the esophagus, gallbladder, colon, breast, uterus, cervix, and prostate gland. Most but not all epidemiologic studies have found a direct relationship between BMI and colon cancer in both men and women. The risks of mortality from breast and endometrial cancer increase with obesity and weight gain after age 18 years. The risk of breast cancer appears to increase with increasing BMI only in postmenopausal women, in premenopausal women, increased BMI may actually protect against breast cancer. Obesity is often correlated with ingestion of a high-fat, high-calorie diet, which is another risk factor for cancer. Therefore, it is difficult to distinguish how much of the relation between obesity and cancer is attributable to obesity per se and how much is attributable to dietary factors.
Genitourinary Disease in Women

Obese women are often affected by irregular menses, amenorrhea, and infertility. Pregnant obese women are at increased risk for gestational diabetes and hypertension, delivery complications, and congenital malformations. The risk of urinary incontinence is also increased in obese women, even after adjustments are made for age and parity. In extremely obese patients, incontinence usually resolves after considerable weight loss, usually achieved by bariatric surgery.
Neurologic Disease

As described earlier, the incidence of ischemic stroke is increased in obesity. Obesity is also associated with idiopathic intracranial hypertension (IIH), also known as pseudotumor cerebri. This syndrome is manifested by headache, vision abnormalities, tinnitus, and sixth nerve paresis. Although the prevalence of IIH increases with increasing BMI, the risk of IIH is increased even in persons who are only 10% above ideal body weight. The observation that weight loss in extremely obese patients with IIH decreases intracranial pressure and resolves most associated clinical signs and symptoms suggests that there is a causal relationship between obesity and IIH.
Cataracts

Overweight and obesity are associated with an increased prevalence of cataracts. Moreover, persons with abdominal obesity are at greater risk than those with lower body obesity, suggesting that insulin resistance may be involved in the pathogenesis of cataract formation.
Gastrointestinal Disease

Gastroesophageal Reflux Disease

The relationship between gastroesophageal reflux disease and obesity is unclear because of conflicting data from different studies. A higher incidence of reflux symptoms in obese than in lean persons has been found in most but not all large epidemiologic studies. In addition, studies that evaluated gastroesophageal acid reflux by 24-hour pH monitoring have reported the presence of a significant relationship and no relationship between BMI and pathologic reflux, defined as the occurrence of esophageal pH less than 4 more than 5% of the time.

Gallstones

The risk of symptomatic gallstones increases linearly with BMI. The Nurses Health Study found that the annual incidence of symptomatic gallstones was 1% in women with a BMI greater than 30 kg/m² and 2% in women with a BMI greater than 45 kg/m². The risk of gallstones increases during weight loss, particularly when weight loss is rapid. This increased risk is related to increased bile cholesterol supersaturation, enhanced cholesterol crystal nucleation, and decreased gallbladder contractility. When the rate of weight loss exceeded 1.5 kg (1.5% of body weight) per week, the risk of gallstone formation increased exponentially.

In obese patients who underwent rapid weight loss with a very-low-calorie diet (<600 kcal/day), a low-fat diet (1 to 3 g/day), or gastric surgery, the respective incidence of new gallstones was approximately 25%, 33%, and 38%. Gallstone formation was also promoted by the low fat content of very-low-calorie diets because more than 4 to 10 g of fat in a meal was needed to stimulate maximal gallbladder contractility. Therefore, increasing the fat content of a very-low-calorie diet can prevent the development of new gallstones. However, increasing dietary fat content may not be as important in preventing gallstones in patients consuming a low-calorie diet as in those consuming a very-low-calorie diet. Administration of ursodeoxycholic acid (600 mg/day) during weight loss markedly decreased gallstone formation.

Pancreatitis

Although obese patients would be expected to be at increased risk for gallstone pancreatitis because of their increased prevalence of gallstones, few studies have addressed this issue. Several studies have shown that overweight and obese patients with pancreatitis have a higher risk of local complications, severe pancreatitis, and death than lean patients. It has been hypothesized that the increased deposition of fat in the peripancreatic and retroperitoneal spaces predisposes obese patients to the development of peripancreatic fat necrosis and subsequent local and systemic complications.

Liver Disease

Obesity is associated with a spectrum of liver abnormalities, which are now referred to as nonalcoholic fatty liver disease (NAFLD). These abnormalities include hepatomegaly, abnormal liver biochemistry, steatosis, steatohepatitis, fibrosis, and cirrhosis. Alanine aminotransferase and aspartate aminotransferase are the most commonly elevated liver enzymes, but elevations of these enzymes generally do not exceed two times the upper limit of normal. Moreover, enzyme levels often do not correlate with the severity of histologic abnormalities. Most of the available data suggest that steatosis affects approximately 75%, steatohepatitis approximately 20%, and cirrhosis approximately 2% of obese patients.

The presence of NAFLD is associated with the presence of abdominal obesity and the metabolic syndrome. However, the factors underlying the development of NAFLD in obese persons are not clear. NAFLD has been hypothesized to result from two or more insults to the liver. The first of these events is steatosis, caused by obesity-induced alterations in lipid metabolism: (1) increased lipolysis of adipose tissue triglycerides, which increases the delivery of free fatty acids to the liver; (2) increased de novo lipogenesis; and (3) possibly inadequate hepatic fatty acid oxidation. The second insult may involve peroxidation of hepatic lipids and injury-related cytokines, which can promote direct cellular injury, inflammation, and fibrosis.

Obese patients with NAFLD are usually advised to lose weight, but it is not known whether weight loss alters the progression of the disease. With a gradual weight loss of 10% or more, abnormalities in liver chemistry resolve and liver size, hepatic fat content, and features of steatohepatitis decrease.
BENEFITS OF INTENTIONAL WEIGHT LOSS

Effect on Mortality

Intentional weight loss improves many of the medical complications associated with obesity. Moreover, many of these beneficial effects have a dose-dependent relationship with the amount of weight lost and begin after only a modest weight loss of 5% of initial body weight. In addition, weight loss can decrease the risk of new obesity-related diseases, such as diabetes.

Type 2 Diabetes Mellitus

In obese patients with type 2 diabetes mellitus, weight loss improves insulin sensitivity and glycemic control. A 1-year study of obese patients with type 2 diabetes treated with oral hypoglycemic agents showed that even a 5% weight loss decreased fasting blood glucose, insulin, and hemoglobin A\textsubscript{1c} concentrations and the required dosage of medication. Loss of 15% or more of their body weight decreased or eliminated the need for hypoglycemic medication. In patients with severe obesity who underwent gastric bypass surgery, the average loss of about 30% of initial body weight promoted marked long-term improvements in glucose homeostasis. In this study, normal fasting blood glucose, insulin, and glycoylated hemoglobin concentrations were achieved by 83% of the patients who had type 2 diabetes and 99% of the patients who had impaired glucose tolerance. However, a subset of obese patients with severe diabetes may not experience improved glycemic control with weight loss.

In obese patients with mild type 2 diabetes mellitus, both energy restriction and weight loss have important beneficial effects on insulin action and glycemic control. The initial negative energy balance associated with dieting acutely improves insulin sensitivity before there is a significant change in body weight. Subsequent weight and fat losses further improve glycemic control and insulin-mediated glucose uptake.

Sustained weight loss can prevent the development of new cases of diabetes. For example, the Swedish Obese Subjects (SOS) Study found that in severely obese patients (initial BMI 41 kg/m\textsuperscript{2}) who underwent gastric surgery, a 16% weight loss reduced the risk of diabetes fivefold over an 8-year period. Data from the Finnish Diabetes Prevention Study demonstrated that changes in lifestyle that resulted in modest (5%) weight loss decreased the 3-year incidence of diabetes by 58% in subjects with impaired glucose tolerance.

Several studies have found that weight loss is more difficult in obese patients with type 2 diabetes than in patients without diabetes. Moreover, successful weight loss may be inversely related to the duration and severity of diabetes. The reasons obese patients with diabetes are less responsive to weight loss therapy are not known but may involve the energy-conserving effects of improved glycemic control (e.g., reduced glycogenesis) and the tendency for weight gain associated with most drug treatments for diabetes.

Dyslipidemia

Weight loss usually decreases serum triglyceride, total cholesterol, and LDL-cholesterol concentrations and increases serum HDL cholesterol concentrations. In serum triglyceride, total cholesterol, and LDL cholesterol concentrations are generally greatest during the first 4 to 8 weeks of a weight loss program. Serum HDL cholesterol concentrations decrease during active weight loss but tend to increase when body weight stabilizes. A greater reduction in LDL cholesterol is observed when weight loss is induced by a program of diet plus exercise than with either treatment alone.

Hypertension

Systolic blood pressure and diastolic blood pressure decrease with weight loss, independent of sodium restriction. In the Trials of Hypertension Prevention, phase II (TOHP II), which is one of the largest intervention studies to date, approximately 1200 overweight and obese patients were randomly assigned to dietary weight loss intervention or usual care. This study showed a dose-response relationship between weight loss and change in blood pressure at 36 months. During the first 6 months, patients who lost weight experienced a marked reduction in blood pressure. Among patients who regained most or all of their lost weight, however, blood pressure steadily increased to near baseline values.

The marked weight loss induced by gastric surgery improved or completely resolved hypertension in about two thirds of extremely obese hypertensive patients. However, data compiled by the SOS Study indicate that the beneficial effect of weight loss on blood pressure may not persist. Much of the improvement in blood pressure observed at 1 and 2 years after gastric surgery disappeared by 3 years. Over the next 5 years, both systolic and diastolic blood pressures gradually increased. These findings imply that current energy balance and the direction of weight change are also important in blood pressure control.

A decreased incidence of hypertension with weight loss has been reported by several large prospective epidemiologic and intervention studies. For example, TOHP II found that persons who maintained a weight loss of at least 4.5 kg at 36 months had a 65% decrease in the risk of hypertension compared with the control group who gained 1.8 kg. The Nurses Health Study observed a direct correlation between the risk of developing hypertension and changes in body weight among normotensive women, who were monitored for 12 to 15 years. With weight losses of 5.0 to 9.9 kg and 10 or more kg, the risk of developing hypertension decreased by 15% and 26%, respectively.

Data from the SOS Study also question the ability of weight loss to prevent development of hypertension. In that study, the preventive effect of weight loss on the development of hypertension, which was observed 2 years after gastric surgery, disappeared at 3 years. In contrast, the SOS Study found a marked effect of weight loss on the incidence of other obesity-related diseases. For example, long-term maintenance of major weight loss after gastric surgery was associated with a marked and persistent reduction in the risk of diabetes.

Cardiovascular Disease

Modest weight loss can simultaneously affect the entire cluster of cardiovascular risk factors associated with obesity. In the Framingham Offspring Study, a weight loss of 5 pounds (2.25 kg) or more over 16 years was associated with 48% and 40% reductions in the sum of risk factors (defined as the highest quintile of systolic blood pressure, serum triglyceride, serum total cholesterol, fasting blood glucose, and BMI) and the lowest quintile of HDL-cholesterol) in men and women, respectively. Improvements in cardiovascular structure and function associated with weight loss included reductions in blood volume and hemodynamic demands on the heart, left ventricular mass and chamber size, and septal wall thickness. Such improvements in cardiac function may be responsible for the reduced frequency of chest pain and dyspnea reported by patients who lost weight after bariatric surgery.

Weight loss may also delay the progression of atherosclerosis. In one study, the progression of carotid artery intimal wall thickening over 4 years was three times higher in untreated obese subjects who did not lose weight than in obese subjects who lost weight after gastric surgery.

Pulmonary Disease

Weight loss results in improved pulmonary function, obstructive sleep apnea, and obesity-hypoventilation syndrome. Even modest weight loss resulted in reduced severity of sleep apnea, improved sleep patterns, and decreased daytime hypersomnolence. More marked weight loss, induced by bariatric surgery, has been
shown to improve obesity-hypoventilation syndrome by correcting resting room air arterial blood gases, lung volume, and cardiac filling pressure. These improvements in sleep apnea and obesity-hypoventilation syndrome are maintained with sustained weight loss; however, pulmonary symptoms recur with weight regain.

Reproductive and Urinary Tract Function in Women

Marked weight loss (>20% of initial body weight) has been shown to result in correction of urinary overflow incontinence, resolution of amenorrhea, and improved fertility.
Effect on Mortality

To date, there is no conclusive evidence that weight loss itself in obese persons reduces mortality. In fact, most epidemiologic studies indicate that weight loss or weight fluctuation increases mortality. However, these studies did not distinguish between intentional and unintentional weight loss, and their results may have been confounded by unintentional weight loss caused by concomitant rather than preexisting illness.

The effect of intentional weight loss on mortality has been addressed in three studies, which obtained baseline data between 1959 and 1960 and follow-up data for an average of 12 years. The composite results of these studies suggest that intentional, and possibly transient, weight loss may increase survival among overweight and obese persons who have type 2 diabetes mellitus. However, these data are not conclusive because weight loss was self-reported and occurred at any time before the initial interview, and possible changes in weight that might have occurred during follow-up were not determined. Therefore, long-term prospective trials are needed to determine the true relationship between intentional weight loss and survival in obese persons.
OBESITY THERAPY

Many obese persons can achieve short-term weight loss by dieting alone, but successful long-term weight maintenance is much more difficult to achieve. Weight cycling and yo-yo dieting are terms used popularly to describe repetitive cycles of weight loss and subsequent regain. 165 Although some adverse consequences have been associated with weight cycling, 166 available data on the health effects of weight cycling are inconclusive and should not deter obese persons from attempting to lose weight. 167 Currently available weight loss treatments include (1) dietary intervention, (2) increased physical activity, (3) behavior modification, (4) pharmacotherapy, and (5) surgery.

Dietary Intervention

For most obese people, negative energy balance is more readily achieved by decreasing food intake than by increasing physical activity. Therefore, dietary intervention is considered the cornerstone of weight loss therapy. Weight loss diets generally involve modifications of energy content and macronutrient composition. However, the degree of weight loss achieved depends primarily on the energy content, rather than the relative macronutrient composition, of the diet.

Energy Content

Weight loss diets can be classified according to their energy content:

1. Balanced-deficit diets of conventional foods usually contain more than 1500 kcal/day and an appropriate balance of macronutrients.
2. Low-calorie diets (LCDs) contain 800 to 1500 kcal/day and are consumed as liquid formula, nutritional bars, conventional food, or a combination of these items.
3. Very-low-calorie diets (VLCDs) contain less than 800 kcal/day and are generally high in protein (70 to 100 g/day) and low in fat (<15 g/day). Such diets may be consumed as a commercially prepared liquid formula and may include nutritional bars. VLCDs consumed as regular foods (mostly lean meat, fish, or fowl) are known as protein-sparing modified fasts.

According to treatment guidelines issued by the National Institutes of Health (NIH), 168 persons who are overweight (BMI of 25.0 to 29.9 kg/m²) and have two or more cardiovascular disease risk factors and persons with class I obesity (BMI of 30 to 34.9 kg/m²) should decrease their energy intake by approximately 500 kcal/day. This deficit in energy intake generally promotes weight loss of 1 pound (0.45 kg) per week and results in about a 10% reduction of initial weight at 6 months. The NIH guidelines recommend a more aggressive energy deficit of 500 to 1000 kcal/day for persons with more severe obesity (BMI of 35.0 kg/m² or higher). In such individuals, this energy deficit generally produces weight loss of 1 to 2 pounds/week and results in a 10% weight loss at 6 months.

Total daily energy requirements can be estimated by using standard equations (e.g., the Harris-Benedict equation 169 or the World Health Organization equation 170), which are based on the patient's size, age, gender, and activity level. However, the use of standard equations is cumbersome and may be unreliable in obese persons. Use of the simple dietary guidelines outlined in Table 33-4 is suggested as an alternative to a specific energy deficit diet based on the individual's daily energy requirements. Patients who follow these guidelines generally lose weight. Because many patients do not fully adhere to their prescribed diet, the energy content of the diet should be regularly adjusted according to the patient's weight loss response.

More than 30 prospective randomized clinical trials have investigated the effectiveness of LCDs for weight loss. 171 The composite results of these trials indicate that an LCD providing 1000 to 1500 kcal/day induces about an 8% weight loss after 16 to 26 weeks of treatment. However, these results may not be typical of the results obtained when an LCD is prescribed in routine clinical practice because trial participants volunteered to enroll in a weight loss study and most study protocols included some form of behavior modification therapy.

The use of VLCDs induced a weight loss of about 15% to 20% in 12 to 16 weeks of treatment, but this weight loss was not usually maintained. 172 In fact, several randomized trials have shown that weight regain is greater after VLCD than LCD therapy. 173 174 175 176 Therefore, 1 year after treatment, weight loss with a VLCD is often similar to that obtained with an LCD. In addition, initial weight losses with a VLCD and an LCD are similar when the diets are consumed in the same manner. For example, the weight loss observed in patients given a liquid diet providing 420 kcal/day was not significantly greater than that observed in persons who consumed a liquid diet providing 800 kcal/day. 177 This finding suggests that patients treated with VLCDs are either less compliant with the diet or sustain a greater decline in energy expenditure than those treated with LCDs. With VLCDs, there is a greater risk of the medical complications associated with dieting, such as hypokalemia, dehydration, and gallstone formation. Patients treated with a VLCD therefore require closer medical supervision than those treated with an LCD.

| TABLE 33-4 – Suggested Energy and Macronutrient Composition of Initial Reduced-Calorie Diet |
|-------------------------------------------------|----------------------------------|
| Body Weight (lb) | Suggested Energy intake (kcal/day) |
| 150199 | 1000 |
| 200249 | 1200 |
| 250299 | 1500 |
| 300349 | 1800 |
| 350 | 2000 |

Macronutrient Composition

| Total fat | 20%-30% of total calories |
| Saturated fatty acids | 8%-10% of total calories |
| Monounsaturated fatty acids | Up to 15% of total calories |
| Polyunsaturated fatty acids | Up to 10% of total calories |
| Cholesterol | <300 mg/day |
| Protein | 15%-20% of total calories |
| Carbohydrate | 55%-65% of total calories |

Macronutrient Composition

Alterating the macronutrient composition of the diet does not induce weight loss unless total energy intake is reduced. Lowfat diets have traditionally been prescribed for weight loss because such diets facilitate energy restriction. Triglycerides, the principal component of dietary fat, increase the palatability and energy density of food. The results of epidemiologic and diet intervention studies suggest that increased dietary fat intake is associated with increases in total energy intake and body...
Conversely, data from a large number of studies suggest that decreasing fat intake is associated with spontaneous decreases in total energy intake and body weight even when carbohydrate and protein intakes are not restricted.

A direct relationship between changes in dietary fat intake and body weight was found in a meta-analysis of 37 intervention studies involving the Step I or Step II low-fat (<30% kcal as fat) diet recommended by the National Cholesterol Education Program to lower cardiovascular risk. Data from another meta-analysis suggest that the amount of weight loss induced by a low-fat diet is directly related to the severity of obesity.

The weight loss effects of a low-fat diet may be related to the effect of dietary fat on energy density. Energy density is defined as the energy (i.e., calories) present in a given weight (grams) of food. Because the energy density of fat is so high, there is a high correlation between dietary fat content and dietary energy density. According to short-term studies lasting up to 14 days, energy intake is regulated according to the weight of ingested food rather than its fat or energy content. For example, the weight of food ingested was the same when lean and obese subjects were given either an ad libitum high-fat, high-energy-density (1.5 kcal/g) diet or a low-fat, low-energy-density (0.7 kcal/g) diet. As a result, energy intake with the high-fat, high-energy-density diet (3000 kcal/day) was nearly double the intake with the low-fat, low-energy-density diet (1570 kcal/day). In other studies, the weight of food ingested also remained the same when subjects were given liquid diets that had the same energy density but varied in fat content (20% to 60%) and when energy density was varied but fat content remained constant.

The results of these short-term studies show that dietary fat content itself does not affect total energy intake, apart from its effects on dietary energy density and food palatability. Whether diets of low energy density can help induce and maintain weight loss remains to be confirmed by long-term studies in obese subjects.

The weight loss effects of varying dietary carbohydrate content have also been investigated. For example, several short-term (<12 weeks) trials have compared the effects of low-carbohydrate and high-carbohydrate diets on weight loss when energy intake was kept constant. These studies suggest that, despite equal energy intakes, initial weight loss during the first 4 weeks may be greater with a low-carbohydrate than with a high-carbohydrate diet because a low-carbohydrate diet induces greater water loss; however, weight loss between 6 and 12 weeks was the same with either diet. Many currently popular low-carbohydrate diets restrict carbohydrate intake (e.g., the Atkins diet restricts carbohydrate intake to 20 g/day) but allow unlimited intake of fat and protein.

Some potentially valid explanations for promotion of weight loss by low-carbohydrate diets, despite unlimited fat and protein intakes, include (1) initial diuresis associated with ketone and urea nitrogen excretion, (2) losses of up to 100 kcal/day in urinary ketones, and (3) most important, decreased energy intake, which may be related to ketosis, diet monotony, or other unknown mechanisms.

Although it has been suggested that the high-fat and high-protein intakes associated with such diets may cause dehydration, electrolyte imbalance, hyperuricemia, calciumuria, kidney stones, glycogen depletion with easy fatigue, and hyperlipidemia potential adverse effects of long-term ingestion of such diets have not been carefully investigated. No serious adverse effects were reported in a 6-month trial involving 41 subjects receiving the Atkins diet. These subjects, in fact, showed a 43% decrease in plasma triglycerides, an 18% increase in plasma HDL-cholesterol, and a 7% decrease in plasma LDL-cholesterol. However, the safety and efficacy of low-carbohydrate diets need to be evaluated in long-term studies using randomized controlled trials.
Physical Activity

Metabolic Rate

Although there is a profound increase in energy expenditure during an actual episode of exercise, the addition of regular exercise to a weight loss program has negligible effects on REE. In a meta-analysis of prospective controlled trials in which obese subjects were randomly assigned to treatment with diet alone or diet plus exercise, the addition of exercise did circumvent the expected decline in REE when REE was adjusted for body mass.\[219\]

Body Composition

The composition of weight loss can be influenced by the addition of exercise to a diet program. Pooled data from two meta-analyses indicated that exercise can reduce the loss of FFM that occurs with weight loss.\[220\] When diet-induced weight loss was about 10 kg, regular exercise of low or moderate intensity reduced the percentage of weight lost as FFM from approximately 25% to 12%. Although the difference in weight lost as FFM was large on a percentage basis, it nonetheless represented only a small (1 kg) difference in the absolute amount of FFM lost. This preservation of FFM with exercise may not necessarily reflect preservation of muscle protein but may instead involve increased retention of body water and muscle glycogen. Indeed, nitrogen balance studies have not been able to detect any nitrogen-sparing effect of exercise during diet-induced weight loss in women.\[221\] Whether there is a difference between the effects of endurance and resistance exercise on FFM conservation is not clear because the available data are limited and conflicting.

Diabetes and Coronary Heart Disease

Endurance exercise increased insulin sensitivity\[223\] and was associated with a decreased risk of development of diabetes\[224\] and death from cardiovascular disease.\[225\]  

Weight Loss

Increasing physical activity alone is not an effective strategy for promoting initial weight loss. Most studies have shown that moderate endurance exercise, such as brisk walking for 45 to 60 minutes, four times a week, for up to 1 year, usually induces only minor weight loss.\[226\] In obese persons, the energy deficit created by exercise is usually much less and requires more effort than the energy deficit created by a reduced-calorie diet. For example, to lose 1 pound of fat, an obese patient would have to walk or run approximately 4.5 miles/day for 1 week or to consume a 500 kcal/day deficit diet for 1 week.\[227\]

Although exercise is not an effective strategy for inducing initial weight loss, increasing physical activity is an important component of successful long-term weight management. Several large-scale, cross-sectional case studies have shown that obese subjects who were successful in maintaining weight loss for 1 year or more engaged in regular exercise.\[228\] In several prospective randomized studies, subjects treated with diet plus exercise who continued to exercise sustained significantly larger long-term weight losses than subjects who stopped exercising or subjects treated with diet alone.\[229\] However, when data were analyzed on an intention-to-treat basis, most prospective randomized trials did not find that exercise had a statistically significant effect on the long-term maintenance of weight loss, presumably because adherence to the exercise program was often poor.\[230\]

It has been reported that obese patients need to expend approximately 2500 kcal/week to maintain weight loss.\[231\] This level of energy expenditure can be accomplished through vigorous activity (aerobics, cycling, or jogging) for approximately 30 minutes/day or more moderate activity (brisk walking) for 60 to 75 minutes/day. Most obese persons cannot easily achieve this level of activity. Therefore, prescribed activity goals should be initially modest and increased gradually over time.
Behavior Modification

Principles

Behavior modification therapy attempts to enable obese patients to recognize and subsequently alter eating and activity habits that promote their obesity. Behavior modification is derived from the classical conditioning principle that behaviors are often triggered by an antecedent event. The association between the antecedent event, such as watching television, and the behavior, such as eating, is strengthened by repetition so that the more often the two are paired, the stronger the association between them.

Behavior modification for the treatment of obesity usually involves multiple strategies to modify eating and activity habits. These strategies include:

1. Stimulus control (avoiding the cues that prompt eating).
2. Self-monitoring (keeping daily records of food intake and physical activity).
3. Problem-solving skills (developing a systematic manner of analyzing a problem and identifying possible solutions).
4. Cognitive restructuring (thinking in a positive manner).
5. Social support (cooperation from family members and friends in altering lifestyle behaviors).
6. Relapse prevention (methods to promote recovery from bouts of overeating or weight regain).

Effectiveness

Treatment by a comprehensive group behavior therapy approach generally results in about a 9% loss of initial weight in 20 to 26 weeks. When treatment ends, weight regain is commonly observed. Although patients generally regain about 30% to 35% of their lost weight in the year after treatment, most patients sustain a clinically significant weight loss of more than 5% of initial body weight. Increasing the duration of behavior therapy programs has only marginally improved total weight loss, but it probably prevents the weight regain that usually occurs when treatment is stopped.
Pharmacotherapy

Overview

Conventional obesity therapy is associated with a high rate of recidivism. Therefore, the most important goal of pharmacotherapy is to maintain long-term weight loss. Pharmacotherapy should not be considered a short-term approach for weight loss because patients who lose weight with drug therapy usually regain weight when the therapy is discontinued. 

Some obese patients do not respond to drug therapy, and long-term success is unlikely if weight loss does not occur within the first 4 weeks of drug treatment. Moreover, weight loss usually levels off by 6 months of treatment and weight begins to increase after 1 year. This observation implies that the efficacy of weight loss medications declines with time or obesity is a progressive disease, or both.

Treatment outcomes are less successful when pharmacotherapy is administered alone than when it is administered as part of a comprehensive weight loss program that includes diet, exercise, and behavior modification (Fig. 33-5). The use of obesity pharmacotherapy alone exposes patients to the full risks of the drug without the full medical benefits of more comprehensive treatment.

Table 33-5 lists the drugs currently approved by the United States Food and Drug Administration for the treatment of obesity. All such approved weight loss drugs act as anorexigens, with the exception of orlistat, which inhibits the absorption of dietary fat. Three anorexigent drugs have been withdrawn from the market because of the increased incidence of either valvular heart disease (fenfluramine and dexfenfluramine) or hemorrhagic stroke (phenylpropanolamine) associated with their use.

All anorexigent drugs, except mazindol, are derived from -phenylethylamine, the amphetamine precursor. The structures of these drugs have been chemically altered to reduce the potential for abuse. Anorexigent medications affect the monoamine (norepinephrine, serotonin, and dopamine) system in the hypothalamus and thereby enhance satiety (level of fullness during consumption of a meal, which influences the amount of food consumed), satiety (level of hunger after consumption of a meal, which influences the frequency of eating), or both. Monoamine neurotransmitters are synthesized from tyrosine and stored in granules that release their contents to presynaptic nerve terminals into the interneuronal cleft between presynaptic and postsynaptic nerves. Only a small portion of the monoamines released into the interneuronal cleft actually bind to postsynaptic receptors and thus transmit a signal from one nerve to the other. Most of the released monoamines are taken back up into the presynaptic nerve terminal, where they are either degraded or repackaged into granules for future release. Weight loss pharmacotherapy is approved for patients with no contraindications to therapy who have BMI greater than 30 kg/m² or a BMI between 27 and 29.9 kg/m² and an obesity-related medical condition. Because a comprehensive review of drug therapy for obesity has been published, we review only data from long-term (>6 months) prospective, randomized, controlled trials that investigated the weight loss efficacy and safety of sibutramine and orlistat, the only drugs currently approved for long-term use in the management of obesity.

Sibutramine

Sibutramine inhibits the neuronal reuptake of norepinephrine, serotonin, and, to a lesser degree, dopamine. It enhances satiety rather than satiety. In humans, sibutramine also appears to promote a small increase in metabolic rate several hours after its administration. The currently recommended initial dose of sibutramine is 10 mg/day. This daily dose can be decreased or increased by 5 mg if tolerance is poor or weight loss is inadequate. Administration of sibutramine at doses between 1 and 30 mg/day for 24 weeks resulted in a dose-dependent weight loss; the weight loss ranged from 0.9% of initial body weight with placebo to 7.7% with sibutramine at 30 mg/day.

Results from two 1-year, randomized, controlled trials of the effectiveness of sibutramine treatment in producing and maintaining weight loss have been reported. The results of one trial have appeared only in abstract form, and the other trial involved only obese subjects with medication-controlled hypertension. In both trials, all participants received minimal adjunctive weight management therapy and the placebo group lost less weight than usually observed in placebo groups from other trials. Subjects treated with sibutramine (10 to 20 mg/day) lost more weight than those treated with placebo. In the first study, 39% of patients randomly assigned to sibutramine therapy lost 10% or more of their initial body weight compared with 9% of those randomly assigned to receive placebo. In the second study of hypertensive obese patients, 13% of patients who received sibutramine therapy lost 10% or more of their body weight compared with 4% of those who received placebo.

Results have also been reported from two prospective, randomized, controlled trials that evaluated the efficacy of sibutramine therapy in long-term weight management after a predetermined amount of weight was lost. In the first trial, obese subjects who lost at least 6 kg after a 4-week VLCD resumed a regular diet with diet counseling and were randomly assigned to 1 year of treatment with placebo or sibutramine.
the year after randomization, sibutramine-treated subjects lost an additional 5.2 kg while placebo-treated subjects gained 0.5 kg. Total weight losses in the study were 12.9 kg in sibutramine-treated subjects and 6.9 kg in subjects treated with placebo. The initial weight loss achieved with the VLCD was maintained or increased in 74% of sibutramine-treated subjects compared with only 41% of placebo-treated subjects.

In the second trial, obese subjects who lost more than 5% of their initial weight after 6 months of treatment with sibutramine (10 mg/day) and a 600 kcal/day deficit diet were randomly assigned to treatment with either sibutramine (increased to 15 or 20 mg/day) or placebo. All subjects received dietary counseling. Nearly half of the subjects who entered the study failed to complete the 18-month treatment program. Among subjects who completed the study, 43% of those treated with sibutramine but only 16% of those treated with placebo maintained 80% or more of their original 6-month weight loss. On average, subjects who continued sibutramine maintained their weight loss for 1 year and then experienced a slight and progressive increase in weight; subjects who were switched to placebo experienced a progressive increase in weight as soon as sibutramine therapy was stopped.

The most common side effects of sibutramine therapy are dry mouth, headache, constipation, and insomnia. Sibutramine also causes small increases in blood pressure (2 to 4 mm Hg) and heart rate (4 to 6 beats/min). However, some patients experience much larger increases in blood pressure or heart rate and require dose reduction or discontinuation of therapy.

**Orlistat**

Orlistat is synthesized from lipstatin, a product of *Streptomyces toxytricini* mold, which inhibits most mammalian lipases. Orlistat binds to lipases in the gastrointestinal tract and thereby blocks the digestion of dietary triglycerides. This inhibition of fat digestion reduces micelle formation and, subsequently, the absorption of long-chain fatty acids, cholesterol, and certain fat-soluble vitamins. The degree of fat malabsorption is directly related in a curvilinear fashion to the dose of orlistat administered. Excretion of about 30% of ingested triglycerides, which is near the maximum plateau value, occurs at a dose of 360 mg/day (120 mg three times a day with meals). Orlistat has no effect on systemic lipases because less than 1% of the administered dose is absorbed.

Many clinical trials of orlistat included treatment with low doses (30 and 60 mg three times a day) that were not effective. Therefore, only data obtained with the standard recommended dose of 120 mg three times a day are reviewed here. The effectiveness of orlistat therapy (120 mg three times a day) in promoting and maintaining weight loss has been evaluated in several prospective, randomized, controlled trials more than 1 year in duration. At 1 year, about one-third more patients treated with orlistat than treated with placebo lost 5% or more of initial body weight; about twice as many patients treated with orlistat than treated with placebo lost 10% or more of initial weight. Subjects who were enrolled in a trial conducted within a primary care practice setting, which did not include behavior therapy or interaction with a dietitian, did not do as well as those enrolled in trials that provided formal behavior modification and dietary counseling.

Successful weight loss was also more difficult to achieve in patients with type 2 diabetes mellitus. The long-term efficacy of orlistat in maintaining initial weight loss after 1 year has been evaluated in several randomized, controlled trials, including second-year extensions of the 1-year trials just discussed. During the second year of these trials, more liberal energy intake was allowed with a goal of preventing weight regain rather than promoting additional weight loss. About half of the initially randomly assigned subjects completed the second year. After 1 year, both placebo-treated and orlistat-treated groups in all trials regained weight. However, at the end of 2 years, relative weight loss was greater with orlistat than with placebo treatment.

The results of several randomized clinical trials suggest that orlistat administration is associated with a reduction of serum cholesterol concentrations that is independent of the effect of weight loss alone. Even after adjusting for percent weight loss, these studies found that subjects treated with orlistat sustained a greater reduction in serum LDL-cholesterol concentrations than those treated with placebo. The mechanism responsible for this effect may be related to orlistat-induced inhibition of dietary cholesterol absorption.

The most common side effects associated with orlistat therapy are gastrointestinal complaints. Approximately 70% to 80% of subjects treated with orlistat experienced one or more gastrointestinal events, compared with approximately 50% to 60% of those treated with placebo. These gastrointestinal events were induced by fat malabsorption, usually occurred within the first 4 weeks of treatment, and were of mild or moderate intensity. Subjects rarely reported more than two episodes despite continued orlistat treatment. Orlistat treatment can also affect fat-soluble vitamin status and the absorption of some lipophilic medications. Therefore, it is recommended that all patients treated with orlistat also receive a daily multivitamin supplement and that orlistat not be taken for at least 2 hours before or after the ingestion of vitamin supplements or lipophilic drugs.
Surgical Therapy

Overview

Gastrointestinal surgery is the most effective approach for inducing major weight loss in extremely obese patients. In 1991, guidelines for the surgical treatment of obesity were established by an NIH consensus conference. According to these guidelines, eligible candidates for surgery include patients with a BMI of 40 kg/m² or higher or those with a BMI of 35.0 to 39.9 kg/m² and one or more severe medical complications of obesity (e.g., hypertension, heart failure, type 2 diabetes mellitus, or sleep apnea). Additional eligibility criteria are inability to maintain weight loss with conventional therapy, acceptable operative risks, absence of active substance abuse, and the ability to comply with the long-term treatment and follow-up required.

Current surgical procedures for obesity can be categorized as those that primarily cause gastric restriction and those that primarily cause nutrient malabsorption and malabsorption. All procedures have been performed laparoscopically, but the laparoscopic approach is technically challenging and usually requires more operating room time.

Types of Surgical Procedures

Gastric Bypass Procedure

The gastric bypass procedure (GBP), also known as Roux-en-Y gastric bypass, involves creating a small (10 to 30 mL) proximal gastric pouch that empties into a segment of jejunum that is anastomosed as a Roux-en-Y limb. The size of the patient determines the length of the Roux-en-Y limb (Fig. 33-6 A) (Figure Not Available). In patients with a BMI less than 50 kg/m², a 45- to 60-cm limb is generally used, but in patients with a BMI of 50 kg/m² or higher, a 150-cm limb (long-limb gastric bypass) promotes better weight loss without increasing the risk of nutrient deficiencies. Although the gastric bypass is primarily a restrictive procedure, some malabsorption does occur as a result of the bypassed stomach, duodenum, and upper jejunum.

Specific complications associated with the gastric bypass procedure

Data from four prospective, randomized trials showed that weight loss several years after surgery was consistently greater with the gastric bypass procedure (loss of 65% of excess weight) than with gastroplasty (loss of 40% of excess weight) (Fig. 33-6 B) (Figure Not Available). Specific complications associated with gastroplasty include staple line disruption, stomal stenosis, and gastroesophageal reflux.

Gastric Bypass versus Gastroplasty

In the laparoscopically inserted adjustable silicone gastric band (LASGB) procedure, a silicone band is placed around the upper stomach, just below the gastroesophageal junction. The band's circumference can be adjusted by inflating or deflating a balloon connected to a subcutaneously implanted port with percutaneous access.

This operation is currently the most popular bariatric surgical procedure performed in Europe and has been approved for use in the United States. The degree of weight loss achieved with this procedure has been similar to that achieved with vertical banded gastroplasty but on average has been much less than that achieved with the gastric bypass procedure.

Associated complications include esophageal dilatation, erosion of the band into the stomach, band slippage, band or port infections, and balloon or system leaks that lead to inadequate weight loss. Esophageal dilatation and dysphagia can result from the placement of the band at the gastroesophageal junction. Although loosening the band usually relieves the dilatation, removal of the band is sometimes necessary. In some patients, the band erodes into the stomach and must be surgically removed. When the posterior stomach wall herniates through the band, band slippage occurs. Band slippage can cause gastric obstruction and necessitates surgical revision.

Partial Bilipancreatic Bypass Procedures

The partial bilipancreatic bypass and the partial bilipancreatic bypass with duodenal switch result in both gastric restriction and maldigestion or malabsorption. Both procedures involve a partial gastrectomy and bypassing a considerable amount of small intestine from biliary and pancreatic secretions. Partial bilipancreatic bypass induces malabsorption of protein, fat, fat-soluble vitamins, iron, calcium, and vitamin B₁₂ and thus promotes more nutritional deficits than gastric restrictive procedures. The incidence of protein deficiency is probably less common and gastrointestinal side effects are not as severe after partial bilipancreatic bypass with duodenal switch than after partial bilipancreatic bypass with distal gastric resection. Presumably, these procedures cause greater weight loss (75% of excess weight) than a standard gastric bypass, but they have never been compared directly in a prospective randomized trial.

Jejunoileal Bypass

Jejunoileal bypass was first described in 1969. This procedure was designed to bypass the major portion of the small intestine and thereby promote weight loss by inducing the malabsorption of ingested nutrients. The procedure is no longer performed because of an unacceptable incidence of serious side effects. The serious side effects of jejunoileal bypass result from protein-calorie malnutrition, bacterial overgrowth and translocation, and excess oxalate absorption (e.g., colitis, interstitial nephritis, migratory arthritis, bypass enteritis, erythema nodosum, oxalate urolithiasis, hypocalcemia, and electrolyte imbalances). Oral metronidazole is effective in treating the complications of jejunoileal bypass related to bacterial overgrowth (e.g., migratory arthritis, elevated liver enzymes, bleeding from inflammation in the bypassed intestine).
About 15% of patients fail to lose more than 40% of their excess weight (10% to 15% of total weight) after a gastric bypass procedure. The percentage of patients who cannot lose this amount of weight after a gastropasty procedure is even greater. The major cause of inadequate weight loss after gastric bypass is the frequent ingestion of high-calorie soft foods and liquids (e.g., ice cream, cookies, milk shakes, and sodas) and high-fat snacks and fried foods (e.g., potato chips and fried potatoes). In patients who have undergone either a stapled gastropasty or gastric bypass, increased food intake may be related to staple line disruption, particularly if the patient is able to eat much larger quantities of food at a time.

**Perioperative Mortality**

The perioperative mortality rate after open surgical procedures for obesity reported in studies involving large numbers of patients was usually less than 1.5%. Approximately 75% of the deaths were due to anastomotic leaks and peritonitis, and 25% were due to pulmonary embolism.

---

### TABLE 33-6 -- Suggested Weight Loss Treatment Options Based on Body Mass Index and Risk Factors

<table>
<thead>
<tr>
<th>BMI Category (kg/m²)</th>
<th>BMI &gt; 25.0</th>
<th>BMI &gt; 27.0</th>
<th>BMI &gt; 30.0</th>
<th>BMI &gt; 35.0</th>
<th>BMI &gt; 40.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diet, physical activity, and behavior therapy</td>
<td>With CHD risk factor or obesity-related disease</td>
<td>With CHD risk factor or obesity-related disease</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Pharmacotherapy*</td>
<td>With obesity-related disease</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Surgery</td>
<td></td>
<td></td>
<td>With obesity-related disease</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>

**BMI, body mass index; CHD, coronary heart disease.**

*Pharmacotherapy should be considered only in patients who are not able to achieve adequate weight loss with available conventional therapy (diet, physical activity, and behavior therapy) and who do not have any absolute contraindications for drug therapy.

Bariatric surgery should be considered only in patients who are unable to lose weight with available conventional therapy and do not have any absolute contraindications for surgery.
TREATMENT GUIDELINES

A practical guide to the management of overweight and obesity was developed by the North American Association for the Study of Obesity in conjunction with the NIH. An overview of these guidelines is shown in Table 33-6. According to these guidelines, all overweight or obese patients who have type 2 diabetes should attempt to modify diet and physical activity behaviors. Certain behaviors are common among patients who have achieved successful long-term weight loss without bariatric surgery. Therefore, these four behaviors should be a goal for all patients:

1. Consume a diet that is low in calories (1300 to 1400 kcal/day) and fat (25% kcal as fat).
2. Engage in high levels of regular physical activity (expending about 2800 kcal/week equivalent to walking about 4 miles/day).
3. Monitor food intake and physical activity.
4. Check weight regularly.

Weight management is a key component in the treatment of overweight or obese patients with type 2 diabetes. Even a modest weight loss of 5% of initial body weight improves glycemic control and reduces the need for hypoglycemic medication. Moreover, modest weight loss improves other diabetes-related risk factors for CHD.

Unfortunately, successful weight management is more difficult to achieve in obese patients with type 2 diabetes than in those without diabetes. In fact, treatment of diabetes itself is usually associated with an increase in body weight. Therefore, the first principle of weight management in patients with diabetes is to use hypoglycemic therapy that is associated with the least amount of weight gain. Metformin is the preferred oral hypoglycemic agent because it produces minimal weight gain or slight weight loss. In addition, providing long-acting insulin at night is associated with less weight gain than more frequent dosing.
References


51. Leiter LA, Mariliss EB. Survival during fasting may depend on fat as well as protein stores. JAMA 1982; 248:23062307.
57. Leeks GE. Fatty acid regulation of very low density lipoprotein production. Curr Opin Lipidol 1997; 8:146153.

1638


References


Lipids are hydrophobic molecules that are insoluble or minimally soluble in water. They are found in cell membranes, which maintain cellular integrity and allow the cytoplasm to be compartmentalized into specific organelles. Lipids function as a major form of stored nutrients (triglycerides), as precursors of adrenal and gonadal steroids and bile acids (cholesterol), and as extracellular and intracellular messengers (e.g., prostaglandins, phosphatidylinositol). Lipoproteins provide a vehicle for transporting the complex lipids in the blood as water-soluble complexes and deliver lipids to cells throughout the body.

Classes of Lipids: Structure and Function

**Fatty Acids**

Fatty acids vary in length and in the number and position of double bonds. Saturated fatty acids lack double bonds (all carbon atoms have a full complement of hydrogen), and unsaturated fatty acids have one or more double bonds. Monounsaturated fatty acids have one double bond, and polyunsaturated fatty acids have two or more. The major fatty acids and their sources in foods are listed in Table 34-1.

Fatty acids are a readily available source of energy. In tissues, they can be esterified to other organic molecules to form complex lipids (e.g., triglycerides). In the blood, they may be transported on lipoproteins as complex lipids, or they may be transported in the nonesterified state as free fatty acids bound to albumin.

**Cholesterol**

Cholesterol is a four-ring hydrocarbon with an eight-carbon side chain. It plays a critical role as a major component of cell membranes and as a precursor of steroid hormones (adrenal and gonadal hormones). Cholesterol is also a precursor of bile acids, which are formed in the liver, stored in the gallbladder, and secreted in the intestine to participate in the absorption of fat. In the blood, about two thirds of the cholesterol is esterified (i.e., has a fatty acid esterified to the hydroxyl group at position 3).

**Complex Lipids**

**Triglycerides (Triacylglycerol)**

---

**TABLE 34-1 -- Major Fatty Acids**

<table>
<thead>
<tr>
<th>Chemical Designation</th>
<th>Common Name</th>
<th>Common Food Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saturated fatty acids (no double bonds)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C12:0</td>
<td>Lauric</td>
<td>Coconut oil</td>
</tr>
<tr>
<td>C14:0</td>
<td>Myristic</td>
<td>Coconut oil, butter fat</td>
</tr>
<tr>
<td>C16:0</td>
<td>Palmitic</td>
<td>Butter, cheese, meat</td>
</tr>
<tr>
<td>C18:0</td>
<td>Stearic</td>
<td>Beef, chocolate</td>
</tr>
<tr>
<td>Monounsaturated fatty acids (one double bond)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C18:19</td>
<td>Oleic</td>
<td>Olive oil, canola oil</td>
</tr>
<tr>
<td>Polyunsaturated fatty acids (two or more double bonds)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Omega-6 fatty acids</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C18:26, 12</td>
<td>Linoleic</td>
<td>Sunflower, corn, soybean, and safflower oils</td>
</tr>
<tr>
<td>C20:46, 8, 11, 14</td>
<td>Arachidonic</td>
<td></td>
</tr>
<tr>
<td>Omega-3 fatty acids</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C20:53</td>
<td>Eicosapentaenoic (EPA)</td>
<td>Salmon, cod, mackerel, tuna</td>
</tr>
<tr>
<td>C22:63</td>
<td>Docosahexaenoic (DHA)</td>
<td>Salmon, cod, mackerel, tuna</td>
</tr>
</tbody>
</table>

*The numeral after the C indicates the number of carbon atoms; the numeral after the colon indicates the number of double bonds. Carbon number 1 is the carboxylic acid carbon, and the carbon atom is the carbon most distant from the carboxyl group. The placement of the double bonds is shown by the designations (e.g., 9 indicates a double bond between carbons 9 and 10). In omega-6 fatty acids the first double bond occurs after the sixth carbon atom from the carbon atom (indicated by 6), and in omega-3 fatty acids it occurs after the third carbon atom from the carbon atom (3).
Triglycerides consist of three fatty acid molecules esterified to a glycerol molecule (see Fig. 34-1). Diglycerides (diacylglycerols) contain two fatty acids, and monoglycerides have only one fatty acid per glycerol molecule. Triglycerides serve to store fatty acids and form large lipid droplets in adipose tissue. They are also transported as a component of certain lipoproteins. When triglycerides are hydrolyzed in adipocytes or on lipoprotein particles, free fatty acid molecules are released to be used as a source of energy.

Phospholipids

Phospholipids have fatty acids esterified at two of the three hydroxyl groups of glycerol (see Fig. 34-1). The third hydroxyl group is esterified to phosphate (this complex lipid is referred to as phosphatidic acid). Typically, in mammalian tissue, the phosphatidic acid is esterified to the hydroxyl group of a hydrophilic molecule, such as choline, serine, or ethanolamine, to form phosphatidylcholine (commonly called lecithin), phosphatidylserine, or phosphatidylethanolamine, respectively. Lysolecithin is phosphatidylcholine from which one of the fatty acids has been removed. The combination of hydrophobic and hydrophilic regions in phospholipids enables them to be miscible at the water-lipid interface and makes them ideal components of membranes and of surface coats of lipoproteins. They are the most hydrophilic of the complex lipids.
Cholesterol Biosynthesis and the Low-Density Lipoprotein Receptor Pathway

Cholesterol is either absorbed from the diet or synthesized by cells in the body. All dietary cholesterol is of animal origin (i.e., from meats, dairy products, and eggs). Plants do not produce cholesterol; plant membranes contain sitosterol, which, except in a rare genetic disease, is not absorbed. Cholesterol is produced in many tissues (e.g., liver, skin, adrenals, gonads, brain, intestine). In most mammals, including humans, about 10% to 20% of the total synthesis of cholesterol occurs in the liver.

Cholesterol Biosynthesis

Cholesterol synthesis, illustrated schematically in Figure 34-2A, begins with acetate. Three molecules of acetate are condensed to form 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA), which is then converted to mevalonic acid by the enzyme HMG-CoA reductase. Through a series of steps, mevalonic acid is converted to cholesterol. The key (rate-limiting) step regulating cholesterol biosynthesis involves HMG-CoA reductase. Competitive inhibitors of this enzyme (the statins) reduce cholesterol biosynthesis and lower plasma cholesterol levels. Increased cholesterol content of cells feeds back on the HMG-CoA reductase and decreases its activity, thereby decreasing cholesterol biosynthesis. Conversely, a deficiency of intracellular cholesterol increases reductase activity and increases cholesterol biosynthesis (see later discussion).

Cholesterol cannot be eliminated by catabolism to carbon dioxide and water; it must be either excreted as free cholesterol in the bile or converted to bile acids and secreted into the intestine. About 50% of the cholesterol entering the intestine is reabsorbed and recirculates to the liver; the remainder is eliminated in the feces. Almost all of the secreted bile acids (97%) are reabsorbed from the intestine and transported back to the liver. This recirculation of cholesterol and bile acids from the intestine to the liver is called the enterohepatic circulation (Fig. 34-2B). The reabsorbed cholesterol and bile acids regulate de novo cholesterol and bile acid synthesis in the liver. For example, if the amount of bile acids returning to the liver is decreased (as occurs in the intestine during treatment with bile acidbinding resins), bile acid synthesis is increased, enhancing the amount of cholesterol being converted to bile acids.

Cholesterol 7-hydroxylase

This enzyme of about 57 kDa (503 amino acids), known as CYP7A (formerly P450 7A1), converts free cholesterol to 7-hydroxycholesterol. This is the rate-limiting step in bile acid synthesis, and it is under feedback regulation by recirculated bile acids. The interruption of bile acid recirculation increases cholesterol 7-hydroxylase activity. This enzyme and HMG-CoA reductase are closely coupled, and their activities usually change in parallel (for a review see references 13 and 17). In this way, the intracellular cholesterol level for bile acid production remains rather constant.

Low-Density Lipoprotein Receptor

Cholesterol levels in the blood are controlled primarily through the low-density lipoprotein (LDL) receptor pathway. This receptor is present on the surface of all cells throughout the body, including hepatocytes, and mediates the uptake of cholesterol-rich lipoproteins (e.g., LDL) from the blood. Specific proteins on the surface of certain lipoproteins (apolipoprotein [apo] B100 and apo-E) interact with the LDL receptor and facilitate lipoprotein internalization by cells. By this mechanism, cells that require cholesterol can obtain the preformed sterol. The LDL receptor also allows the liver (the principal site for LDL catabolism) to take up LDL and eliminate cholesterol from the body (discussed in "Lipoprotein Receptors Controlling Lipoprotein Metabolism").

The number of LDL receptors on the cell surface is tightly regulated. If the cholesterol content of a cell is elevated, fewer receptors are synthesized (i.e., receptor expression is down-regulated). On the other hand, if a cell requires cholesterol, expression of LDL receptors is up-regulated and synthesis increases. This system keeps the intracellular cholesterol concentration relatively constant and prevents excessive and possibly toxic accumulation. Within the cell, cholesterol can be esterified by the enzyme acyl-CoA:cholesterol acyltransferase (ACAT).

Acyl-CoA:Cholesterol Acyltransferase

ACAT is an enzyme of the endoplasmic reticulum (ER) (about 45 to 50 kDa, 550 amino acids) that catalyzes the formation of cholesteryl esters from long-chain fatty acyl-CoA (e.g., oleoyl-CoA) and free cholesterol substrates. When lipoproteins enter the cell by receptor-mediated endocytosis and are degraded within the lysosomes, the free cholesterol released can be transported to the ER, where it is esterified by ACAT. There are two ACAT enzymes. ACAT1 is present in macrophages, steroidogenic tissues, and sebaceous glands, and its action in macrophages has been implicated in foam cell formation and atherogenesis. ACAT2 acts to promote the absorption of dietary cholesterol. Agents that inhibit intestinal ACAT activity may provide a means to limit cholesterol absorption by the intestine.

Cholesteryl ester hydrolysis by cholesterol ester hydrolase generates free cholesterol, either for efflux from the cells or to serve as a biosynthetic substrate (e.g., for steroid hormones and cell membranes) within the cells. The pool of intracellular cholesterol and cholesteryl esters is dynamic.
Metabolism of Dietary Lipids

The digestion of dietary fats begins in the stomach and continues in the proximal small intestine. Triglycerides are hydrolyzed to free fatty acids and small amounts of monoglycerides and diglycerides, cholesteryl esters are hydrolyzed to free cholesterol, and phospholipids are converted primarily to lysophosphatidylcholine. Bile salt micelles disperse and partially solubilize water-insoluble lipids; this facilitates the intestinal transport and delivery of lipids to the unstirred water layer of intestinal epithelial cells, where they can be taken up by the cells. Bile acids also activate pancreatic lipase, which participates in the hydrolysis of triglycerides. Long-chain fatty acids are taken up primarily by the enterocytes of the duodenum and proximal jejunum, reesterified into triglycerides, and used in the biosynthesis of intestinal lipoproteins (chylomicrons), which are delivered to the mesenteric lymph and enter the general circulation with the thoracic duct lymph. Medium-chain fatty acids (10 carbons) are absorbed into the portal blood without being esterified and are cleared directly from the blood by the liver. Bile acids are reabsorbed primarily from the ileum, enter the portal blood, and are taken up by the liver.
Triglyceride and Free Fatty Acid Metabolism

Storage and Use

Free fatty acids are released from triglycerides of chylomicrons and very-low-density lipoproteins (VLDLs) through the action of lipoprotein lipase (LPL). LPL is bound to the capillary endothelial cells adjacent to adipose, muscle, and breast tissue, where it liberates free fatty acids from lipoprotein triglyceride. The level of LPL in tissues differs under different physiologic circumstances so that free fatty acids are directed to tissues requiring them as substrates or energy sources. During fasting, for example, LPL activity decreases in adipose tissue and increases in heart muscle. In the breast, LPL levels are low until parturition, when they increase 10-fold to promote milk formation.

Figure 34-3 Triglyceride synthesis and the DGAT reaction. A, Two major pathways of triglyceride synthesis have been described: the glycerol-phosphate pathway and the monoglyceride pathway, which is prominent in the small intestine. B, DGAT catalyzes a reaction in which 1,2-diacylglycerol and fatty acyl-CoA react to form triacylglycerol at the surface of the endoplasmic reticulum. (From Farese RV Jr, Cases S, Smith SJ. Triglyceride synthesis: insights from the cloning of diacylglycerol acyltransferase. Curr Opin Lipidol 2000; 11:229-234.)

In adipose tissue, high levels of glucose and insulin promote the conversion of free fatty acids to triglyceride for storage. Insulin stimulates LPL activity and fatty acid esterification through the formation of glycerol phosphate and decreases free fatty acid release through the inhibition of hormone-sensitive lipase. Insulin deficiency, as in diabetes mellitus, is associated with decreased LPL activity. Insulin and glucose also stimulate the biosynthesis of free fatty acids in the liver and, to a lesser degree, in adipocytes when dietary fat is replaced by carbohydrates. As a result, hepatic free fatty acids are converted to triglyceride and packaged into VLDL particles (discussed in the section on plasma lipoproteins).

Acyl-Coenzyme A:Diacylglycerol Acyltransferase

Triglyceride (triacylglycerol) synthesis is catalyzed by the enzyme acyl-CoA:diacylglycerol acyltransferase (DGAT) (Fig. 34-3). A DGAT gene, which is expressed in all tissues, has been identified. Interestingly, the inactivation of this gene in mice has revealed that multiple pathways exist for triglyceride synthesis. These alternative mechanisms might include a second DGAT or fatty acyl-CoA-independent mechanisms. The gene inactivation study also revealed that DGAT plays an important role in energy metabolism.

Fatty Acid Release from Adipose Tissue

The net release of free fatty acids and glycerol from adipose triglyceride stores occurs during various physiologic conditions, including stress, exercise, fasting, and uncontrolled diabetes mellitus. This release occurs in response to hormones (Table 34-2), most of which act by means of cyclic adenosine monophosphate to activate a hormone receptor-coupled protein kinase that in turn activates a hormone-sensitive lipase (Fig. 34-4). In contrast to numerous hormones, insulin inhibits rather than stimulates hormone-sensitive lipase in adipose tissue. Growth hormone liberates free fatty acids by a different mechanism, which requires enhanced synthesis of hormone-sensitive lipase.

After triglyceride hydrolysis in adipose tissue, the released free fatty acids bind to albumin and circulate in the plasma. Released glycerol is taken up by the liver and kidney for triglyceride synthesis or for gluconeogenesis. The fate of the free fatty acid-albumin complexes is determined in part by the blood flow. With intense exercise and diminished blood flow to the splanchnic bed, free fatty acids are targeted to muscle. Depending on the metabolic state, free fatty acids taken up by the liver are reused for triglyceride or phospholipid synthesis (exported on VLDL), oxidized to carbon dioxide, or converted to ketone bodies.

TABLE 34-2 -- Hormones That Affect Lipolysis in Vitro

<table>
<thead>
<tr>
<th></th>
<th>Hormones That Affect Lipolysis in Vitro</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rapid Stimulation</td>
<td>Catecholamines (-1 agonists)</td>
</tr>
<tr>
<td></td>
<td>Corticotropic</td>
</tr>
<tr>
<td></td>
<td>Glucagon</td>
</tr>
<tr>
<td></td>
<td>Placental lactogen</td>
</tr>
<tr>
<td></td>
<td>Prolactin</td>
</tr>
<tr>
<td></td>
<td>Secretion</td>
</tr>
<tr>
<td></td>
<td>Thrytropin</td>
</tr>
<tr>
<td></td>
<td>Vasoactive intestinal peptide</td>
</tr>
<tr>
<td></td>
<td>Vasopressin</td>
</tr>
<tr>
<td>Slow Stimulation</td>
<td>Glucocorticoids</td>
</tr>
<tr>
<td></td>
<td>Growth hormone</td>
</tr>
<tr>
<td>Suppression</td>
<td>Insulin</td>
</tr>
<tr>
<td></td>
<td>Gastric inhibitory polypeptide</td>
</tr>
<tr>
<td></td>
<td>Oxytocin</td>
</tr>
<tr>
<td></td>
<td>Prostaglandin</td>
</tr>
<tr>
<td></td>
<td>Somatomedins</td>
</tr>
</tbody>
</table>


Fatty Acid Oxidation and Ketogenesis

Both oxidation and ketogenesis of fatty acids take place in the mitochondria, except for very-long-chain fatty acids (C-24 and C-26), which are oxidized in peroxisomes. Because free fatty acids and their CoA derivatives can penetrate only the outer leaflet of the mitochondrial membrane, they are converted to carnitine...
derivatives within the mitochondrial membrane to allow transport across the inner membrane. Once inside the mitochondria, they are reconverted to CoA derivatives and undergo beta oxidation, which produces acetyl-CoA and the reduced forms of nicotinamide-adenine dinucleotide (NADH) and flavin-adenine dinucleotide (FADH).

With a normal flux of free fatty acids, the NADH and FADH enter the electron transport system, resulting in the formation of adenosine triphosphate and water. The condensation of acetyl-CoA with oxaloacetic acid yields citrate, which can enter the citric acid cycle, where it is oxidized to carbon dioxide or is transported out of the mitochondria and converted again to free fatty acids. If free fatty acid flux to the liver is massively increased, as in insulin-deficient states such as prolonged fasting or uncontrolled diabetes mellitus, the production of VLDL triglyceride from free fatty acids is limited. As a result, NADH, FADH, and acetyl-CoA accumulate in the mitochondria and give rise to the products of ketogenesis: acetoacetate, -hydroxybutyrate, and acetone.

Ketogenesis occurs in several steps. Initially, acetyl-CoA condenses in two steps to form acetoacetyl-CoA and then HMG-CoA. The latter is cleaved to acetoacetate and acetyl-CoA, which leads to the liberation of CoA and its use in beta oxidation of free fatty acids. Acetoacetate can be reduced by NADH to form -hydroxybutyrate; the NAD produced can be used for continued beta oxidation of fatty acids. Alternatively, the acetoacetate can decompose to form acetone. The ketones are released into the plasma and, if they accumulate, cause ketoacidosis.

Fatty Acid Biosynthesis

Under normal conditions, the diet supplies sufficient fatty acids through the ingestion of fat. However, increases in the ratio of carbohydrate to fat in the diet stimulate fatty acid synthesis by the liver and adipose tissue. Fatty acids are synthesized from two carbon units of acetyl-CoA. Because acetyl-CoA is produced in the mitochondria, it must first be converted to citrate by condensation with oxaloacetate and then transported into the cytosol, where it is reconverted to acetyl-CoA and oxaloacetate. Eight acetyl-CoA units are condensed to form palmitic acid (16 carbon atoms) in a series of reactions involving the enzymes fatty acid synthase and acetyl-CoA carboxylase. Longer fatty acids, such as stearic acid (18 carbon atoms) or oleic acid (18 carbon atoms with a single double bond), are synthesized from palmitic acid by chain extension. In this way, fatty acid synthesis can meet most of the body's requirements.

Certain essential polyunsaturated fatty acids cannot be synthesized in humans and must be supplied in the diet. These include linoleic acid (18 carbon atoms with two double bonds) and linolenic acid (18 carbon atoms with three double bonds). Essential fatty acids are required for a number of special functions, including prostaglandin synthesis.
PLASMA LIPOPROTEINS: APOLIPOPROTEINS, RECEPTORS, AND ENZYMES

General Structure and Major Classes of Lipoproteins

Lipoproteins function as vehicles to transport lipids in the blood in the form of soluble complexes of lipids and proteins. The lipids include triglycerides, cholesteryl esters, free cholesterol, and phospholipids. About 10 different protein moieties, called apolipoproteins, are associated with various lipoproteins and are given letter designations (Table 34-3). Lipoproteins also transport fat-soluble vitamins (A, D, and E), drugs (e.g., probucol, cyclosporine), some viruses, and certain antioxidant enzymes (e.g., paraoxonase and platelet-derived activating factor hydrolase).

Lipoproteins are spherical particles with a core of mostly hydrophobic lipids (triglycerides and cholesteryl esters) and a surface layer of more hydrophilic constituents, namely protein, free cholesterol, and phospholipids (Fig. 34-5). Six major classes of lipoproteins play different roles in lipid transport (Table 34-4), and the specific apolipoproteins on the surface determine the fate of the lipoproteins. To understand lipoprotein metabolism and the diseases associated with lipid abnormalities, it is necessary to consider the roles of the individual apolipoproteins in regulating lipid metabolism. Some of their properties are summarized in Table 34-4 and Figure 34-6.

### TABLE 34-3 – Characteristics and Major Functions of Human Apolipoproteins

<table>
<thead>
<tr>
<th>Apolipoprotein</th>
<th>Average Plasma Concentration (mg/dL)</th>
<th>Chromosome</th>
<th>Gene (bases)</th>
<th>Molecular Weight (× 1000)</th>
<th>Mature Protein (amino acids)</th>
<th>Major Sites of Synthesis</th>
<th>Major Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI</td>
<td>130</td>
<td>11</td>
<td>1863</td>
<td>29</td>
<td>243</td>
<td>Liver, intestine</td>
<td>Structural protein/HDL. Cofactor for LCAT. Ligand for putative HDL receptor.</td>
</tr>
<tr>
<td>AII</td>
<td>40</td>
<td>1</td>
<td>1330</td>
<td>17 (dimer)</td>
<td>77</td>
<td>Liver</td>
<td>Inhibits apo-E binding to receptors (through the EAII complex).</td>
</tr>
<tr>
<td>AIV</td>
<td>40</td>
<td>11</td>
<td>2600</td>
<td>45</td>
<td>376</td>
<td>Intestine</td>
<td>May facilitate cholesterol efflux from cells. Activator of LCAT. Possible role in triglyceride metabolism.</td>
</tr>
<tr>
<td>B100</td>
<td>85</td>
<td>2</td>
<td>43,000</td>
<td>513</td>
<td>4536</td>
<td>Liver</td>
<td>Structural protein/VLDL and LDL. Ligand for LDL receptor.</td>
</tr>
<tr>
<td>B48</td>
<td>Variable</td>
<td></td>
<td>241</td>
<td>2152</td>
<td>Intestine</td>
<td>Structural protein/chylomicrons.</td>
<td></td>
</tr>
<tr>
<td>CII</td>
<td>6</td>
<td>19</td>
<td>4653</td>
<td>6.6</td>
<td>57</td>
<td>Liver</td>
<td>Modulates remnant binding to receptors. Activates LCAT.</td>
</tr>
<tr>
<td>CIII</td>
<td>3</td>
<td>19</td>
<td>3320</td>
<td>8.9</td>
<td>79</td>
<td>Liver</td>
<td>Cofactor for LPL.</td>
</tr>
<tr>
<td>E</td>
<td>5</td>
<td>19</td>
<td>3597</td>
<td>34</td>
<td>299</td>
<td>Liver, brain, skin, testes, spleen</td>
<td>Ligand for LDL and remnant receptors. Local lipid redistribution. Reverse cholesterol transport (HDL with apo-E).</td>
</tr>
<tr>
<td>Apo(a)</td>
<td>Variable</td>
<td></td>
<td>400800</td>
<td>40006000</td>
<td>Liver</td>
<td>Modulates thrombosis/fibrinolysis.</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>10</td>
<td>3</td>
<td>12,000</td>
<td>20</td>
<td>169</td>
<td>Liver, intestine</td>
<td>Activator of LCAT (?).</td>
</tr>
</tbody>
</table>

Apo-E, apolipoprotein E; HDL, high-density lipoprotein; LCAT, lecithin:cholesterol acyltransferase; LDL, low-density lipoprotein; LPL, lipoprotein lipase; VLDL, very-low-density lipoprotein.
Major Apolipoproteins Regulating Lipoprotein Metabolism

Apolipoprotein B

In human plasma, apo-B occurs in two forms, apo-B100 and apo-B48, which are derived from a single gene **B100** on the short arm of chromosome 2. The human apo-B gene comprises 29 exons and 28 introns and is approximately 45 kilobases in length. A unique ribonuclease acid (RNA) editing mechanism is responsible for the synthesis of apo-B100 and apo-B48 from the apo-B messenger RNA (mRNA) (Fig. 34-7) for a review see references 51 and 57. An editing protein (or proteins) interacts with the apo-B mRNA in the human intestine to change a single nucleotide, resulting in the synthesis of a truncated form of apo-B (apo-B48). In humans, this modification of the apo-B mRNA occurs only in the intestine and not in the liver; therefore, the liver produces the full-length apo-B100. Apo-B100 (but not apo-B48) is also expressed in the yolk sac of mammals.

The editing of apo-B mRNA results in the change of cytosine-6666 in the apo-B100 mRNA to a uracil. This cytosine is part of the codon CAA, which encodes a glutamine at amino acid residue 2153 in apo-B100, whereas the codon UAA is a stop codon and terminates translation of the protein chain (see Fig. 34-7). Therefore, apo-B48 possesses only 2152 amino acids, compared with the 4536 in apo-B100. Apo-B100, a 513-kd protein, is synthesized in the liver; it serves as a structural protein of VLDL and of intermediate-density lipoproteins (IDLs) and is the exclusive protein constituent of LDL. Each VLDL, IDL, and LDL particle contains one molecule of apo-B100. The primary structure of apo-B contains many hydrophobic and amphipathic sequences, forming alpha helices and beta strands, that occur throughout the molecule and appear to function as lipid-binding domains. In addition to its structural role, apo-B100 functions as a ligand for the LDL receptor.

Apo-B48, a 241-kd protein, is a structural constituent of chylomicrons. Each chylomicron appears to possess one or two apo-B48 molecules. The apo-B48 is a 17-kd protein, which is derived from apo-B100 by the selective chemical modification of the apo-B100 mRNA in the human intestine to change this single nucleotide, resulting in the synthesis of the truncated form of apo-B (apo-B48) (see Fig. 34-7).

Table 34-4

<table>
<thead>
<tr>
<th>Type</th>
<th>Density (g/mL)</th>
<th>Electrophoretic Mobility</th>
<th>Site of Origin</th>
<th>Major Lipids</th>
<th>Major Apolipoproteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chylomicrons</td>
<td>&lt;0.95</td>
<td>Origin</td>
<td>Intestine</td>
<td>85% Triglyceride</td>
<td>B48, Al, AIV (E, CI, CII, CIII)</td>
</tr>
<tr>
<td>Chylomicron remnants</td>
<td>&lt;1.006</td>
<td>Pre-</td>
<td>Liver</td>
<td>60% Triglyceride, 20% cholesterol</td>
<td>B48, E</td>
</tr>
<tr>
<td>VLDL</td>
<td>&lt;1.061</td>
<td>Liver</td>
<td>Derived from VLDL</td>
<td>35% Cholesterol, 25% triglyceride</td>
<td>B100, E</td>
</tr>
<tr>
<td>IDL</td>
<td>1.0191</td>
<td>Liver</td>
<td>Derived from IDL</td>
<td>60% Cholesterol, 5% triglyceride</td>
<td>B100</td>
</tr>
<tr>
<td>LDL</td>
<td>1.0631</td>
<td>Liver, intestine, plasma</td>
<td>25% Phospholipid, 20% cholesterol, 5% triglyceride (50% protein)</td>
<td>Al, AII, CI, CII, CIII, E</td>
<td></td>
</tr>
<tr>
<td>HDL2</td>
<td>1.0631.125</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDL3</td>
<td>1.1251.21</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

HDL, high-density lipoprotein; IDL, intermediate-density lipoprotein; LDL, low-density lipoprotein; VLDL, very-low-density lipoprotein.

*Small, partially lipolyzed VLDL and IDL are often called VLDL remnants.*

Possibly two molecules of apo-B48. Because it lacks the carboxyl-terminal domain of apo-B100, apo-B48 cannot bind to the LDL receptor. The carboxyl-terminal domain of apo-B100 in the region of amino acids 3000 to 3700 is critical for the binding of apo-B100 to the LDL receptor (Fig. 34-8). Selective chemical modification of the apo-B100 of LDL demonstrated that the positively charged (basic) amino acids arginine and lysine are important in the interaction of LDL with its receptor. When apo-B100 was sequenced, several regions enriched in arginines and lysines became candidates for receptor binding. It is now apparent that the basic residues in the region of amino acids 3359 to 3369 are critical for receptor binding. However, it is also clear that the carboxyl-terminal region of apo-B100 in the vicinity of amino acid 3500 can modulate receptor-binding activity. Patients expressing apo-B48 have hypercholesterolemia and high LDL levels. This genetic disorder, familial defective apo-B100 (see later discussion), is caused by the substitution of glutamine for arginine at amino acid 3500.

Figure 34-5 General structure of lipoproteins (a schematic representation of very-low-density lipoprotein, VLDL).

Figure 34-6 Polyacrylamide gel showing the various apolipoproteins characteristic of each type of plasma lipoprotein particle. HDL, high-density lipoprotein; LDL, low-density lipoprotein; VLDL, very-low-density lipoprotein. (Modified from Mahley RW, Interaneri TL. Lipoprotein receptors and cholesterol homeostasis. Biochim Biophys Acta 1983; 737:197-222. With permission from Elsevier Science-NL, Sara Burgerhartsstraat 25, 1055 KV Amsterdam, The Netherlands.)

Role of Apolipoprotein B in Lipid Metabolism

Apo-B100 and apo-B48 play critical roles in the biosynthesis of apo-B-containing lipoproteins (for a review, see references 51 and 57). In addition, the apo-B100 in VLDL interacts with the LDL receptor. Although it is also a constituent of VLDL and IDL, apo-B100 does not play a major role in the binding of these lipoproteins to LDL receptors. Apo-E is responsible for most of the receptor-mediated clearance of VLDL and IDL; presumably, the lipid or apolipoprotein content of the VLDL and IDL masks or alters the conformation of the receptor-binding domain of the apo-B100 on these particles. Apo-B100 is, however, the major (or exclusive) protein moiety of LDL and is responsible for directing the clearance of these lipoproteins through the LDL receptor pathway.
Overexpression of apo-B in transgenic mice increases the levels of LDL and other apo-B-containing lipoproteins, resulting in increased susceptibility to diet-induced atherosclerosis. Knockout

**Apolipoprotein E**

Apo-E mediates the interaction of apo-E-containing lipoproteins with the LDL receptor and with the chylomicron remnant (apo-E) receptor, presumably the LDL receptor-related protein (LRP). As a consequence, apo-E plays a critical role in determining the metabolic fate of several classes of lipoproteins and is of central importance in cholesterol metabolism. In addition, apo-E appears to participate in cholesterol transport to cells undergoing proliferation and repair and may modulate lymphocyte response and smooth muscle cell proliferation (see references 77 and 78). Apo-E, a 34-kDa protein composed of 299 amino acids, circulates in the plasma both as a constituent of chylomicrons, chylomicron remnants, VLDL, and LDL and as a component of a minor subclass of high-density lipoproteins (HDLs), referred to as HDL with apo-E or HDL-E (see Fig. 34-8). Normal plasma apo-E levels range from 30 to 70 µg/mL, approximately half of which is associated with HDL and serves as a reservoir of apo-E for redistribution to chylomicrons and VLDLs as they enter the plasma. In lymph and interstitial fluid, apo-E is associated with lipid complexes (phospholipid-apo-E discs) or with HDL.

Approximately 75% of plasma apo-E is synthesized by hepatocytes, and the remainder is synthesized in a variety of tissues. Macrophages can synthesize and secrete the protein, especially when they are loaded with cholesterol, and are responsible for a portion of the apo-E found in interstitial fluid. Smooth muscle cells of arteries and keratinocytes in the skin also synthesize apo-E (see Table 34-3). The tissue with the second highest level of apo-E mRNA (after the liver) is the brain, where apo-E is synthesized primarily by astrocytes. Cerebrospinal fluid contains apo-E derived from the brain (approximately 0.3 mg/dL, or 5% to 10% of plasma apo-E levels). Apo-E appears to play a key role in cholesterol transport in both the central and peripheral nervous systems and may be involved in the pathogenesis of Alzheimer's disease (for a review, see references 77 and 78).

The apo-E gene is located on chromosome 19 and is part of a gene cluster that includes the genes for apo-CI and apo-CII. The apo-E gene locus has multiple alleles, and the three major forms of apo-E—apo-E2, apo-E3, and apo-E4—are encoded by three different alleles, referred to as epsilon2, epsilon3, and epsilon4, that occur in several populations with a frequency of about 8%, 77%, and 15%, respectively (see Fig. 34-9). There are three homoyzous (E2/2, E3/3, and E4/4) and three heterozygous (E3/2, E4/2, and E4/3) phenotypes. About 60% of individuals are homozygous for apo-E3.

These genetic polymorphisms are caused by amino acid differences at two sites in the protein (see Fig. 34-9). Apo-E3 has cysteine at position 112 and arginine at position 158, whereas apo-E2 has cysteine at both positions and apo-E4 has arginine. In addition, apo-E displays a second type of polymorphism, post-translational glycosylation. Carbohydrate attachment at threonine-194 and the presence of multiple sialic acid residues give rise to minor acidic isoforms.

Apo-E functions in both receptor binding and lipid binding, and the different isoforms have different activities. Apo-E3 and apo-E4 are equally capable of interacting with LDL receptors, but the binding of apo-E2 to LDL receptors is impaired and is associated with the development of type III hyperlipoproteinemia under certain conditions. Apo-E isoforms also interact differently with specific types of lipids and lipoproteins. Apo-E binds preferentially to large, triglyceride-rich lipoproteins (e.g., VLDL), whereas apo-E3 and apo-E2 bind preferentially to smaller, phospholipid-rich HDL.

The apo-E primary translational product is a 317-amino-acid protein; an 18-amino-acid signal peptide is cleaved before the mature protein (299 amino acids, relative molecular mass 32,000).
34 kDa) is secreted into plasma. The molecule has two domains (Fig. 34-10). The amino-terminal domain (residues 1 to 191) contains the receptor-binding region, and the carboxyl-terminal domain (residues 192 to 299) appears to have three amphipathic alpha helices (one face being hydrophilic and the other hydrophobic) and is responsible for lipid binding. Residues 242 to 272 are key in the binding of apo-E to lipoproteins. Paradoxically, the lipid-binding region of apo-E resides in the carboxyl-terminal domain, but the amino acid differences that distinguish the three major apo-E isoforms are in the amino-terminal domain (residues 112 and 158). The fact that the isoforms display different specificities for different types of lipoproteins (i.e., apo-E4 for VLDL and apo-E3 and apo-E2 for HDL) suggests that the amino-terminal and carboxyl-terminal domains interact so that specific residues in the amino terminus alter the conformation and specificity of the lipid-binding domain in the carboxyl terminus for certain types of lipoproteins (for a more complete discussion, see references 23 and 24).

The amino acids of apo-E that mediate its binding to the LDL receptor are in the vicinity of residues 134 to 160 (Fig. 34-11). Positively charged arginines and lysines between amino acids 136 and 150 may interact with the negatively charged glutamyl and aspartyl acids in the ligand-binding region of the LDL receptor. As shown by x-ray crystallography, the amino-terminal domain of apo-E (residues 1 to 191) forms a four-helix bundle (Fig. 34-12). The fourth helix encompasses residues 130 to 165, the area envisioned to contain the receptor-binding region. The basic residues in the vicinity of amino acids 134 to 150 are oriented away from the surface of the molecule and are probably involved in the direct interaction of apo-E with the LDL receptor. 25 26

The identification of naturally occurring defects of apo-E that are defective in receptor binding has provided key insights into the specific residues involved (see Fig. 34-11). The most common variant that is defective in binding is apo-E2 (Arg-158 Cys). This substitution appears to impair receptor binding secondarily by altering the conformation of residues in the 136 to 150 region of apo-E. Other variants that are defective in binding involve single amino acid substitutions: Arg-136 Ser, Arg-142 Cys, Arg-145 Cys, and Lys-146 Gln or Glu. Site-directed mutagenesis showed that Arg-150 also plays a key role in receptor binding. A rare apo-E mutation, apo-E Leiden, involves a duplication of seven amino acids (residues 121 to 127) inserted in tandem at the junction between helices 3 and 4. This insertion probably disrupts receptor binding by altering the conformation of the 136 to 150 receptor-binding region.

Apo-E also binds to heparin and to heparan sulfate proteoglycans (HSPGs). As discussed later, binding of apo-E to HSPG is important in the clearance of remnant lipoproteins by the LDL pathway. Residues in the 136 to 150 region of apo-E are responsible for the ionic interaction with the sulfate groups of heparin-like molecules and for binding to the LRP.

Apolipoprotein E in Lipid Metabolism

Apo-E functions in two aspects of lipid and cholesterol transport. The first, involving chylomicron and VLDL metabolism, provides a global transport role for apo-E. The knockout of apo-E by gene targeting in mice results in marked hyperlipidemia and the development of severe atherosclerosis, confirming the importance of this protein in cholesterol homeostasis and lipid transport. The second aspect involves the redistribution of lipids (including cholesterol) among cells within a tissue or organ. This local transport role redistributes lipids from cells with excess cholesterol to those requiring cholesterol, phospholipids, and other lipids for repair, proliferation, or other purposes. This pathway may involve lipid-laden HDL and apo-E that can acquire tissue lipids or apo-ElpId complexes formed in the interstitial fluid. As stated previously, apo-E is synthesized and secreted by a variety of cells and is available in interstitial fluid to transport lipids. Cells requiring cholesterol up-regulate their LDL receptors, and apo-E targets the apo-E-containing HDL or lipid complexes to cells deficient in necessary lipids. For example, the local transport pathway for apo-E is involved in lipid redistribution within a liver after injury and during regeneration.

Apolipoprotein A-I

Apo-A-I is a 29-kDa protein encoded by a gene on the long arm of chromosome 11, part of a cluster that includes genes for apo-CII and apo-AIV. The apo-A-I gene is 1863 base pairs in length, and its mRNA encodes a 267-amino-acid protein that includes an 18-amino-acid prepeptide and a 6-amino-acid prepeptide. The prepeptide is cleaved extracellularly to yield the mature circulating form of 243 amino acids (Table 34-3).

Apo-A-I is synthesized by the human intestine and liver and is a constituent of chylomicrons and HDL. It binds to lipids of these lipoproteins, mainly through a series of 22-amino-acid amphipathic alpha helices separated by helix-breaking proline residues. The polar face of the amphipathic helix is exposed to the aqueous environment, whereas the nonpolar face binds to the lipid (primarily phospholipids) on the surface of the particle. There are eight complete 22-amino-acid amphipathic helices and two 11-amino-acid repeats in apo-A-I.

In addition to its role as a structural protein in HDL, apo-A-I activates lecinthin:cholesterol acyltransferase (LCAT), which esterifies free cholesterol on HDL particles. It may facilitate the interaction of LCAT with phosphatidylcholine, the substrate of LCAT, and activate the enzyme. The specific regions of apo-A-I involved in LCAT activation have been identified, including the amino acids responsible for catalytic activity. Other apolipoproteins, such as apo-AIV and apo-CI, which have similar lipid-binding properties, can also activate LCAT (discussed in detail in a later section).

Apolipoprotein A-II

Apolipoprotein A-II is a 34-kDa protein encoded by a gene on the long arm of chromosome 1. The mRNA encodes a 100-amino-acid protein, but the mature circulating form of apo-A-II is 77 amino acids in length. In the plasma, human apo-A-II exists primarily as a homodimer. A cysteine at residue 6 of apo-A-II forms a disulfide bond with a second apo-A-II molecule. Heterodimers of apo-A-I and apo-E occur only in persons with apo-E2 and apo-E3, which possess free cysteine residues. Heterodimer formation interferes with the function of the individual proteins.
with the ability of apo-E to bind to the LDL receptor.

Apo-AII is synthesized primarily in the liver. It is found together with apo-AI on a subfraction of HDL referred to as LpAI/AII particles. Apo-AII may play a role in the activation of hepatic lipase and the inhibition of LCAT. The genetic absence of apo-AII in two sisters did not produce any obvious phenotypic effects and did not cause low HDL levels.

Overexpression of apo-AII in mice leads to an increased susceptibility to atherosclerosis, possibly because apo-AII displaces apo-AI from HDL. This could interfere with the normal ability of apo-AI-containing HDL to transport cellular cholesterol to the liver for excretion. Therefore, apo-AII is considered a proatherogenic apolipoprotein.

C Apolipoproteins

The genes for apo-CI and apo-CII reside on chromosome 19 near the gene that encodes apo-E, whereas the apo-CIII gene is part of the apo-AI and apo-AIV gene cluster on chromosome 11 (for a review, see references 8 and 45). The C apolipoproteins (see Table 34-3 and Fig. 34-6) readily exchange among various lipoproteins and are synthesized primarily by the liver. (Apo-CI is also produced by macrophages and, in small amounts, by the intestine.) HDLs appear to serve as a reservoir for the C apolipoproteins, which can then be transferred to triglyceride-rich lipoproteins. The C apolipoproteins appear to regulate triglyceride metabolism and to influence the inverse relation between triglyceride levels and HDL cholesterol (HDL-C). Apo-CI (6.6 kDa) modulates the uptake of triglyceride-rich lipoprotein (chylomicron remnants, VLDL, and IDL) by interfering with the ability of apo-E to mediate binding to lipoprotein receptor pathways. Similarly, apo-CIII (8.8 kDa) may prevent the normal interaction of triglyceride-rich, apo-E-containing lipoproteins with receptors and cell-surface HSPG. Apo-CI and apo-CIII may displace apo-E from the particles. Apo-CII (8.9 kDa) is a cofactor for LPL, and mutations in the apo-CII gene result in a marked hypertriglyceridemia (discussed later).

Overexpression of apo-CI, apo-CII, or apo-CIII in transgenic mice results in hypertriglyceridemia (for a review, see references 45 and 110). In the case of apo-CI and apo-CIII, the resulting hyperlipidemia appears to be caused by displacement of apo-E from triglyceride-rich particles, which results in impaired receptor-mediated uptake, displacement of apo-CII, and impaired lipolytic processing. A polymorphism of the apo-CIII gene promoter region in mice is also associated with increased levels of apo-CIII and hypertriglyceridemia.

The hypertriglyceridemia that follows the overexpression of apo-CII was initially puzzling because apo-CII is a cofactor that activates LPL-mediated hydrolysis of triglycerides. However, the triglyceride-rich lipoproteins that accumulate in the plasma are poor in apo-E and do not interact well enough with cell-surface HSPG to allow lipase activity to occur or with receptors in the proteoglycan-rich matrices of the cell surface to allow uptake. Therefore, either overproduction or underproduction of apo-CII can cause hypertriglyceridemia.
Lipoprotein Receptors Controlling Lipoprotein Metabolism

Low-Density Lipoprotein Receptor Gene Family

Mammalian members of this gene family, in addition to the LDL receptor itself, include the LRP, the glycoprotein 330 (gp330)/megalin receptor, the VLDL receptor, and the apo-E receptor. Non-mammalian members include the chicken vitellogenin receptor, a Caenorhabditis elegans receptor, and the Y1 protein in the fruit fly Drosophila melanogaster. These receptors share common structural motifs, including a single transmembrane domain, a short cytoplasmic tail, and an extracellular ligand-binding domain that contains various numbers of cysteine-rich repeats of approximately 40 amino acids each. It has become apparent that the functions of this gene family extend far beyond mediating lipid uptake by cells and include serving as transducers of extracellular signals involved in normal brain development.

Low-Density Lipoprotein Receptor

The LDL receptor, a glycoprotein with an apparent molecular weight of 160,000, is expressed on the surface of most cells and especially in liver. It functions in the uptake of apo-B and apo-E-containing lipoproteins, including LDL, chylomicron remnants, VLDL, VLDL remnants, IDL, and HDL. Most HDL particles lack apo-E and do not interact with the LDL receptor. Cells can acquire cholesterol from the plasma by taking up these lipoproteins through the LDL receptor. The LDL receptor gene was first identified in 1973, and its gene was characterized in 1985 in the laboratory of Nobel laureates Joseph L. Goldstein and Michael S. Brown. Two proteins on the lipoprotein surface, apo-B100 and apo-E, bind to the LDL receptor, which for this reason is sometimes referred to as the apo-B100/apo-E receptor.

After the lipoprotein binds to the LDL receptor, the resulting complex becomes localized to a specialized area of the cell membrane called a coated pit. The “coat” contains a protein complex called clathrin, which clusters the receptors in a region of the cell membrane that can invaginate and form an intracellular vesicle to contain the lipoprotein. As these internalized vesicles, or endosomes, move into the cytoplasm, the internal environment becomes progressively more acidic, causing the receptor and the lipoprotein to dissociate. The lipoproteins are degraded in the lysosomes, and the unoccupied receptors recycle to the cell surface.

The LDL receptor is synthesized in the ER as a protein of 839 amino acids with an apparent molecular weight of 120,000. Glycosylation of the protein in the ER and in the Golgi apparatus increases its weight to about 160,000. The LDL receptor has five distinct structural and functional domains. Mutations within these domains disrupt the normal function of the receptor in lipoprotein metabolism and cause the genetic disorder familial hypercholesterolemia (FH).

Ligand-Binding Domain

The ligand-binding domain of the LDL receptor consists of the 292 amino acids at the amino terminus. This region of the molecule is rich in cysteines and contains glutamic and aspartic acids that mediate the binding to apo-B and apo-E. It is composed of seven repeats of approximately 40 amino acids each. Each repeat contains six cysteines that form three intrarepeat disulfide bonds, resulting in a very stable structure. In addition, each repeat contains a Ser-Asp-Glu triple that mediates the interaction of apo-B and apo-E-containing lipoproteins with the LDL receptor. The ligand-receptor binding is an ionic interaction between positively charged arginines and lysines in apo-B100 and apo-E and negatively charged aspartic and glutamic acids in the ligand-binding domain of the LDL receptor.

Site-directed mutagenesis and analysis of naturally occurring mutants of the LDL receptor associated with FH have provided insights into the roles of specific repeats and residues in ligand binding. Ligand-binding domain repeat 1 does not play a major role in the binding of either apo-B-containing (LDL) or apo-E-containing (VLDL) lipoproteins. The deletion of repeats 2 through 7, however, markedly impairs the binding of LDL. The binding of -VLDL is mediated by apo-E and is impaired only if repeat 5 is deleted. Therefore, the requirements for the binding of LDL are more stringent than those for the binding of -VLDL.

Single amino acid substitutions of critical residues in the ligand-binding repeats also impair binding activity. For example, in patients with FH Puerto Rico, in which the second of the ligand-binding triplet (Ser-Asp-Glu) in repeat 4 is changed to a leucine, the LDL fails to bind, although apo-E-containing -VLDL binds with near-normal affinity. In FH Heterozygote, in which lysine is substituted for the glutamic acid of the ligand-binding triplet of repeat 5, neither LDL nor -VLDL binds normally.

The defect in the Watanabe heritable hyperlipidemic (WHHL) rabbit with hypercholesterolemia and accelerated atherosclerosis involves a deletion of four amino acids in repeat 4. Although this defect is associated with a reduced number of receptors reaching the cell surface, those that do reach the surface retain the ability to bind -VLDL (apo-E) but not LDL (apo-B).

As demonstrated in the WHHL rabbit, mutations in the ligand-binding domain can also disrupt the normal transport of the LDL receptor to the cell surface. The decreased transport of the receptor from the ER to the Golgi apparatus and to the cell surface is undoubtedly caused by improper folding of the molecule and increased intracellular degradation. For example, FH Afrikaner, which is caused by the presence of a glutamic acid rather than aspartic acid in the triplet of repeat 5, results in defective transport and lack of normal expression of the receptor on the cell surface.

Epidermal Growth Factor Precursor Homology Domain

This region is composed of 400 amino acids and is about 33% identical to the sequence of the human EGF precursor. It contains three cysteine-rich repeats (A, B, and C), each approximately 40 amino acids in length. The repeats are not homologous to the 40-amino-acid repeats of the ligand-binding domain but...
are related to the EGF. Repeat A is involved in the binding of LDL, and its deletion markedly inhibits LDL binding (-VLDL binding is retained). The EGF precursor domain also plays a role in allowing the LDL receptor or receptors to dissociate from the lipoproteins and recycle to the cell surface. Detection of the EGF precursor homology domain allows normal binding and internalization of -VLDL but prevents dissociation of the ligand and receptor. As a result, receptors do not recycle to the cell surface. The role of the EGF precursor domain was established by site-directed mutagenesis studies, and FH Osaka was subsequently found to have the same deletion.

**O-Linked Sugar Domain**

This domain is composed of 58 amino acids, primarily serines and threonines, many of which are sites for the attachment of O-linked carbohydrate chains. No functional role for this domain has been described, and its deletion has no functional consequences.

**Membrane-Spanning Domain**

The 22 amino acids in this domain are hydrophobic and serve to anchor the receptor within the plasma membrane. Truncation mutations that exclude this region are characteristic by secretion of the receptor from the cell so that lipoproteins are not internalized.

**Cytoplasmic Domain**

The carboxyl-terminal region of the LDL receptor is composed of 50 amino acids and contains the sequence NPXY (N, asparagine; P, proline; X, any amino acid; Y, tyrosine), which is responsible for clustering the receptors in coated pits and mediating internalization of the receptors by the cells. One of the early mutations associated with FH (J.D. allele, FH Ban) provided insights into the role of a critical residue for directing internalization. In the mutant form of the receptor, tyrosine-807 is changed to a cysteine. Site-directed mutagenesis demonstrated that this position must be occupied by an aromatic amino acid (tyrosine, phenylalanine, or tryptophan) for normal internalization. The tetrameric sequence Asn-Pro-Val-Tyr, in which tyrosine-807 occurs, is the signal directing the receptors to the coated pit.

**Regulation of the Low-Density Lipoprotein Receptor Gene**

The LDL receptor gene is 45 kilobases in length and is located on the distal portion of the short arm of chromosome 19. Synthesis of the LDL receptor is regulated by deoxyribonucleic acid (DNA) sequences in the 5'-flanking region of the LDL receptor gene. A sequence of approximately 10 bases in this region, called the sterol regulatory element (SRE), and two other repeats that bind the transcription factor Sp1 are necessary for the regulation of the LDL receptor mRNA levels. If intracellular sterol levels are high, LDL receptor mRNA is not transcribed. When the sterol content of the cells decreases, the expression of LDL receptors on the cell surface increases, causing increased uptake of apo-B-100-containing lipoproteins and increased delivery of cholesterol to the cells. The LDL receptor gene "senses" the sterol level of the cell and appropriately controls receptor mRNA production and protein biosynthesis to meet the needs of the cell.

The control mechanism of LDL receptor expression has been elucidated in considerable detail. Currently, there are three structurally related transcription factors, SRE-binding proteins (SREBPs) 1α, 1c, and 2, which regulate the level of LDL receptors and other genes encoding enzymes involved in the biosynthesis of cholesterol, unsaturated fatty acids, and triglycerides. SREBP-1a and SREBP-1c arise from the same gene but use different promoters and have different first introns; SREBP-2 arises from a separate gene. The intact 125-kDa SREBPs are three-domain integral membrane proteins containing two membrane-spanning regions. The amino-terminal domain of SREBPs represents transcription factors of the loop-helix leucine zipper family and contains sequences that recognize the SREs on the genes that they control.

To become active transcription factors, the intact SREBPs must be cleaved in the correct order by two proteases in a post-ER compartment and then translocated to the nucleus to interact with the SREs. The first protease, designated site-1 protease (S1P), cleaves the loop connecting the amino-terminal and carboxyl-terminal domains, both of which remain membrane bound after cleavage. The second protease, site-2 protease (S2P), further cleaves the amino-terminal domain just within the first membrane-spanning region, releasing the transcription factor to enter the nucleus and interact with the SREs. Sterol control is exerted through a two-domain regulatory protein, SREBP cleavage activating protein (SCAP), that is required for S1P cleavage of the SREBP. SCAP is membrane associated (eight transmembrane regions) and is tightly complexed to the SREBPs through its carboxyl-terminal domain. Five of the eight membrane-spanning segments serve as a sterol-sensing domain. It is not clear whether the sensing domain interacts directly or indirectly with sterols. What is known is that sterols regulate the ability of SCAP to transport SREBPs to the post-ER compartment where S1P is located. SCAP cycles between the ER and Golgi apparatus, and whether SCAP transports the SREBPs to the S1P compartment is dependent on the processing of its N-linked carbohydrates by the Golgi apparatus. In sterol-depleted cells, SCAP cycles to the Golgi apparatus and its N-linked carbohydrates are modified; the modified SCAP returns to the ER to transport the SREBPs. Sterol blocks the movement of SCAP from the ER to the Golgi apparatus, preventing carbohydrate modification and the ability of SCAP to transport the SREBPs for S1P cleavage.

**Low-Density Lipoprotein Receptor-Related Protein**

The LRP is an integral membrane receptor composed of two components: a 515-kDa amino-terminal extracellular domain and an 85-kDa cytoplasmic and membrane-spanning domain (the precursor protein, composed of 4525 amino acids, is cleaved after synthesis). This large protein is equivalent structurally to approximately four LDL receptors and possesses 31 ligand-binding domains. The LRP contains the four structural motifs characteristic of other members of the LDL receptor gene family: multiple ligand-binding repeats, EGF repeats and EGF precursor homology domains, a single membrane-spanning region, and two NPXY internalization signals. The LRP is expressed primarily in liver (parenchymal cells), brain (neurons), and placenta (syncytiotrophoblast cells).

The LRP interacts with approximately 18 ligands and has several functions. With respect to lipoprotein metabolism, the LRP binds with high affinity to apo-E-rich chylomicron remnants and VLDL remnants and internalizes them. Interaction of these lipoproteins with the LRP requires the addition of multiple apo-E molecules per particle, which serves as ligands. Initial binding of the lipoprotein to cell-surface HSGLP is necessary to facilitate the interaction or transfer of the apo-E-rich remnants to the LRP (discussed further in "Chylomicron Remnant Receptors in Remnant Catabolism"). The LRP does not bind LDL.

The LRP can also interact with LPL and hepatic lipase. This interaction could mediate the hepatic binding and uptake of remnant lipoproteins possessing these enzymes on their surface. Other ligands for the LRP that are not directly related to lipid metabolism include 2-macroglobulin, plasminogen activators and inhibitors, and bacterial toxins. Knockout of the LRP in mice is lethal, demonstrating its critical importance, but the reason for the lethality remains to be elucidated.

A receptor-associated protein (RAP) of 39 kDa can be isolated along with purified LRP and effectively competes with all the ligands for the LRP binding. This protein also binds to the gp330 and VLDL receptors (described later) and blocks ligand binding to these receptors as well. However, RAP does not appear to be secreted from the cells, and it may serve as an intracellular chaperone that occupies the ligand-binding sites for transport of the LRP to the cell surface. The knockout...
of RAP by gene targeting in mice markedly reduces the expression of LRP in both liver and brain, further suggesting an intracellular transport role for this protein. Alternatively, it may participate in the intracellular recycling of the receptors. Regardless of its physiologic role, RAP inhibits the interaction of LRP and its ligands both in cultured cells and in intact animals.

Glycoprotein 330

The gp330/megalin receptor, also referred to as the major Heymann nephritis antigen, is a large protein (about 600 kDa) that possesses many of the structural motifs of the LDL receptor. It is expressed in the proximal tubules of the kidney and the ependymal cells in the brain and is not present in liver. Although gp330 binds apo-E containing lipoproteins and LDL, its role in lipoprotein metabolism is unknown. The knockout of gp330 by gene targeting does not have an obvious effect on lipoprotein metabolism, but it causes developmental abnormalities of the central nervous system (holoprosencephaly).

Very-Low-Density Lipoprotein Receptor

The VLDL receptor closely resembles the LDL receptor except that it has an eighth ligand-binding repeat. The VLDL receptor (about 130 kDa) binds apo-E containing lipoproteins and is present primarily in muscle, fat, and brain. In the nervous system, it is present in the choroid plexus and in some neurons. It is absent from liver, and its role in lipoprotein metabolism remains to be determined. It has been suggested, because the receptor is present in tissues that metabolize VLDL-derived fatty acids, that it may function to deliver triglyceride-rich lipoproteins to target tissues.

Apolipoprotein E Receptor 2

The apo-E receptor 2 (106 kDa) is the newest member of the LDL receptor family to be described; it is expressed primarily in the brain and to a lesser extent in the placenta and can be expressed as various splice variants. Although this receptor, like the LDL receptor, contains seven cysteine-rich repeats in the ligand-binding domain, the repeats are more closely related structurally to the VLDL receptor. Because the receptor is primarily expressed in the brain, it is likely to play a role in lipoprotein metabolism in the central nervous system. In addition to its roles in lipoprotein metabolism, the apo-E receptor 2 and the VLDL receptor have been implicated in normal brain development by transducing extracellular signals.

Scavenger Receptors

Originally, it was thought that a single scavenger receptor existed on macrophages. Also known as the acetyl-LDL receptor, this receptor was characterized by its ability to interact with chemically modified LDL but not with native LDL. LDL particles that had been modified by acetylation, acetoacetylation, or malondialdehyde were taken up by high-affinity cell-surface receptors on macrophages, resulting in marked cholesterol accumulation. As a result of cloning efforts, it became apparent that the scavenger receptor actually represented a large family of receptors with specificities for a broad range of unrelated ligands and involvement in a spectrum of physiologic processes, including atherosclerosis, host defense, and central nervous system disorders.

Currently, there are five subclasses (A to E) of the scavenger receptor family. Class A receptors include types I, II, and III and MARCO. The type I and type II receptors are generated by alternative splicing of the mRNA encoded by a gene on chromosome 8. The predicted structure is that of a trimer (220 kDa) composed of three identical subunits (each about 77 kDa). The type I receptor contains six domains: a cytoplasmic, amino-terminal domain (50 amino acids); a transmembrane domain (26 amino acids); a spacer (74 amino acids); an alpha-helical coiled-coil domain (121 amino acids); a collagen-like domain (72 amino acids with a Gly-X-Tyr repeat); and a cysteine-rich domain (110 amino acids). The type II scavenger receptor is identical to the type I receptor except that it lacks the carboxyl-terminal cysteine-rich domain; its collagen-like domain is responsible for ligand binding. Clusters of positively charged residues (lysines) appear to mediate the interaction with the chemically modified lipoproteins (see section on the LDL receptor and oxidized lipids for a discussion of the role of the scavenger receptor in atherogenesis). In addition to binding acetylated LDL, class A scavenger receptors bind anionic proteins, polynucleotides, and bacterial endotoxins (lipopolysaccharides). Their main function appears to involve the clearance of microbial pathogens, senescent cells, and altered lipoproteins.

Class B scavenger receptors include CD36 and murine SR-BI (human homologue CLA-1). These receptors possess two membrane-spanning regions and bind both oxidized and native lipoproteins. CD36 is expressed on the surface of platelets, capillary endothelial cells, adipose cells, circulating monocytes, and other cell types. The role of the SR-BI as a receptor for HDL and its involvement in reverse cholesterol transport are discussed in the section "Transport Facilitated by a Cell-Surface Binding Protein." CD36 is also implicated in platelet adhesion and aggregation, phagocytosis of apoptotic cells, and clearance of Plasmodium falciparum-infected cells for a review, see reference 112).

Class C is represented by a single member, SR-C from D. melanogaster. It contains domains that are homologous to the vertebrate complement control protein and a mucin-like domain. Class D and E scavenger receptors are also represented by single members, Lox-1 and endothelial scavenger receptor, respectively. Lox-1 is characterized by a C-type lectin structure and the endothelial scavenger receptor by multiple EGF repeats. Regardless of the class, all scavenger receptors share the common property of binding oxidized or modified LDLs, or both.
Enzymes and Transfer Proteins Involved in Lipid and Lipoprotein Metabolism

Lipoprotein Lipase

Human LPL is a protein composed of 448 amino acids (approximately 50 kDa). It is synthesized by adipocytes, by myocytes in skeletal and cardiac muscle, and by macrophages but is not produced by hepatocytes. After secretion from adipocytes and myocytes, LPL is transported to the surface of capillary endothelial cells of these tissues, where it attaches to HSPG and interacts with chylomicrons and VLDL in the circulation and mediates the hydrolysis of their triglycerides to release free fatty acids for use by the tissues. The fatty acids are stored as triglyceride in adipocytes and used as a source of energy in muscle and for triglyceride synthesis in the formation of hepatic VLDL.

The active form of LPL is a dimer. Although its crystal structure is not known, LPL has a high degree of homology with another serine esterase, pancreatic lipase, whose structure is known. Based on similarities between LPL and pancreatic lipase, a model for LPL function has been suggested (Fig. 34-16), and five functional domains have been identified in LPL on the basis of structural and mutational studies.

Contrasting Lipoprotein Lipase and Hepatic Lipase

"Metabolic Pathways Involving High-Density Lipoproteins."

Hepatic Lipase

Hepatic lipase has several roles in lipoprotein metabolism. First, it hydrolyzes triglycerides and possibly excess surface phospholipids in the final processing of chylomicron remnants. As suggested, this enzyme may be active in the space of Disse. It binds heparan sulfate and facilitates the interaction of remnant lipoproteins with the LRP, thereby delivering these lipoproteins to the receptor for internalization by hepatocytes. Second, it completes the processing of IDL to LDL (discussed in the section on IDL). Third, it participates in the conversion of HDL to HDL₂ by the removal of triglyceride and phospholipid from HDL₁ (discussed in "Metabolic Pathways Involving High-Density Lipoproteins"). High levels of hepatic lipase activity decrease total HDL levels.

LPL requires apo-C-II as a cofactor to stimulate its catalytic activity, but apo-C-II is not a cofactor for hepatic lipase. In contrast, apo-E may facilitate both triglyceride binding and uptake of lipoproteins associated with the enzyme.

Heparin-Binding Site

The heparin-binding site mediates the interaction of LPL with HSPG on endothelial cells. Clusters of positively charged arginines and lysines on one face of LPL, particularly those in the carboxyl terminus, appear to mediate this interaction.

Lipid-Binding Site

The domain of the protein that allows the enzyme to interact with the surface of the chylomicron lies in the carboxyl terminus, particularly around residues 245 to 253.

Apolipoprotein CII-binding Site

Apo-C-II, an essential cofactor for LPL, binds to the carboxyl terminus at a site that has not been identified precisely.

Catalytic Site

This site mediates the hydrolysis of triglycerides, primarily to fatty acids and monoglyceride, and is postulated to involve serine-132, aspartic acid-156, and histidine-241, which are at the bottom of a hydrophobic channel that is covered by a flap or catalytic lid. The lid may mediate the interaction with the lipid substrate by assuming an open or closed conformation. LPL is a serine esterase with triglyceride hydrolase activity and, to a lesser extent, phospholipase activity.

LRP-Binding Site

The LRP-binding site is distinct from the heparin-binding site and involves the carboxyl-terminal domain. Through its interaction with the LRP, LPL can facilitate the binding and uptake of lipoproteins associated with the enzyme.

LPL is synthesized by adipocytes and is present primarily on liver endothelial cells and on HSPG in the space of Disse. Hepatic lipase is transported from the liver to the capillary endothelium of the adrenals, ovaries, and testes, where it functions in the release of lipids from lipoproteins for use in these organs. Its activity is increased by androgens and reduced by estrogens. Little is known about the structural domains of hepatic lipase except by analogy to similar domains within LPL, but the catalytic triad includes serine-145, aspartic acid-171, and histidine-256.

Hepatic lipase has several roles in lipoprotein metabolism. First, it hydrolyzes triglycerides and possibly excess surface phospholipids in the final processing of chylomicron remnants. As suggested, this enzyme may be active in the space of Disse. It binds heparan sulfate and facilitates the interaction of remnant lipoproteins with the LRP, thereby delivering these lipoproteins to the receptor for internalization by hepatocytes. Second, it completes the processing of IDL to LDL (discussed in the section on IDL). Third, it participates in the conversion of HDL to HDL₂ by the removal of triglyceride and phospholipid from HDL₁ (discussed in "Metabolic Pathways Involving High-Density Lipoproteins"). High levels of hepatic lipase activity decrease total HDL levels.

Contrasting Lipoprotein Lipase and Hepatic Lipase

LPL requires apo-C-II as a cofactor to stimulate its catalytic activity, but apo-C-II is not a cofactor for hepatic lipase. In contrast, apo-E may facilitate both triglyceride binding and phospholipid hydrolysis by hepatic lipase and may be a cofactor for its enzymatic activity. In other respects, the enzymes are similar. After intravenous injection of heparin, both enzymes are released from endothelial cells of the liver and peripheral tissues and are referred to as postheparin lipase. Therefore, measurements of total plasma lipolytic activity after heparin injection reflect the activities of both enzymes.

Mutations that impair or inactivate LPL cause hypertriglyceridemia (discussed later). Likewise, deficiency of apo-C-II prevents normal activation of LPL and also causes hypochondylidemia. Hepatic lipase deficiencies result in a variable and diverse pattern of lipoprotein changes, including the accumulation of remnant lipoproteins, IDL, and LDL. These changes are predictable on the basis of the functional roles of hepatic lipase. Knockout of the LPL gene in mice causes a particularly severe hypertriglyceridemia that becomes evident as soon as the newborns begin to suckle and causes death within the first 24 hours. On the other hand, knockout of the hepatic lipase gene causes less severe manifestations, including changes in HDL and increased plasma phospholipid levels. In the mouse, LPL may take on some of the functions subserved by hepatic lipase in other species. Overexpression of human hepatic lipase in transgenic mice markedly decreases HDL and IDL.

Lecithin:Cholesterol Acyltransferase

LCAT circulates in association with HDL in the plasma and functions to esterify free cholesterol. In humans, most of the cholesteryl esters in plasma lipoproteins are formed by the action of LCAT. The major substrate for LCAT is the small HDL particle; to a lesser extent, LDL is also a substrate. The enzyme catalyzes the transfer of long-chain fatty acids from phosphatidylcholine (linoleic acid at position 2 of lecithin preferred) to the hydroxyl group at position 3 on cholesterol. The
structure and function of LCAT are discussed more thoroughly in the context of HDL metabolism.

Cholesteryl Ester Transfer Protein

The cholesteryl ester transfer protein (CETP) transfers cholesteryl esters from the larger HDL to VLDL, IDL, and remnant lipoproteins. In return, triglyceride from these lipoproteins is transferred to HDL. LCAT and CETP function in concert in HDL metabolism, and the structure and function of CETP are discussed further in the section on HDL.
PLASMA LIPOPROTEINS: STRUCTURE, FUNCTION, AND METABOLISM

Chylomicrons

Characteristics

Chylomicrons (density [d] < 0.95 g/mL) are the largest of the plasma lipoproteins (> 1000 Å in diameter) and readily float after ultracentrifugation of plasma. They are composed of 98% to 99% lipid (85% to 90% triglyceride) and 1% to 2% protein (see Table 34.4). Chylomicrons are present in post-prandial plasma (but absent after an overnight fast) and contain several apolipoproteins, including apo-B48, apo-Al, apo-AIV, apo-E, and the C apolipoproteins (see Fig. 34-6). The distinctive apolipoprotein is apo-B48, a form of apo-B that has an apparent molecular mass 48% that of apo-B100. Because it is the only form of apo-B synthesized by the intestine, apo-B48 is a marker for human lipoproteins produced by the intestinal epithelium.

Origin

Chylomicrons are produced by the epithelial cells of the small intestine (duodenum and proximal jejunum) when dietary fat and cholesterol are presented to the brush border of the epithelial cell membranes as bile acid micelles. Free fatty acids and monoglycerides taken up by the intestinal epithelial cells are synthesized into triglycerides in the ER in the apical region of the intestinal cells. Triglycerides, phospholipids, and cholesterol (absorbed or synthesized by the intestinal cells) are used for chylomicron formation in the Golgi apparatus, where some of the apolipoproteins undergo final carbohydrate processing, and the chylomicrons are secreted into the space along the lateral borders of the intestinal cells. From there, they enter the mesenteric lymph and proceed through the thoracic duct lymph to the general circulation. Newly synthesized chylomicrons possess apo-B48, apo-Al, and apo-AIV (intestinally synthesized apolipoproteins); they acquire apo-E and C apolipoproteins in the lymph and blood, primarily from HDL.

Metabolic Fate

In the circulation, LPL catalyzes the release of free fatty acids from chylomicron triglycerides and converts them into triglyceride-poor, cholesterol-enriched chylomicron remnants (Fig. 34-17). The free fatty acids are taken up by various tissues to be stored as triglyceride, oxidized as an energy source, or reutilized in hepatic lipoprotein triglyceride synthesis. Hepatic lipase, acting primarily as a phospholipase and secondarily as a glyceride hydrolase, also plays a role in the final preparation of chylomicron remnants for uptake by hepatocytes. Chylomicron remnants are cleared rapidly from the plasma by the liver.

Sequestration of chylomicron remnants within the space of Disse (see Fig. 34-18) appears to involve binding of the remnant lipoproteins to HSPG mediated by apo-E (or possibly LPL or hepatic lipase). The microvilli-covered surface of hepatocytes is coated with HSPG, which is abundant in the space of Disse. HSPGs bind apo-B48 by an ionic interaction between negatively charged sulfate groups of HSPG and basic amino acids within the 136 to 150 region of apo-E. The absence of proteoglycans on the cell surface impairs uptake of the particles. Apo-E secreted by the hepatocytes appears to be bound to the cell-surface HSPG and further enhances the apo-E-mediated binding of remnant lipoproteins.

Chylomicron remnants may be further processed by lipases or other enzymes in the space of Disse. LPL is carried into the space of Disse on chylomicron remnants, and hepatic lipase produced by the liver may be localized there. These lipases facilitate the binding and uptake of remnants by interacting with the LRP.

The actual uptake of the particles by hepatocytes may involve two or more receptors (see Fig. 34-18), the LDL receptor, which interacts with lipoproteins containing apo-B100 and apo-E, and possibly a unique apo-E or chylomicron remnant receptor, now known to be the LRP.

Whereas remnant particles with LPL or hepatic lipase on their surfaces may interact by means of these molecules with the HSPG in the space of Disse and facilitate binding and uptake by the LRP, apo-E mediates interactions with HSPG and the LRP or the LDL receptor are critical in remnant metabolism. Patients with apo-E mutations that prevent interaction with HSPG or lipoprotein receptors develop hyperlipidemia characterized by remnant lipoprotein accumulation despite having normal lipase activity. In addition, knockout of apo-E by gene targeting in mice causes a massive accumulation of remnant lipoproteins.
or chylomicron remnants in patients with absent or defective LDL receptors could reflect the fact that remnant clearance requires several steps, as described. For example, sequestration of the particles in the space of Disse (HSPG binding) is normal in patients with defective LDL receptors and could prevent the accumulation of remnants in plasma. Furthermore, the HSPG/LRP pathway can compensate for deficiency of LDL receptors. Both receptors probably function in the uptake of the remnants, and in the absence of one the other continues to function.

### Chylomicron Remnant Receptors in Remnant Catabolism

Evidence suggests that the LRP is the chylomicron remnant (apo-E) receptor, which belongs to the LDL receptor gene family (discussed previously). As noted earlier, the LRP binds with high affinity to apo-E-enriched lipoproteins but does not bind LDL to a significant extent. Apo-E must be added to remnant lipoproteins before they bind to the LRP with high affinity. Apo-E exists in the space of Disse in high concentration, probably because it is secreted by hepatocytes and binds to HSPG in the space of Disse. The HSPG may serve as a reservoir for apo-E, allowing enrichment of the remnants with this apolipoprotein. These and other observations have led to the hypothesis that apo-E functions in a process called secretion-capture. It is envisioned that apo-E combines with lipids or lipoproteins and directs them to cells expressing LDL receptors or the LRP. In the liver, the LRP and apo-E could interact in this way to capture chylomicron remnants. The LRP is also present in other tissues, including brain, and may function locally in the uptake of lipids. The secretion-capture role of apo-E functions in peripheral nerve injury and repair and in the normal maintenance of neurons.

As already stated, LRP-mediated uptake of remnants requires the initial interaction of apo-E-containing lipoproteins with cell-surface HSPG. HSPGs are hydrolyzed by treating cells with heparinase in vitro or by infusing heparinase into the portal vein of mice, apo-E-rich remnants do not bind to the cell surface and do not interact with the LRP even though the receptor is present. After the lipoproteins interact with HSPG, the remnants may be transferred to the LRP for internalization by the cells, or the HSPG/LRP complex may be internalized. This two-step process involving cell-surface proteoglycans and receptors is referred to as the HSPG/LRP pathway, a similar two-step process has also been described for growth factors. It is also possible that HSPG alone can mediate remnant uptake directly without the LRP.

The LRP interacts not only with apo-E-containing lipoproteins but also with an unrelated protein, 2-macroglobulin, a broad-spectrum endopeptidase inhibitor that is involved in clearing of proteases from the plasma by the liver. The binding of protease activates 2-macroglobulin, which then binds to the LRP. Activated 2-macroglobulin competes with remnants for binding to the LRP. Radiolabeled chylomicron remnants are rapidly cleared from the plasma of mice, but the injection of 2-macroglobulin along with the chylomicron remnant impairs remnant clearance. This finding indicates that 2-macroglobulin and the remnants are binding to the same receptor, the LRP.

Herz and colleagues demonstrated that the LRP is the chylomicron remnant (apo-E) receptor and documented its importance in remnant catabolism. These studies used RAP, which blocks the interaction of all ligands with the LRP, to demonstrate the role of the LRP in remnant clearance in mice. Knockout of RAP in mice results in a loss of LRP expression in the liver. Double-knockout mice, in which both RAP and the LDL receptor are missing, develop hyperlipidemia characterized by the accumulation of remnant lipoproteins in the plasma.

In summary, the catabolism of chylomicron remnants involves several steps and several components: sequestration, further lipolytic processing, and receptor-mediated endocytosis utilizing both the LDL receptor pathway and the HSPG/LRP pathway.
Very-Low-Density Lipoproteins

Characteristics
VLDLs are particles 300 to 700 Å in diameter that float on ultracentrifugation at a density of less than 1.006 g/mL (see Table 34-4). They are composed of 85% to 90% lipid (about 55% triglyceride, 20% cholesterol, and 15% phospholipid) and 10% to 15% protein. The distinctive apolipoprotein is apo-B100, the hepatic form of apo-B. VLDLs also contain apo-E and C apolipoproteins (see Fig. 34-8). VLDLs have pre- or pre-electrophoretic mobility and were previously called pre-lipoproteins.

Origin
VLDLs are synthesized by the liver, and their production is stimulated by increased delivery of free fatty acids to the hepatocytes, either from a high intake of dietary fat or from the mobilization of fatty acids from adipose tissue with fasting or uncontrolled diabetes mellitus. Triglycerides and phospholipids to be used in the formation of VLDL are synthesized in the liver, whereas VLDL cholesterol can be synthesized de novo or remobilized from LDL cholesterol (LDL-C). The VLDL particles are first visible at the junction of the rough ER and the smooth ER (transitional elements) before they enter the Golgi apparatus. Several of the apolipoproteins undergo carbohydrate processing within the Golgi apparatus. Large Golgi secretory vesicles migrate to the brush border surface of the hepatocytes, fuse with the plasma membrane, and release the VLDL particles into the space of Disse, where they enter the plasma (Fig. 34-20). The major protein constituents of the newly synthesized VLDLs are apo-B100, apo-E, and small amounts of the C apolipoproteins. In plasma, VLDLs acquire additional C apolipoproteins and apo-E, primarily from HDL.

Control of Very-Low-Density Lipoprotein Secretion Rate
The quantity of VLDL secreted from the liver is not controlled by changes in apo-B100 mRNA levels. Apo-B100 is constitutively expressed and is not highly variable. Newly synthesized apo-B100 is subject to two fates: (1) It can be combined with lipid to form VLDL particles, or (2) it can be degraded, in which case a VLDL particle is not secreted. If there is a stimulus for VLDL production, such as the delivery of free fatty acids to the liver, the balance is shifted away from apo-B100 degradation to the formation and secretion of apo-B100 containing VLDL.

Biosynthesis of Very-Low-Density Lipoproteins
Newly synthesized apo-B100 is translocated across the rough ER membrane. If not sufficiently lipidated as it is translocated, apo-B100 is destined to be degraded. If sufficient lipid is available, the apo-B100 binds the lipid as it enters the ER and forms triglyceride-rich particles. These particles can increase in size, enter the secretory pathway, and exit from the cell as mature VLDLs (see Fig. 34-20). MTP, phospholipid transfer protein (PLTP), and additional triglyceride, cholesterol, and phospholipid may be added as the VLDL precursor passes through the lumen of the rough ER. At the junction of the rough and smooth ER, lipid-rich particles lacking apo-B have been identified in rat liver, and these particles may fuse with the apo-B containing VLDL precursors to form the mature particle. However, because rat liver synthesizes both apo-B100 and apo-B48 containing VLDLs, the fusion step may apply only to apo-B48 VLDL and may be more relevant to apo-B48 containing chylomicron synthesis by the intestine.

Microsomal Triglyceride Transfer Protein
MTP is produced in the liver at sites where apo-B100 containing VLDLs are synthesized and in the intestine at sites where apo-B48 containing chylomicrons are synthesized. MTP (97 kDa) occurs as a heterodimer complex with the 58-kDa protein disulfide isomerase, an association required for MTP activity. Protein disulfide isomerase reshuffles disulfide bonds of cysteine residues and therefore may play a role in altering the conformation of apo-B for lipidation. In addition to transferring triglycerides to these lipoprotein particles, MTP transfers cholesterol esters and phospholipids.

More than a dozen mutations of MTP interfere with its activity. Defective MTP is responsible for the lipid disorder abetalipoproteinemia, a condition in which patients essentially lack apo-B containing lipoproteinemia in plasma. Therefore, MTP is critical for the biosynthesis of both apo-B100 VLDLs in the liver and apo-B48 chylomicrons in the intestine.

The luminal surfaces of the hepatocytes express LDL receptors and the LRP, and VLDLs possess both apo-B100 and apo-E that can react with these receptors. How then do VLDLs traverse the space of Disse and enter the blood? First, lipids such as phosphatidylethanolamine on the surface of the newly secreted VLDLs may alter the reactivity of the lipoproteins with the receptors. Newly secreted VLDLs are rich in phosphatidylethanolamine, but VLDLs in the circulation are poor in phosphatidylethanolamine. This phospholipid may prevent the particle from interacting with the receptors (i.e., specific lipids may mask the receptor-binding domains of apo-B100 and apo-E).

Second, other apolipoproteins may mask the receptor-binding domains of the VLDL apo-B100 and apo-E. Although they are present in small amounts on newly secreted VLDLs, the C apolipoproteins may alter the conformation or availability of apolipoproteins that interact with lipidoprotein receptors. Specifically, when VLDLs are formed in the liver and acquire apo-E, the C apolipoproteins may be positioned so as to mask the apo-E and thereby block its ability to react with the receptor or with proteoglycans in the space of Disse. Alternatively, apo-E associated with the particles intracellularly may not be available to bind to the receptors; only newly acquired apo-E obtained from HDL may have the appropriate conformation for receptor binding.

Metabolic Fate
VLDL triglycerides are hydrolyzed by the actions of LPL and hepatic lipase. They are converted to smaller and smaller particles that become increasingly rich in cholesterol (see Fig. 34-17). The products of VLDL catabolism are IDLs (d = 1.006 to 1.019 g/mL). IDLs retain apo-B100 and apo-E but have lost most of the C apolipoproteins. IDLs are processed to LDLs (d = 1.019 to 1.063 g/mL) by LDL with final processing by hepatic lipase. Approximately half of VLDLs are converted to LDLs, and the remainder are cleared directly by the liver as VLDL remnants (small VLDL) and IDLs (see Fig. 34-17). The uptake of VLDL remnants and IDLs by liver parenchymal cells is mediated by apo-E, and the uptake of LDL by the LDL receptor is mediated by apo-B100.
Intermediate-Density Lipoproteins

IDLs (d = 1.006 to 1.019 g/mL) are normally present in low concentrations in the plasma and are intermediate in size and composition between VLDL and LDL. Their primary proteins are apo-B100 and apo-E. The IDLs are precursors of LDLs and represent metabolic products of VLDL catabolism in the plasma by the action of lipases. As shown in Figure 34-17, IDLs may be further processed by hepatic lipase or removed from the plasma by the LDL receptor. IDLs are often considered to be VLDL remnants and to be atherogenic.
Low-Density Lipoproteins

Characteristics

LDLs (\(a = 1.019\) to 1.063 g/mL), which are about 200 Å in diameter, are the major cholesterol-carrying lipoproteins in the plasma; about 70% of total plasma cholesterol is in LDL. LDLs are composed of approximately 75% lipid (about 35% cholesteryl ester, 10% free cholesterol, 10% triglyceride, and 20% phospholipid) and 25% protein (see Table 34-4). Apo-B100 is the principal protein in these particles, along with trace amounts of apo-E (see Fig. 34-6). LDLs have -electrophoretic mobility and were previously referred to as lipoproteins.

Origin

LDLs are the end products of lipase-mediated hydrolysis of VLDLs (see Fig. 34-17). Moreover, as the triglyceride-rich core of the larger VLDL particles is removed, the surface lipids and proteins are remodeled and excess surface constituents are transferred to HDL, resulting in the formation of a small, cholesterol-rich LDL devoid of almost all apolipoproteins except apo-B100.

Metabolic Fate

About 75% of LDL is taken up by hepatocytes. Other tissues take up smaller amounts of LDL. Approximately two thirds of the uptake is mediated by the LDL receptor, and the remainder is mediated by a poorly defined process that does not involve receptors. LDLs are considered to be atherogenic.

Apolipoproteins B and E Determine Rate of Plasma Lipoprotein Clearance

The rate of clearance of lipoproteins from the plasma is determined by the apolipoprotein that mediates the interaction with the receptor and by the number of receptors expressed on the cell surface (primarily in the liver). VLDL and IDL are rapidly cleared from the plasma (their half-lives are measured in minutes to a few hours). Apo-E mediates their binding to the LDL receptors. Multiple apo-E molecules per lipoprotein can interact with more than one receptor or with multiple sites on a receptor. Multiple interactions enhance binding affinity and increase the clearance of these particles from the plasma. The clearance of LDL is mediated by apo-B100. The affinity of apo-B100 for the LDL receptor is lower than that of apo-E, and clearance of LDL is much slower (with a half-life of 2 to 3 days). Compared with apo-B100-containing LDLs, apo-E-containing lipoproteins have 20-fold greater affinity for the LDL receptor.

Role for Lipoprotein Cholesterol in Cellular Metabolism

All cells can synthesize cholesterol de novo. However, LDL serves as a source of cholesterol for many cells. Cholesterol taken up by the liver has several fates: membrane biosynthesis, VLDL biosynthesis, excretion as cholesterol in the bile, and conversion to bile acids. Cholesterol is used as a precursor for steroid hormone production in the adrenals, ovaries, and testes. In other peripheral tissues, cholesterol is used in membrane biosynthesis for cell repair and proliferation.

Factors Affecting Low-Density Lipoprotein Levels in the Blood

Plasma LDL levels can be increased through two primary mechanisms: (1) increased VLDL biosynthesis and secretion caused by increased flux of free fatty acids to the liver from dietary fats or from mobilization from adipose tissue and (2) decreased LDL catabolism. Decreased catabolism can result from (1) decreased LDL receptor levels in hepatic and extrahepatic tissues (LDL receptor expression is down-regulated when cells have enough cholesterol for their metabolic needs or when diets are high in saturated fat and cholesterol), (2) increased numbers of high-affinity apo-E-containing lipoproteins that compete with LDL for receptor interaction (as discussed previously), (3) defective LDL receptors incapable of normal interaction with apo-B100, and (4) defective apo-B100 incapable of normal interaction with LDL receptors.
High-Density Lipoproteins

Characteristics

HDLs are small particles (70 to 120 Å in diameter) that float at densities of 1.063 to 1.21 g/mL. They are somewhat arbitrarily divided into two major subclasses: HDL2 (d = 1.063 to 1.125 g/mL) and HDL3 (d = 1.125 to 1.21 g/mL). HDLs contain about 50% lipid (25% phospholipid, 15% cholesteryl ester, 5% free cholesterol, and 5% triglyceride) and 50% protein (see Table 34-4). Their major apolipoproteins are apo-AI (65%), apo-AII (25%), and smaller amounts of the C apolipoproteins and apo-E (see Fig. 34-4). Apo-E is a minor component of a subclass of HDL referred to as HDL-E, but about 50% of total plasma apo-E is in this HDL fraction. The major classes of HDLs lack apo-E and therefore do not interact with the LDL receptor. HDLs serve as a reservoir for apo-E and the C apolipoproteins to be distributed to other lipoproteins when they enter the plasma (e.g., chylomicrons, VLDLs). Subclasses of HDL may contain only apo-AI (called LpAI) or apo-AI and apo-AII (called LpAII). Although LpAI and LpAII do not correspond directly to the ultracentrifugal fractions, LpAI corresponds primarily to HDL3 and LpAII to HDL2. The HDLs as a class have -electrophoretic mobility and were previously referred to as lipoproteins.

Origin

HDLs originate from three major sources (Fig. 34-21). First, the liver secretes an apo-AI-phospholipid disc called nascent or precursor HDL. Second, the intestine directly synthesizes a small apo-AI-containing HDL particle. Third, HDLs are derived from surface material (primarily apo-AI and phospholipid) that comes from chylomicrons and VLDLs during lipolysis. As chylomicrons and VLDLs become depleted of cholesterol and triglyceride, excess material is shed from the surface of the particle in combination with apo-AI to form small HDL discs. The phospholipid transfer protein facilitates the shedding of the surface material during lipolysis of triglyceride-rich lipoproteins to generate the HDL precursors.

Maturation of High-Density Lipoproteins

The nascent or precursor HDL particles exist as apo-AI-phospholipid discs. Designated pre-1, pre-2, and pre-3, these discs are excellent acceptors of free cholesterol from cell membranes with excess cholesterol or from other lipoproteins. The pre-phospholipid discs can accommodate only a limited amount of free cholesterol. However, esterification of the cholesterol with a long-chain fatty acid increases its hydrophobicity, and the newly formed cholesteryl ester moves away from the surface of the disc, beginning the process of forming a cholesteryl ester-rich core and converting the disc to a sphere. The enzyme in plasma that converts free cholesterol to cholesteryl ester is LCAT.

The small, spherical, mature HDL particles (HDL3) also serve as acceptors for free cholesterol; as more free cholesterol is acquired and esterified, the particles increase in size, forming HDL2. These HDL subclasses can include LpAI, or they can be converted to LpAII particles by the addition of apo-AI.

In some animals and to a lesser extent in humans, HDL4 can be further enriched in cholesteryl ester and at the same time acquire apo-E (Fig. 34-22). These apo-E-containing HDL4 are a minor but metabolically active subclass of HDL. The presence of apo-E targets the HDL4 to cells expressing the LDL receptor. Typical HDLs lack apo-E and do not interact with the LDL receptor. The HDL4 represent a major HDL class in many lower species and in humans with abetalipoproteinemia or CETP deficiency.

HDL4 can also arise from a precursor particle that displays -electrophoretic mobility and is called LpE. This particle is approximately 80% protein and 20% lipid (primarily sphingomyelin).

Acquisition of Cholesterol by High-Density Lipoproteins

HDLs, especially HDL3, precursors of mature HDL, and lipid-poor apo-AI, can acquire cholesterol from cells by two mechanisms: aqueous transfer from cells and transport facilitated by a cell-surface binding protein.

Aqueous Transfer from Cells

The HDLs come in close contact with cells having excess cholesterol and acquire free cholesterol (not cholesteryl ester) from the cell surface. Free cholesterol follows a physicochemical concentration gradient from the cell to the HDL particle, from a high concentration of free cholesterol in the membranes of cells with excess cholesterol to a low concentration at the surface of the HDL. This process is referred to as passive desorption.

Transport Facilitated by a Cell-Surface Binding Protein

At least two cell-surface proteins facilitate the efflux of free cholesterol from cells possessing excess cholesterol. The class B, type I scavenger receptor (SR-BI) binds HDL particles to the cell surface. This receptor may alter the organization of the cell membrane lipids facilitating the efflux of free cholesterol from the membrane to the lipoprotein. The HDLs are not internalized by the cell and are released into the circulation when the particle is enriched in cholesterol. The second receptor that participates in the efflux of cholesterol from cells is the ATP binding cassette transporter A1 (ABCA1). It appears to bind apo-AI or a pre-HDL disc to the cell membrane and facilitate the transfer of free cholesterol and phospholipid from the cell to enrich the HDL precursors in these lipids. Mutations in ABCA1 prevent the efflux of cholesterol from cells, resulting in absence of mature HDL and rapid catabolism of apo-AI and causing the lipid disorder called Tangier disease.
HDLs function in the redistribution of lipids among lipoproteins and cells by a process called reverse cholesterol transport. In exchange for the function of cholesterol esters to the liver, HDLs acquire cholesterol from cells and transport it to the liver for excretion or to other cells that require cholesterol. The scheme is shown in Figure 34-22.

The mechanism for the transfer of the fatty acid to cholesterol has not been well defined. LCAT has two different enzymatic activities. First, lecithin cleavage (phospholipase activity) involves the ester bond of the fatty acid in position 2 of lecithin, which is usually linoleic acid (C18:2), yielding lyssolecithin and the fatty acid. The fatty acid becomes covalently linked to serine-181 in the LCAT molecule. Second, transesterification (transacylase activity) involves the transfer of the fatty acid attached to LCAT to the 3-hydroxyl position of cholesterol, forming a cholesteryl ester. The mechanism for the transfer of the fatty acid to cholesterol has not been well defined.

Much has been learned about the normal function of HDL in lipoprotein metabolism by studying patients who have low or undetectable activity of this enzyme in plasma. HDL deficiency can be caused by mutations that affect the structure of LCAT or of apo-AI. The disorder is manifested by low levels of cholesterol esters, low levels of HDL, and clinical features ranging from mild symptoms such as corneal clouding (caused by accumulation of free cholesterol in the cornea) to severe disorders such as renal failure (see "Lecithin:Cholesterol Acyltransferase Deficiency").

**Metabolic Pathways Involving High-Density Lipoproteins**

HDLs function in the redistribution of lipids among lipoproteins and cells by a process called reverse cholesterol transport. In exchange for the function of cholesterol esters to the liver, HDLs acquire cholesterol from cells and transport it to the liver for excretion or to other cells that require cholesterol. The scheme is shown in Figure 34-22. HDLs acquire cholesterol from cells and transport it to the liver for excretion or to other cells that require cholesterol. The scheme is shown in Figure 34-22.

The channels appear to represent flaps of cell-surface membrane that form a 150- to 250-Å-wide cleft in which the lipoproteins are trapped at least transiently. Within these channels, cholesterol ester and free cholesterol can be extracted from HDLs without the lipoprotein particle entering the cell or being degraded. The SR-BI can facilitate the transfer of cholesteryl esters from HDLs to cells without the lipoprotein particle entering the cell or being degraded. The SR-BI appears to function by transferring cholesteryl ester through a hydrophilic channel formed in the cell membrane. Hepatic lipase may be involved in the selective uptake of cholesterol from the HDL by hydrolyzing the phospholipids on the particles and creating a chemical gradient that promotes the transfer of cholesterol from the particle to the cell. Recall that hepatic lipase is localized in the space of Disse of the liver and in the adrenal glands and ovaries. In the kidneys, apo-AI is removed in preference to cholesterol; this apo-AI may be dissociated from the HDL particle, filtered, and degraded. Ultimately, intact HDL can be taken up by hepatocytes and degraded. Although HDL, represents a small fraction of total HDL, it is taken up directly by LDL receptors and degraded in the liver through the LDL pathway.

The catabolism of HDL is not entirely understood. Specific lipid moieties of HDL can be taken up by cells without removal of the intact particle from the plasma compartment. For example, cholesterol esters are removed from the particle by selective uptake and preferentially delivered to the liver, adrenal glands, and gonads. The SR-BI can facilitate the transfer of cholesteryl esters from HDLs to cells without the lipoprotein particle entering the cell or being degraded. The SR-BI appears to function by transferring cholesteryl ester through a hydrophilic channel formed in the cell membrane. Hepatic lipase may be involved in the selective uptake of cholesterol from the HDL by hydrolyzing the phospholipids on the particles and creating a chemical gradient that promotes the transfer of cholesterol from the particle to the cell. Recall that hepatic lipase is localized in the space of Disse of the liver and in the adrenal glands and ovaries. In the kidneys, apo-AI is removed in preference to cholesterol; this apo-AI may be dissociated from the HDL particle, filtered, and degraded. Ultimately, intact HDL can be taken up by hepatocytes and degraded. Although HDL, represents a small fraction of total HDL, it is taken up directly by LDL receptors and degraded in the liver through the LDL pathway.

**Selective Uptake of Cholesterol by Steroidogenic Cells**

HDL is more efficient than LDL in delivering cholesterol to steroidogenic cells of the adrenal, ovary, and testis. In these organs, the lipoproteins concentrate on the surface of cells in micellar structures. The channels appear to represent flaps of cell-surface membrane that form a 150- to 250-Å-wide cleft in which the lipoproteins are trapped at least transiently. Within these channels, cholesterol ester and free cholesterol can be extracted from HDLs without the lipoprotein particle entering the cell or being degraded. The SR-BI appears to function by transferring cholesteryl ester through a hydrophilic channel formed in the cell membrane. Hepatic lipase may be involved in the selective uptake of cholesterol from the HDL by hydrolyzing the phospholipids on the particles and creating a chemical gradient that promotes the transfer of cholesterol from the particle to the cell. Recall that hepatic lipase is localized in the space of Disse of the liver and in the adrenal glands and ovaries. In the kidneys, apo-AI is removed in preference to cholesterol; this apo-AI may be dissociated from the HDL particle, filtered, and degraded. Ultimately, intact HDL can be taken up by hepatocytes and degraded. Although HDL, represents a small fraction of total HDL, it is taken up directly by LDL receptors and degraded in the liver through the LDL pathway.

**Cholesteryl Ester Transfer Protein**

CETP facilitates the transfer of cholesteryl esters from HDL to the lower density, triglyceride-rich lipoproteins (primarily VLDL, IDL, and remnants). CETP plays a pivotal role in lipid metabolism and may affect susceptibility or resistance to the development of atherosclerosis. For example, humans, nonhuman primates, and rabbits have significant amounts of CETP activity in their plasma. As a consequence, they form only small amounts of HDL,; they dispose of most of their HDL cholesterol esters by delivering them to lower density lipoproteins. Ultimately, most of the cholesteryl esters leave the plasma by the LDL pathway. These species are susceptible to atherosclerosis and tend to have higher levels of LDL. On the other hand, rats, mice, and dogs have no CETP activity, readily form HDL, and can deliver the cholesterol directly to the liver by the apo-E-mediated pathway. These animals have very low levels of LDL and are resistant to the development of atherosclerosis. These observations suggest that high levels of CETP activity accelerate atherosclerosis and that inhibition of CETP may be beneficial in treating certain types of hyperlipidemia.

However, this concept has been brought into question by the observation that Japanese Americans with a deficiency of CETP have increased HDL but nevertheless develop CHD. The HDL in these subjects tends to be the large HDL, and levels of the smaller HDL, are decreased. If these data concerning the atherogenicity of low CETP activity are confirmed, it may mean that low levels of HDL, which serves as the most potent acceptor of cellular cholesterol, are a major risk factor for...
CHD in these subjects; alternatively, high levels of the large apo-E-containing HDL may be atherogenic.

Data obtained through overexpression of CETP in transgenic mice do not clarify whether high levels are protective or detrimental. In one study, overexpression of CETP led to accelerated atherogenesis, but in a study in which CETP was overexpressed in hypertriglyceridemic mice expressing high levels of apo-CIII, there was less atherosclerosis even though the mice were hyperlipidemic and had low HDL levels. The potential therapeutic value of lowering CETP to retard atherogenesis must be questioned until these inconsistencies are sorted out.

High-Density Lipoproteins as Antiatherogenic Lipoproteins

Numerous studies have demonstrated that high levels of HDL-C are associated with a lower incidence of CHD. Conversely, low levels of HDL-C are associated with a higher incidence of CHD. The protective mechanism involving HDL may be related to its role in reverse cholesterol transport, which results in redistribution of cholesterol away from the artery wall. Although low HDL-C is a major CHD risk factor, it must be kept in mind that the HDLs are a heterogeneous group of molecules having different metabolic roles. Some may be protective (e.g., LpAI, HDL₂), and others may not be (e.g., LpAI/AII). As the complex nature of HDL is unraveled, it may be possible to define an antiatherogenic spectrum of HDL particles and determine the metabolic and therapeutic measures needed to alter these HDLs selectively.
LIPIDS AND ATHEROSCLEROSIS

Atherosclerosis causes a reduction of blood flow and insufficient delivery of oxygen and nutrients to affected organs. Insufficient oxygen results in ischemia or infarction, leading to angina or myocardial infarction in the case of restricted blood flow to the heart muscle, to stroke with reduced blood flow to the brain, or to intermittent claudication with restricted blood flow to the lower extremities. CHD is the leading cause of death in the United States and in western Europe.

The restricted arterial blood flow in atherosclerosis is caused by changes in the vessel wall characterized by lipid deposition and cell proliferation. Narrowing of the vessel lumen may lead to obstruction and, more important, to unstable plaques susceptible to ulceration or fissure formation causing thrombosis. The deposited lipids are derived from plasma lipoproteins, and elevated plasma cholesterol represents a major risk factor. Other important risk factors include low HDL levels, cigarette smoking, hypertension, male sex, diabetes mellitus, obesity, stress, and lack of exercise. The discussion here focuses on plasma lipoproteins and the cholesterol-diet-heart hypothesis.

Cholesterol-Diet-Heart Hypothesis

For the last 40 years, evidence linking high plasma cholesterol concentrations with an increased risk for CHD has been accumulating, and the evidence is now overwhelming and indisputable. The cholesterol-diet-heart hypothesis states (1) that increased plasma cholesterol concentrations increase the risk of CHD, (2) that diets high in fat (especially saturated fat) and cholesterol result in increased levels of plasma cholesterol, and (3) that lowering plasma cholesterol levels results in a decreased risk of CHD.

Animal Models

Numerous animal models have demonstrated that diets enriched in cholesterol and saturated fat elevate plasma cholesterol levels and lead to atherosclerosis with many features of the human disease.

Studies in monkeys are particularly relevant. Feeding monkeys a diet that approximates the typical Western diet (500 mg cholesterol per day and 20% of calories as saturated fat) resulted in elevation of plasma LDL concentrations and atherosclerotic lesions almost identical to those seen in humans. Eliminating the saturated fat and cholesterol from the diet reduced LDL levels and caused lesions to regress. In addition to elevations in LDL concentrations, cholesterol-fat feeding in animals caused accumulation of VLDL. These cholesterol-enriched remnant lipoproteins are derived from lipoproteins secreted by the intestine and liver and also accumulate in type III hyperlipoproteinemia (discussed later). Single-gene mutations in animals have also demonstrated the link between hypercholesterolemia and atherosclerosis. The WHHL rabbit is a model of FH in which the deposited cholesterol from LDL is increased in the vessel walls and the aorta of these animals develops severe spontaneous atherosclerosis, even when fed a low-fat diet.

Epidemiologic Evidence

Several epidemiologic studies have demonstrated a relation between the plasma cholesterol level and the risk of CHD. For example, the Multiple Risk Factor Intervention Trial (Fig. 34-23) showed that there is increased risk at levels above 5.2 mmol/L (200 mg/dL). The Seven Countries Study also demonstrated a relation between an increased incidence of CHD and high plasma cholesterol levels (Fig. 34-24). The causal relationship between elevated plasma cholesterol levels and accelerated atherosclerosis is established. Epidemiologic studies have linked the intake of high levels of dietary fat, especially saturated fats, with increased plasma cholesterol levels. Likewise, diets high in cholesterol also tend to increase plasma cholesterol levels. Therefore, restriction of saturated fat and cholesterol is the cornerstone of dietary therapy to reduce elevated blood cholesterol levels.

Familial Hypercholesterolemia

Some of the most compelling evidence for the deleterious effects of elevated plasma cholesterol has come from studies of FH. This genetic disorder results from a series of mutations in the LDL receptor that cause LDL to accumulate in the plasma as a result of defective clearance of LDL by the receptors. These studies conclusively demonstrate that increased plasma concentrations of LDL cause atherosclerosis.

Experimental Evidence in Humans

Final compelling evidence supporting the cholesterol-diet-heart hypothesis came from several human clinical trials examining the efficacy of several lipid-lowering drugs in reducing CHD (reviewed in the section on treatment of lipid disorders, clinical trials). In all groups examined, including patients with and without preexisting
CHD over a range of initial plasma cholesterol levels, the results unequivocally demonstrated that lowering plasma cholesterol levels reduces the risk of CHD.

In summary, the current evidence overwhelmingly supports the cholesterol-diet-heart hypothesis and upholds the conclusion of the Cholesterol Consensus Conference on Lowering Blood Cholesterol to Prevent Coronary Heart Disease organized by the National Heart, Lung, and Blood Institute that the cause-and-effect relation between cholesterol and CHD is clearly established.
Atherogenic Lipoproteins

In addition to LDL, almost all classes of lipoproteins that contain apo-B (VLDL, -VLDL, IDL, Lp(a), and oxidized LDL) are considered to be atherogenic. A common feature of these atherogenic lipoproteins is that they contain various amounts of cholesterol esters and either apo-B-100 or apo-B-48. In addition, Lp(a) contains apo(a), a protein that is disulfide-linked to apo-B and is homologous to plasminogen; apo(a) may contribute to atherogenesis by mechanisms related to thrombosis. Furthermore, the atherogenic potential of LDL differs among the various LDL size and density subclasses, with the small, dense LDL subclass being the most atherogenic.

Apo-B-containing remnant lipoproteins appear to be especially atherogenic because -VLDLs, which accumulate in the plasma of cholesterol-fed animals and in patients with type III hyperlipoproteinemia, are associated with accelerated formation of atherosclerotic lesions. These particles, representing chylomicron remnants and VLDL remnants (IDL), are taken up by macrophages, including, presumably, macrophages in the artery wall, in a nonsaturable manner. This uptake results in massive intracellular accumulation of cholesteryl esters in the form of lipid droplets. The lipid-engorged macrophages resemble the foam cells of the early fatty streak (discussed subsequently).

Another related class of potentially atherogenic apo-B-containing lipoproteins is triglyceride-rich lipoproteins, which are associated with postprandial lipemia after ingestion of a fatty meal. Whereas chylomicrons and large, triglyceride-rich VLDLs are not believed to be atherogenic, remnants derived from these particles are.

The Low-Density Lipoprotein Paradox and Oxidized Lipids

Because LDL-C levels are a strong predictor of CHD and atherosclerosis, it was expected that LDL would be taken up avidly by macrophages, leading to the formation of foam cells. However, in vitro experiments showed that only low levels of normal plasma LDL are taken up by macrophages. This low uptake presumably occurred because of the highly regulated LDL receptor pathway; the delivery of LDL-C to macrophages down-regulates LDL receptor expression, thereby protecting the cells from overaccumulation of LDL-C. These results led to the so-called LDL paradox: How do LDLs contribute to atherosclerosis if only limited quantities are taken up by macrophages? The explanation turned out to be that LDLs that have been modified are taken up by macrophages in an unregulated manner through receptors unrelated to the LDL receptor. These receptors for modified LDL were originally referred to as acetyl-LDL receptors but are now commonly referred to as scavenger receptors (discussed in "Scavenger Receptors").

In vitro experiments have demonstrated that a number of chemical modifications, including acetylation, acetoacetylation, and reaction with malondialdehyde, circumvent the LDL receptor pathway and cause massive amounts of modified LDL to enter macrophages by means of scavenger receptors. Furthermore, macrophages can alter LDL so that these particles can be taken up by macrophages in an unregulated manner. Other cells, including smooth muscle cells, can also modify LDL.

The physiologically important LDL modification probably involves oxidation and results in lipid peroxidation. The oxidized-LDL hypothesis proposes that unsaturated lipids on the particle undergo oxidative modification, which subsequently leads to oxidation of apo-B, which alters the protein's affinity for cell-surface receptors. As a corollary, antioxidant vitamins (such as A, C, and E), drugs (such as probucol), and enzymes (such as paraoxonase) may limit these oxidative processes. It appears that production of reactive oxygen species (i.e., free radicals) is an integral part of the modification and may be related to the general aging process, in which lipid peroxidation may be a component. Two products of lipid peroxidation, 4-hydroxynonenal and malondialdehyde, modify amino acids of apo-B100, resulting in its fragmentation. Modification is inhibited by antioxidants. Phospholipase and lipoxigenases have also been implicated in LDL modification. In addition to macrophage uptake, oxidized LDLs may participate directly in atherogenesis because they are cytotoxic and may serve as chemoattractants for circulating monocyte-macrophages and are immunogenic.

The role of oxidized LDL as a major contributor to atherosclerosis remains to be proved in vivo. However, what appear to be oxidatively modified forms of LDL have been identified in atherosclerotic lesions and inflammatory fluid. Also, epitopes of malondialdehyde- and 4-hydroxynonenal-modified apo-B100 have been observed in lesions. Furthermore, the formation of oxidized LDL, which may contribute to atherogenesis in a number of ways, is an attractive solution to the LDL paradox.

The most probable mechanism by which oxidized or modified LDLs are taken up by macrophages is by one or more of the scavenger receptors (see "Scavenger Receptors"). The roles of the various scavenger receptors and their relative importance in atherogenesis are still being clarified (for a review, see reference ). However, the receptor whose primary function is to take up oxidized or modified LDL within the artery wall and to contribute to the development and progression of atherosclerosis may well remain to be identified.
Overview of Atherogenesis

Until the last several years, two theories of atherogenesis prevailed. The first, referred to initially as the lipid infiltration hypothesis, proposed that excess blood lipids in the form of lipoproteins infiltrate into the arterial wall. This theory was supported by epidemiologic evidence and the identification of atherogenic lipoproteins. The second theory was the endothelial injury hypothesis, which proposed that injury to the endothelial surface is required and results in removal of these cells, exposing a thrombogenic surface to which platelets adhere. Platelet adherence was suggested to result in the release of platelet-derived growth factor, which would stimulate the smooth muscle cell proliferation and migration that are characteristic of early lesion formation. Endothelial denudation would also remove the endothelial barrier, allowing lipoproteins to enter the vessel wall more readily. The current view of atherosclerosis development combines features of both hypotheses. However, loss of endothelial cells is not required for an atherosclerotic lesion to develop, and atherogenic lipoproteins can and do penetrate intact endothelium to enter the artery wall.

A unifying modification of this general view of atherogenesis is the response-to-retention hypothesis. According to this view, one key event in atherosclerosis is the retention of atherogenic, cholesterol-rich lipoproteins bound to arterial proteoglycans in the arterial subendothelium. Oxidation and other modifications of the retained lipoproteins could initiate a series of responses that lead to the transformation of healthy, normal arteries into diseased, lesioned arteries. An important component of the artery wall involved in the retention is chondroitin 6-sulfate proteoglycans, which are known to bind LDL. Although oxidation may be a major fate of the retained lipoproteins, it is not the only possibility. Nonpancreatic secretory phospholipase A₂ has been proposed to generate proinflammatory modified lipid components from aggregated but nonoxidized lipoproteins. Also, induction of cell-adhesion molecules such as vascular cell adhesion molecule 1, leading to recruitment of inflammatory cells, appears to be triggered by nonoxidized, retained lipoproteins.

Consistent with a prominent role for inflammation, atherosclerotic lesions share many features with wound healing or inflammation: (1) proliferation of smooth muscle cells and accumulation of macrophages; (2) formation, by smooth muscle cells, of a connective tissue matrix composed of elastic fibers, collagen, and proteoglycans; and (3) deposition of lipid, primarily cholesterol, both intracellularly and extracellularly. The pattern of lesion formation does not occur randomly within the arterial tree but is focal in nature. Susceptible regions are more permeable to plasma components, and endothelial cell turnover is greater, although the endothelial surface appears to be intact. The focal nature of atherosclerosis suggests that local hemodynamic factors are involved.

Normally, the endothelium forms a relatively impermeable barrier. The endothelial cells and the relatively narrow region beneath them (subendothelial space), which contains an occasional smooth muscle cell, constitute the intima of the artery wall (Fig. 34-25A). Beneath the intima is a layer of many smooth muscle cells, the media, which constitutes the bulk of the arterial wall. The adventitia is the outermost layer of the artery wall and is composed of loose connective tissue.

A current model of atherogenesis is depicted in Figure 34-25B through E. The major cell types involved include endothelial cells, smooth muscle cells, and inflammatory mononuclear cells, such as macrophages and possibly lymphocytes. One of the initial events is the focal attachment of circulating monocytes to the endothelial surface (see Fig. 34-25B). It is not entirely clear which factors are responsible for the adherence of monocytes, although oxidized or modified LDLs or other atherogenic lipoproteins retained in the subendothelium are probably a major initiating factor; areas of micronjury may contribute as well. The monocytes modify the endothelial surface and induce the expression of leukocyte adhesion molecules such as vascular cell adhesion molecule 1. Once adhered, the monocytes migrate between endothelial cells, enter the subendothelial space, and differentiate into macrophages (see Fig. 34-25B). In addition, LDLs and other atherogenic lipoproteins can enter this space. Within the wall, LDLs may become entrapped in the matrix and undergo oxidation or further chemical modifications. The macrophages take up the oxidized or modified LDLs and begin to take on the appearance of foam cells as lipid accumulates. These initial steps set in motion a chain of events that includes the expression of growth factors (mediators of cell proliferation and chemotaxis) and cytokines (mediators involved in inflammation and immunity).

Monocyte chemoattractant protein 1 (MCP-1), produced by endothelial and smooth muscle cells, plays a role in further monocyte recruitment into lesions and may be induced by the presence of oxidized LDL. The role of MCP-1 in macrophage recruitment in the early stages of lesion development was established in MCP-1-deficient mice. MCP-1 stimulates monocyte chemotaxis, and the absence of MCP-1 results in a substantial decrease in monocyte infiltration into lesions of apo-E-deficient mice. When the MCP-1-deficient mice were crossed with apo-E-deficient mice, a common mouse model of atherosclerosis, and fed a high-fat Western-type diet for 5 weeks, fewer macrophages were present in the aortas of the double transgenic mice than in the apo-E-deficient control mice. After 5 to 26 weeks of the diet, the double transgenic mice displayed significantly smaller lesions. These results establish a role for MCP-1 in the recruitment of macrophages into the artery wall at an early stage of lesion formation and establish MCP-1 as an important factor for atherogenesis.

Other growth factors that have been implicated in atherosclerosis include platelet-derived growth factor, basic fibroblast growth factor, insulin-like growth factors, interleukin-1, tumor necrosis factor, and transforming growth factor. These mitogenic factors, which can stimulate smooth muscle cell proliferation, are not expressed in the normal artery wall but are present in developing lesions. Several of these mitogens are also chemoattractants with the potential to attract smooth muscle cells or monocytes-macrophages into a developing lesion. Inflammatory response cytokines include interleukin-1, interferon, tumor necrosis factor, interleukin-2, and the colony-stimulating factors. It is unlikely that the various factors act in isolation from each other, but they probably act through a network of cellular interactions operating in a paracrine or autocrine manner.

The first grossly visible atherosclerotic lesion is referred to as a fatty streak (see Fig. 34-25C). Macrophages accumulate in abundance in the subendothelial space and are converted to foam cells, presumably through the uptake of oxidized LDLs or remnant lipoproteins. The recruitment of monocytes continues, and smooth muscle cells begin to migrate into the intima. Fatty streaks probably come and go, depending on the local stimuli present in the artery wall.
As the cycle of interactions continues, the fatty streak matures into a proliferative or fibrous plaque, which is raised and begins to extend into the lumen of the vessel (see Fig. 34-25D). The foam cells begin to necrose, probably because of the cytotoxicity of the accumulated lipid; as the lesion progresses, cholesterol crystals develop. The death of foam cells leads to extracellular lipid deposition, accompanied by collagen synthesis and smooth muscle cell migration and proliferation. In the continued presence of factors that promote atherogenesis (e.g., high plasma concentrations of atherogenic lipoproteins), the plaque progresses to the complicated lesion stage (see Fig. 34-25E).

The surfaces of complicated lesions may become thrombogenic as endothelial cells are lost and the subendothelial space is exposed. Platelets can adhere to this exposed surface, promoting thrombus formation. Alternatively, a fissure forms in the unstable plaques and blood actually dissects into the artery wall, leading to the formation of a large thrombus. At late stages in complicated lesions, T lymphocytes infiltrate the lesion and there is evidence of an autoimmune response characterized by lymphocyte infiltration of the adventitia. Calcification is also a feature of late lesions. Advanced lesions can weaken the elasticity and integrity of the artery wall, with the potential to lead to an aneurysm of the vessel. As experiments in humans have shown, removal or reduction of the atherogenic stimulus can result in plaque regression, leaving a remnant devoid of lipid that resembles a wound scar.
HYPERLIPIDEMIA: DEFINITIONS AND OVERVIEW

Plasma lipid levels vary among individuals of different populations owing to genetic and dietary factors. For example, the mean plasma cholesterol concentration for Western men is 5.4 mmol/L (210 mg/dL), whereas for Japanese men it is 4.3 mmol/L (165 mg/dL). Historically, hyperlipidemia has been arbitrarily defined from population distributions as the upper 5% to 10% of values (i.e., the 90th to 95th percentile). For Western adults, cholesterol concentrations higher than 6.2 mmol/L (240 mg/dL) or triglyceride concentrations higher than 2.3 mmol/L (200 mg/dL) constitute significant hyperlipidemia.

In agreement with these definitions, guidelines from the 2001 National Cholesterol Education Program (NCEP) suggest that plasma cholesterol levels less than 5.2 mmol/L (200 mg/dL) are desirable, that those between 5.2 and 6.2 mmol/L (200 to 240 mg/dL) are borderline elevated, and that levels greater than 6.2 mmol/L (240 mg/dL) are high. Because plasma lipid levels increase with age, cutoff values for hyperlipidemia in children are lower (5.2 mmol/L [200 mg/dL] and 1.6 mmol/L [140 mg/dL], respectively, for cholesterol and triglycerides). Conversely, hypolipidemia can be defined as plasma cholesterol concentrations less than 3.4 mmol/L (130 mg/dL). Although these boundaries are arbitrary, the designation of plasma cholesterol concentrations higher than 6.2 mmol/L (240 mg/dL) as hyperlipidemic has support from clinical observations indicating that the risk for atherosclerotic CHD increases markedly when the cholesterol concentration reaches or exceeds this level.

Hyperlipidemia is caused by increased concentrations of plasma lipoproteins. One or more classes of lipoproteins may accumulate in the blood stream because of increased production or secretion into the circulation or because of decreased clearance or removal from the circulation; in some cases, both processes coexist. Alterations in metabolic processes are often related to alterations in the proteins involved in lipoprotein metabolism (see "Plasma Lipoproteins: Apolipoproteins, Receptors, and Enzymes"). Alterations resulting from genetic defects are classified as primary disorders of lipid metabolism. Alternatively, other factors that alter lipoprotein metabolism, such as diabetes mellitus or hypothyroidism, lead to increased plasma lipoprotein concentrations; these are classified as secondary disorders of lipid metabolism. Often, hyperlipidemia results from mixed primary and secondary causes, such as when diabetes mellitus occurs in a subject who has an inherited defect in one of the proteins involved in lipoprotein metabolism. In cases in which no known cause of hyperlipidemia can be identified, the disorder is classified as sporadic or possibly polygenic in origin.

When considering the causes of hyperlipidemia, it is useful to classify the various possibilities by the pattern of plasma lipoprotein elevation (or the lipoprotein phenotype). This was once commonly done by performing plasma electrophoresis to separate and determine the relative concentrations of the plasma lipoproteins (Table 34-5). Although such a classification is still useful in some instances, it has generally lost utility because it has been recognized that certain disorders can manifest different phenotypes at different times in the same person and that different phenotypes may occur in different family members with the same disorder. Furthermore, classification based on electrophoresis is not necessary to determine a therapeutic plan. Nevertheless, many of the terms still used to describe lipoprotein disorders (e.g., hypopalphalipoproteinemia, hypobetalipoproteinemia, abetalipoproteinemia, type III hyperlipoproteinemia, dysbeta1lipoproteinemia) are derived from the patterns observed on plasma electrophoresis gels.

It is also possible to create a differential diagnosis on the basis of whether the concentration of plasma cholesterol, triglycerides, or both is elevated. Table 34-6 illustrates such a diagnostic strategy. More extensive lists of primary and secondary disorders are given in Table 34-7 and Table 34-8. Although it is often not essential to diagnose a genetic disorder in a hyperlipidemic subject for treatment purposes, the understanding of genetic causes may have important implications for family members. The recognition of secondary disorders is of great importance because therapy should be directed, at least in part, toward correcting the underlying disorder.
PRIMARV DISORDERS OF HYPERLIPIDEMIA

Familial Hypercholesterolemia

FH is a relatively common disorder caused by mutations in the LDL receptor gene that result in LDL receptor malfunction or absence in cells of the liver and peripheral tissues, leading to elevation of plasma LDL and total cholesterol concentrations. Heterozygous subjects typically have plasma cholesterol concentrations that are twofold to threefold above average, and homozygous subjects have cholesterol concentrations that are elevated threefold to sixfold.

Table 34-6 -- Differential Diagnosis of Hyperlipidemia Including Common Secondary Disorders

<table>
<thead>
<tr>
<th>Type of Disorder</th>
<th>Major Plasma Lipid Abnormality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary</td>
<td>Increased Cholesterol</td>
</tr>
<tr>
<td>Familial hypercholesterolemia</td>
<td>Familial combined hyperlipidemia</td>
</tr>
<tr>
<td>Familial defective apo-B100</td>
<td>Type III hyperlipoproteinemia (dysbetalipoproteinemia)</td>
</tr>
<tr>
<td>Polygenic hypercholesterolemia</td>
<td>Apo-CII deficiency</td>
</tr>
<tr>
<td>Secondary</td>
<td>Increased Cholesterol and Triglyceride</td>
</tr>
<tr>
<td>Hypothyroidism</td>
<td>Family hyperlipidemia</td>
</tr>
<tr>
<td>Nephrotic syndrome</td>
<td>Diabetes mellitus</td>
</tr>
<tr>
<td></td>
<td>Apo-CII deficiency</td>
</tr>
</tbody>
</table>

Apo-B100, apolipoprotein B100; LPL, lipoprotein lipase.

Clinical Features

Heterozygosity for FH occurs with a frequency of about 1 in 500 in the population and is found in many ethnic groups. Typically, the plasma cholesterol concentration is higher than 7.8 mmol/L (300 mg/dL), and the LDL-C concentration is higher than 6.5 mmol/L (250 mg/dL). Plasma triglyceride concentrations are not elevated. The typical plasma lipid abnormalities are characterized by increased plasma cholesterol, slightly increased triglyceride, and decreased HDL cholesterol concentrations. FH is associated with a type IIa pattern, reflecting increased concentrations of -migrating LDLs.

The characteristic physical finding in approximately 75% of affected subjects is the presence of tendon xanthomas (Fig. 34-26C and D). Xanthomas of the Achilles tendon can cause recurrent episodes of Achilles tendinitis. These xanthomas may be subtle and apparent only as a thickening of the tendon (see Fig. 34-26C). Other common physical findings include xanthelasma (see Fig. 34-26A; see color section) and premature arcus corneae (i.e., in persons younger than 40 years). A minority of affected subjects have no physical findings. Premature coronary artery disease is common, the average age of onset of coronary disease is 45 years in men and 55 years in women.

Homozgyosity for FH is rare, occurring at a frequency of about 1 in 10^6 in the population (i.e., 250 persons in the United States population). These subjects come to clinical attention early in life because of marked hypercholesterolemia or premature CHD. Typical plasma cholesterol concentrations range from 15.5 mmol/L (600 mg/dL) to 25.9 mmol/L (1000 mg/dL), and LDL-C concentrations range from 14.2 mmol/L (550 mg/dL) to 24.6 mmol/L (950 mg/dL). In addition to the xanthelasma and tendon xanthomas found in heterozygotes, FH is associated with a type IIa pattern, reflecting increased concentrations of -migrating LDLs.

Table 34-7 -- Major Genetic Hyperlipoproteinemas Resulting from Single-Gene Mutations

<table>
<thead>
<tr>
<th>Disorder</th>
<th>Mutant Gene</th>
<th>Inheritance</th>
<th>Estimated Population Frequency</th>
<th>Lipoprotein Pattern</th>
<th>Xanthomas</th>
<th>Pancreatitis</th>
<th>Premature Vascular Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Familial LPL deficiency</td>
<td>LPL</td>
<td>Autosomal recessive</td>
<td>1/10^6</td>
<td>I, V</td>
<td>Eruptive</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Familial apo-CII deficiency</td>
<td>Apo-CII</td>
<td>Autosomal recessive</td>
<td>1/10^6</td>
<td>I, V</td>
<td>Eruptive</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Familial hypercholesterolemia</td>
<td>LDL receptor</td>
<td>Autosomal dominant</td>
<td>1/500 (heterozygous); Ila (rarely IIb) Tendon; xanthelasma</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Familial hypercholesterolemia</td>
<td>Apo-B</td>
<td>Autosomal dominant</td>
<td>1/1000</td>
<td>Ila</td>
<td>Tendon</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Familial type III hyperlipoproteinemia</td>
<td>Apo-E</td>
<td>Autosomal recessive (rarely dominant)</td>
<td>1/10000</td>
<td>III</td>
<td>Palmar; tuberous</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Familial combined hyperlipoproteinemia</td>
<td>Unknown</td>
<td>Autosomal dominant</td>
<td>1/100</td>
<td>Ila, IIb, IV (rarely V)</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Familial hypertriglyceridemia</td>
<td>Unknown</td>
<td>Autosomal dominant</td>
<td>Uncertain</td>
<td>IV (rarely V)</td>
<td>+(7)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Apo, apolipoprotein; LDL, low-density lipoproteins; LPL, lipoprotein lipase.

Table 34-8 -- Clinical Disorders Associated with Secondary Hyperlipidemia

<table>
<thead>
<tr>
<th>Disorder</th>
<th>Lipoprotein Type</th>
<th>Elevated Plasma Lipoprotein</th>
<th>Proposed Mechanism</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diabetes mellitus</td>
<td>IV, V</td>
<td>VLDL, chylomicrons</td>
<td>Increased VLDL production; decreased VLDL catabolism</td>
<td>See text</td>
</tr>
<tr>
<td>Hypothyroidism</td>
<td>Ila (rarely III)</td>
<td>LDL (rarely -VLDL)</td>
<td>Decreased LDL clearance</td>
<td>See text</td>
</tr>
<tr>
<td>Estrogen therapy</td>
<td>IV (rarely V)</td>
<td>VLDL</td>
<td>Increased VLDL production (especially in genetically predisposed)</td>
<td>See text</td>
</tr>
<tr>
<td>Glucocorticoid therapy</td>
<td>Ila or IIb</td>
<td>VLDL, LDL</td>
<td>Increased VLDL production with conversion to LDL</td>
<td>247249</td>
</tr>
</tbody>
</table>

Apolipoprotein B100, apolipoprotein CII; LPL, lipoprotein lipase.
**Nonendocrine**

<table>
<thead>
<tr>
<th>Alcohol</th>
<th>IV (rarely V)</th>
<th>VLDL (rarely chylomicrons)</th>
<th>Increased VLDL production (especially in genetically predisposed)</th>
<th>See text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nephrotic syndrome</td>
<td>Ila or Iib</td>
<td>VLDL, LDL</td>
<td>Increased VLDL production</td>
<td>See text</td>
</tr>
<tr>
<td>Uremia</td>
<td>IV</td>
<td>VLDL</td>
<td>Decreased VLDL clearance</td>
<td>259</td>
</tr>
<tr>
<td>Biliary obstruction or cholestasis</td>
<td></td>
<td>LP-X</td>
<td>Diversion of biliary cholesterol and phospholipid into circulation</td>
<td>260</td>
</tr>
<tr>
<td>Hepatitis</td>
<td>IV</td>
<td>VLDL</td>
<td>Decreased LCAT</td>
<td>261, 262</td>
</tr>
<tr>
<td>Systemic lupus erythematosus</td>
<td>I</td>
<td>Chylomicrons</td>
<td>Antibodies bind heparin and thereby decrease LPL activity</td>
<td>263</td>
</tr>
<tr>
<td>Monoclonal gamopathy</td>
<td>Ila, III, IV</td>
<td>VLDL, IDL, LDL</td>
<td>Antibodies bind lipoproteins and interfere with catabolism</td>
<td>264, 265</td>
</tr>
</tbody>
</table>

LCAT, lecithin:cholesterol acyltransferase; LDL, low-density lipoproteins; LPL, lipoprotein lipase; VLDL, very-low-density lipoproteins.


Homozygous individuals frequently have planar xanthomas, which are almost unique to this disorder and almost always noticed by age 6. These xanthomas are raised plaques of cholesterol deposits that occur in the skin at areas of trauma, such as the elbows and knees. Symptoms of CHD may occur before age 10, and, if not treated, these homozygous individuals usually die from myocardial infarction by age 20. Myocardial infarction has been reported as early as age 18 months. Homozygotes are also susceptible to both valvular and supravalvular aortic stenosis.

**Origin and Pathogenesis**

FH is an autosomal dominant disorder caused by mutations in the LDL receptor gene. Many different types of mutations have been described, including null mutations or nonsense mutations that affect the production of a functional protein, mutations that affect the ability of the receptor to bind to its ligands on lipoproteins, and mutations in which receptors bind LDL normally but are unable to internalize the lipoprotein. A milder phenotype occurs when the ability to bind LDL is impaired but not absent. Different LDL receptor mutations occur in different ethnic groups; for example, there is an increased prevalence (about 60%) of a large deletion mutation in French Canadians with heterozygous FH.

The lack of LDL receptors impairs the clearance of lipoproteins that rely on the LDL receptor for this purpose; these include LDLs, in which apo-B100 is the ligand, and remnant lipoproteins (IDLs) that are cleared by apo-E. This results in a twofold to threefold increase in the plasma cholesterol concentration in heterozygotes and in a threefold to sixfold increase in homozygotes. The high levels of LDL in the plasma are taken up by scavenger receptors on macrophages in a nonsaturable manner, possibly after the LDL undergoes oxidative modification.

In one study, heterozygous FH accounted for 4% of men who survived myocardial infarction before age 60. The differential diagnosis includes familial defective apo-B100, which has many of the same phenotypic characteristics, including tendon xanthomas. The pattern of isolated high LDL-C also occurs in the more common disorder of polygenic hypercholesterolemia, but tendon xanthomas are not usually a feature of the latter.

The diagnosis of FH is primarily a clinical diagnosis because tests to detect one of the many LDL receptor gene mutations or to demonstrate diminished LDL receptor function are performed only in specialized research laboratories. The diagnosis can be confirmed in the laboratory by culturing skin fibroblasts and demonstrating a reduced ability of LDL to bind to receptors on the cells. As a consequence, cholesterol accumulates in tissue macrophages in the arterial wall, tendons, and skin and causes the pathologic processes observed in these tissues.

**Diagnosis**

The diagnosis of heterozygous FH is suggested by the presence of high plasma levels of total cholesterol and LDL-C, normal plasma triglycerides, tendon xanthomas, and a family history of premature CHD. Up to 25% of subjects do not have xanthomas. Heterozygous FH should be suspected in any person with premature heart disease. In one study, heterozygous FH accounted for 4% of men who survived myocardial infarction before age 60. The differential diagnosis includes familial defective apo-B100, which has many of the same phenotypic characteristics, including tendon xanthomas. The pattern of isolated high LDL-C also occurs in the more common disorder of polygenic hypercholesterolemia, but tendon xanthomas are not usually a feature of the latter.

The diagnosis of FH is primarily a clinical diagnosis because tests to detect one of the many LDL receptor gene mutations or to demonstrate diminished LDL receptor function are performed only in specialized research laboratories. The diagnosis can be confirmed in the laboratory by culturing skin fibroblasts and demonstrating a reduced ability of LDL to bind to receptors on the cells. As a consequence, cholesterol accumulates in tissue macrophages in the arterial wall, tendons, and skin and causes the pathologic processes observed in these tissues.

**Treatment**

Treatment of heterozygous FH consists of a diet low in total and saturated fat (approximately 20% and 6% of calories, respectively) and low in cholesterol (<2.6 mmol/day [100 mg/day]) and drug therapy. Dietary modifications usually result in only minor decreases in the plasma cholesterol levels (5% to 15%). With the development of more potent HMG-CoA reductase inhibitors, adequate cholesterol lowering in these patients can sometimes be achieved by a single therapeutic drug. However, combinations of two or three drugs are often needed to reduce plasma cholesterol to desired levels. Effective drug combinations usually include
low doses of bile acid sequestrants together with HMG-CoA reductase inhibitors or niacin or all three agents combined. Bile acid sequestrants and HMG-CoA reductase inhibitors both work by depleting cholesterol from hepatic cells (see later discussion), thereby causing increased expression of functional LDL receptors (from the normal allele) on cells, which in turn lowers the plasma cholesterol.

Because high Lp(a) concentrations appear to be an adverse risk factor for patients with heterozygous FH, the plasma Lp(a) level should be determined; if it is elevated, niacin should be considered in the drug regimen because this agent can lower plasma Lp(a) levels. On the basis of studies in animal models of FH, antioxidant agents may be of therapeutic benefit for preventing CHD, but this has not been demonstrated directly in humans. Probucol treatment has resulted in xanthoma regression, ileal bypass surgery, which, like bile acid sequestrants, causes decreased reabsorption of bile acids from the gut, may be considered in patients who cannot tolerate lipid-lowering drugs.

The age at which treatment should begin in heterozygous FH is somewhat controversial. On the one hand, development of atherosclerosis in these subjects is a long process that begins early in life, and one could argue that treatment should begin during the early stages of lesion development. On the other hand, CHD is usually not symptomatic until the third or fourth decade of life in men and 10 years later in women. Because established CHD is reversible, one could argue that medicines can be withheld until after age 25 in men or age 35 in women. A rational approach may be to use diet therapy and bile acid sequestrants, which do not have systemic toxicity, in the early years and to add more potent drug combinations later. The presence of additional risk factors (e.g., high plasma Lp(a) levels, low plasma HDL-C levels, smoking) in an affected subject is an indication for more aggressive treatment at a young age.

Unless the causative mutation is such that there is some residual LDL binding, drug therapy for FH homozygotes is usually ineffective for lowering plasma cholesterol, except with high doses of atorvastatin and simvastatin. However, the most effective means of therapy in these patients is selective removal of LDL from the plasma by extracorporeal pheresis combined with LDL immunoadsorption performed every 1 to 3 weeks. Experimental therapies include liver transplantation, which provides functional LDL receptors, portacaval shunting, and gene therapy.
Familial Defective Apolipoprotein B100

Familial defective apo-B100 is a relatively common disorder caused by a mutation in apo-B100, the ligand for the LDL receptor, that results in high plasma LDL and total cholesterol levels and increased susceptibility to CHD. It is phenotypically similar to FH.

Clinical Features

This disorder occurs with a frequency of 1 in 500 to 1 in 750 in white persons with hypercholesterolemia. The prevalence of familial defective apo-B100 was 0.08% in an ethnically diverse, unselected population. Familial defective apo-B100 had not been described in nonwhite populations until 1993, when it was detected in an individual of Chinese ancestry. The clinical features of heterozygous familial defective apo-B100 overlap extensively with those of heterozygous FH and include isolated elevations of plasma LDL-C (type IIa pattern), tendon xanthomas, xanthelasmas, and premature CHD.

Although there is extensive overlap, familial defective apo-B100 is usually milder in its manifestations than FH. Subjects who are homozygous for the familial defective apo-B100 mutation also appear to have a milder clinical phenotype than FH homozygotes. In one report, a 66-year-old man and his 69-year-old sister with homozygous familial defective apo-B100 had total plasma cholesterol levels of 9.6 mmol/L (370 mg/dL) and 11.9 mmol/L (460 mg/dL), respectively, and only the man had evidence of CHD. Presumably, the less severe phenotype is related to the fact that the binding of apo-B to LDL receptors is defective but not totally absent in familial defective apo-B100, whereas the apo-E-mediated clearance of remnant particles, which is impaired in FH, is normal in persons with familial defective apo-B100.

Origin and Pathogenesis

Familial defective apo-B100 is caused by a mutation in apo-B100 that impairs its ability to bind to the LDL receptor. To date, a single mutation, the substitution of glutamine for arginine at amino acid 3500, accounts for almost all cases of familial defective apo-B100. Apo-B allele haplotype analysis of DNA from affected subjects has indicated that almost all cases can be traced back to an original founder. Only after extensive screening was this mutation detected in persons with a different apo-B haplotype for the allele carrying the mutation, one of whom was of Chinese ancestry and one in a kindred from Germany.

The mutation located at apo-B amino acid 3500 disrupts the conformation of the protein in the receptor-binding domain and reduces receptor binding of LDL from heterozygotes to levels that are about one third of normal in tissue culture assays. Isolation of the binding-defective LDL from affected subjects demonstrated that it binds with 4% to 9% of normal activity to LDL receptors. Decreased affinity of the defective apo-B100 for its receptor delays the clearance of LDL from the plasma (by about 50%) and leads to elevation of plasma LDL-C levels. Defective LDL particles accumulate in the plasma in increased proportions relative to normal LDL. A second mutation at amino acid 3500 (tryptophan for arginine) has been reported.

Another mutation located near the receptor-binding region of the apo-B molecule (a substitution of cysteine for arginine at amino acid 3531) also impairs binding of apo-B to the LDL receptor. This mutation decreases LDL binding to LDL receptors by 35% to 40% in tissue culture assays and is associated with moderate elevations in plasma LDL-C levels.

Diagnosis

As in heterozygous FH, the diagnosis of familial defective apo-B100 is suggested by the presence of increased plasma LDL-C and normal triglyceride levels, especially in the presence of tendon xanthomas and a family history of premature CHD. Without specialized testing, however, familial defective apo-B100 is clinically indistinguishable from FH. Because familial defective apo-B100 is caused primarily by one mutation, in contrast to the many mutations that cause FH, it is possible to screen easily for the familial defective apo-B100 mutation using a polymerase chain reaction-based assay of genomic DNA isolated from blood. This test is available only in specialized laboratories.

Treatment

Treatment of familial defective apo-B100 is similar to that of heterozygous FH and consists of a low-fat, low-cholesterol diet and a combination drug regimen. Drugs that either decrease LDL production (e.g., niacin) or increase the expression of LDL receptors to facilitate clearance of the normal apo-B100-containing particles effectively lower the plasma LDL-C level. In two patients with homozygous familial defective apo-B100 whose LDL had receptor-binding affinities 10% to 20% of normal, treatment with HMG-CoA reductase inhibitors markedly reduced plasma cholesterol levels.

Family members at risk should also be screened for the dominant mutation.
Familial Combined Hyperlipidemia

Originally described in 1973, familial combined hyperlipidemia is a common disorder of unknown genetic cause associated with elevations of plasma cholesterol and triglyceride levels and increased susceptibility to CHD. It is inherited as an autosomal dominant trait. The phenotype of familial combined hyperlipidemia overlaps with and may be the same as that in familial hyperapobetalipoproteinemia, in subjects with plasma elevations of small, dense LDL particles, and in subjects with syndrome X or the metabolic syndrome, a disorder that includes insulin resistance, increased plasma levels of small, dense LDL, elevated plasma triglycerides, and low plasma HDL levels.

Clinical Features

The features of familial combined hyperlipidemia include moderate elevations of plasma cholesterol, triglycerides, or both within subjects of an affected kindred, corresponding to lipoprotein pattern type IIa, IIb, or IV on plasma electrophoresis. The predominant lipid abnormality may vary in a single person over time or among affected family members; the variable phenotype in this disorder led in part to the decreased utility of pattern typing of hyperlipoproteinemia by plasma electrophoresis in the clinical evaluation. Levels of HDL-C are often moderately decreased, especially in the setting of increased plasma triglycerides.

Although it was originally thought that lipid abnormalities usually develop after puberty, it is now known that the phenotype can be detected in children. Neither xanthomas nor xanthelasma is a feature of familial combined hyperlipidemia. Associated metabolic disturbances may include glucose intolerance, obesity, and hyperuricemia. Premature CHD is a common feature; in one study of male survivors of myocardial infarction, familial combined hyperlipidemia was found in 11.3% of those younger than 60 years of age. CHD is often present in men by age 50.

Origin and Pathogenesis

Although familial combined hyperlipidemia is a common disorder (estimated prevalence, 0.5% to 2.0%), neither its genetic cause nor its metabolic pathogenesis is clear. Given the dominant pattern of inheritance, it was initially presumed that familial combined hyperlipidemia is caused by a single gene defect. Now it is believed that multiple genes may be involved, with one or two genes playing a large role and other genes acting as modifiers. A candidate gene locus was identified in syntenic chromosomal regions in mice and humans, suggesting that an important contributing gene may reside at this locus. The phenotype of this disorder has also been mapped to loci on chromosomes 11, and 19, and several different loci may underlie the expression of the phenotype. Heterozygous LPL deficiency may constitute a subset of familial combined hyperlipidemia.

A significant problem with mapping this disorder is the difficult task of assigning phenotypes to individuals because of the fluctuating and indistinct clinical features. Familial hyperapobetalipoproteinemia, a disorder characterized by high plasma levels of apo-B and normal plasma cholesterol levels, overlaps with the phenotype of familial combined hyperlipidemia, as do both the familial syndrome characterized by small, dense LDL and syndrome X characterized by insulin resistance and other metabolic abnormalities. The metabolic defect that leads to the hypercholesterolemia or hypertriglyceridemia, or both, is also unclear, but overproduction of apo-B may be a contributing factor; apo-B overproduction can result in elevations in plasma VLDL, LDL, or both. Similarly, the pathogenesis of the low HDL in this disorder remains unclear. Low HDL-C is commonly associated with hypertriglyceridemia. This finding could be related either to decreased substrate for HDL formation because of impaired catabolism of the apo-B-containing lipoproteins or to enhanced CETP-mediated cholesteryl ester transfer from HDL to the apo-B-containing lipoproteins.

Diagnosis

Familial combined hyperlipidemia should be suspected in subjects with moderate hypertriglyceridemia or moderate hypercholesterolemia (lipoprotein type IIa, IIb, or IV), or both, especially in the setting of a family history of premature CHD. Xanthomas are not a feature of this disorder. Low plasma HDL-C, obesity, insulin resistance, and hyperuricemia are often present. The diagnosis is a clinical one; it requires demonstration of the clinical phenotype in the affected subject and family members and exclusion of other primary or secondary disorders. Secondary disorders that produce a similar phenotype include diabetes mellitus, nephrotic syndrome, and occasionally hypothyroidism. Many patients with diabetes mellitus and combined hyperlipidemia may have a genetic susceptibility related to inheritance of familial combined hyperlipidemia.

Treatment

Weight reduction and dietary treatment can help correct metabolic abnormalities, such as obesity and insulin resistance, that contribute to the hyperlipidemia. Drug therapy should be directed at the predominant lipid abnormality. For example, plasma elevations of total cholesterol and LDL-C can be treated with HMG-CoA reductase inhibitors, niacin, or bile acid sequestrants. Of these, HMG-CoA reductase inhibitors may be preferable because niacin can cause or worsen glucose intolerance and hyperuricemia and bile acid sequestrants can cause hypertriglyceridemia. Gemfibrozil can lower triglyceride and raise HDL-C levels, and it reduced the incidence of coronary events in the Helsinki Heart Study. Low HDL-C levels can be treated with niacin, gemfibrozil, or HMG-CoA reductase inhibitors. Because familial combined hyperlipidemia is associated with premature CHD, affected family members should be identified.
Type III Hyperlipoproteinemia (Familial Dysbetalipoproteinemia)

Type III hyperlipoproteinemia, or familial dysbetalipoproteinemia, is an uncommon disorder of lipoprotein metabolism characterized by moderate to severe hypertriglyceridemia and hypercholesterolemia caused by the accumulation of cholesterol-rich remnant particles in the plasma. Premature peripheral vascular disease and coronary artery disease are common. The cause is mutations in apo-E that result in defective binding to lipoprotein receptors. The disorder is associated with the apo-E2 isoform (described previously) and in most instances is inherited as an autosomal recessive trait that requires a secondary exacerbating metabolic factor (either genetic or environmental) for expression of the phenotype. Several rare apo-E mutations result in the dominant expression of the disorder.

Clinical Features

Type III hyperlipoproteinemia is usually diagnosed in adulthood and is rarely detected in persons younger than 20 years, with the exception of those with the rare autosomal dominant apo-E mutations. The disorder is more common in men and is usually not manifested in women until after menopause. It is characterized by moderately severe elevations in plasma triglyceride and cholesterol levels; typically, these values range from 3.4 to 4.5 mmol/L (300 to 400 mg/dL) and 7.8 to 10.3 mmol/L (300 to 400 mg/dL), respectively. Concentrations of HDL-C are normal, and LDL-C is almost always reduced.

Xanthomas are present in more than half of affected subjects. The presence of palmar xanthomas, which are planar xanthomas in the palmar creases (see Fig. 34-26F), is virtually pathognomonic for this disorder. Tuberous or tuberoeruptive xanthomas (see Fig. 34-26E; see color section) are also common but are less specific for this disorder. Tendon xanthomas and xanthelasma occur in some patients. Premature vascular disease is common, and peripheral vascular disease occurs in about 10% of patients with type III hyperlipoproteinemia. Type III hyperlipoproteinemia accounts for 0.2% to 1.0% of lipid disorders associated with myocardial infarction in persons younger than 60 years. Coexisting metabolic conditions that exacerbate the phenotype of type III hyperlipoproteinemia, such as obesity, alcohol consumption, diabetes mellitus, and hypothyroidism, are often present.

Origin and Pathogenesis

Type III hyperlipoproteinemia is caused by the plasma accumulation of cholesterol-rich remnants of VLDL, IDL, and chylomicrons. The clearance defect is caused by mutant apo-E that binds defectively to remnant receptors, including LDL receptors (discussed in “Roles of Apolipoprotein E in Lipid Metabolism”). The remnants that accumulate have lost much of their triglyceride through LPL-mediated triglyceride hydrolysis and therefore are cholesterol rich. The predominant remnant particles are termed -VLDL and can be isolated in the VLDL ultracentrifugation density range (<1.006 g/mL). In contrast to normal VLDLs, which migrate as pre-particles, these remnants are characterized by -migration on agarose gel electrophoresis.

Homozygosity for the apo-E2 isoform occurs at a frequency of about 1 in 100 in the general population. Despite this high frequency, the type III hyperlipoproteinemia phenotype is relatively rare: the dyslipidemia develops in about 1 in 10 to 1 in 100 apo-E2 homozygotes (overall prevalence of 1 in 10,000). Manifestation of the dyslipidemia appears to require the presence of a secondary factor, such as a metabolic condition that contributes to the phenotype of impaired remnant clearance. Such conditions can be caused by lipoprotein overproduction syndromes, such as obesity, diabetes mellitus, or alcohol consumption, or by conditions that further impair lipoprotein clearance, such as hypertriglyceridemia. In conditions characterized by VLDL overproduction, the increased generation of remnant particles through catabolism of VLDL overwhelms the ability of the remnant receptors to clear them from the plasma. Because the manifestations are so sensitive to conditions that increase hepatic lipoprotein production, this disorder is also one of the most sensitive to therapeutic modalities directed at decreasing hepatic lipoprotein production, such as dietary modifications, weight loss, and alcohol cessation.

In addition to homogygosity for the apo-E2 isoform, six mutations in the apo-E gene are known to lead to the type III hyperlipoproteinemia phenotype in an autosomal dominant fashion. In these dominant disorders, the phenotype is present at an early age and does not require a coexisting metabolic condition as an exacerbating factor. These rare dominant disorders are of great interest because they show that different mutations in a single protein can give rise to either recessively or dominantly inherited phenotypes.

When a given apo-E mutation interacts with the HSPG/LRP pathway is believed to determine whether the mutation gives rise to a dominant or recessive phenotype. For example, in recessive expression of the disorder, as occurs with apo-E2 homozygosity, apo-E2 interacts poorly with the LDL receptor but almost normally with the HSPG/LRP pathway so that in the absence of exacerbating secondary factors remnant lipoproteins are cleared efficiently. However, in the presence of a secondary factor that overwhelms or even slightly impairs the normal pathways, apo-E2 cannot clear remnants efficiently. In dominant expression of the disorder, apo-E interacts poorly with both the LDL receptor and the HSPG/LRP pathway, and remnant lipoprotein clearance is impaired even in the heterozygous state. Another contributing factor is the particular lipoprotein fraction with which the apo-E molecule preferentially associates (discussed in “Roles of Apolipoprotein E in Lipid Metabolism”).

The accumulation of cholesterol-rich remnant lipoproteins in the plasma leads to the deposition of cholesterol in tissue macrophages, which avidly bind and take up -VLDL. The deposition of -VLDL-derived cholesterol in macrophages leads to foam cell formation and accumulation, manifested as skin xanthomas and as atherothrombotic vascular disease. In addition to the frequent occurrence of CHD, -VLDL hyperlipidemia appears to cause a disproportionately high incidence of peripheral vascular disease. The onset of premature CHD occurs at about 40 years of age in men and at about 50 years of age in women.

Diagnosis

The diagnosis of type III hyperlipoproteinemia should be suspected in persons with moderately severe elevations in both plasma triglyceride and cholesterol concentrations. Typically the cholesterol and triglyceride levels are in the range of 3.4 to 4.5 mmol/L (300 to 400 mg/dL) and 7.8 to 10.3 mmol/L (300 to 400 mg/dL), respectively. Because this disorder is most commonly recessive, there is often no family history of hyperlipidemia or premature CHD. The presence of palmar or tuberous xanthomas makes the diagnosis highly likely. In the absence of palmar or tuberous xanthomas, the specific diagnosis is more difficult and requires specialized testing. If available, a measurement of the VLDL cholesterol level allows detection of cholesterol-rich remnant particles. The VLDL cholesterol/triglyceride ratio is a useful screen; in type III hyperlipoproteinemia, this ratio is usually greater than 0.3 (when hyperlipidemia is present). The normal VLDL cholesterol/triglyceride ratio is typically about 0.2 (i.e., the VLDL cholesterol concentration is about 20% of the plasma triglyceride level, as estimated by the Friedewald formula; see later discussion). The ratio is elevated because -VLDL remnants are rich in cholesterol and cause the VLDL fraction to be cholesterol rich. Electrophoresis of plasma samples on agarose gel typically demonstrates a broad band in the migrating lipoprotein region (type III pattern), hence the names broad- disease, dysbetalipoproteinemia, and type III hyperlipoproteinemia. Patients suspected of having type III hyperlipoproteinemia can be evaluated for apo-E2 homozygosity either by isoelectric focusing of plasma (see Fig. 34-9) or by apo-E genotyping of DNA obtained from leukocytes. The other rare dominant mutations in apo-E can be diagnosed only in specialized laboratories.

Treatment

Because type III hyperlipoproteinemia is greatly influenced by coexisting metabolic conditions, a vigorous search should be made to identify and treat obesity, alcohol consumption, diabetes mellitus, and hypothyroidism. If such conditions can be identified and treated successfully, the lipid abnormalities can often be resolved and plasma lipids returned to normal without the use of drug therapies. Type III hyperlipoproteinemia associated with hypothyroidism, in particular, responds dramatically to thyroid hormone replacement therapy. Dietary therapy should be aimed at restricting total fat, saturated fat, and cholesterol (Step 1 or Step 2 American Heart...
Association diet) and at caloric restriction to reduce weight if appropriate. Treatment with estrogen should be considered in postmenopausal women because it can dramatically diminish the hyperlipidemia, probably by stimulating LDL receptor-mediated remnant particle clearance.

If dietary therapy and treatment of coexisting metabolic conditions yield unsatisfactory results, drug therapy should be initiated using either niacin, fibric acid derivatives, or HMG-CoA reductase inhibitors, all of which are effective in the treatment of this disorder. By decreasing VLDL synthesis and secretion, niacin lowers triglyceride and VLDL cholesterol levels by about 40% and LDL-C levels by 20% and can raise HDL-C concentrations by 20%. The fibric acid derivatives, gemfibrozil and clofibrate, can also lower triglyceride and VLDL cholesterol levels. Through their ability to increase LDL receptor levels, the HMG-CoA reductase inhibitors can also lower the plasma cholesterol levels. Cases that are refractory to treatment with individual drugs can be treated with combinations of fibric acid derivatives and HMG-CoA reductase inhibitors, although this combination should be used cautiously because of the risk of myopathy. Because the disorder is associated with premature vascular disease, first-degree relatives such as siblings (or offspring if the affected person's spouse has an apo-E2 allele) should be screened.
Lipoprotein Lipase Deficiency

LPL deficiency is a rare, recessive disorder that results from mutations in the LPL gene. These abnormalities cause LPL deficiency and severe hypertriglyceridemia by blocking the clearance of triglyceride-rich lipoproteins from the plasma. Massive accumulations of these proteins in the plasma, known as the chylomicronemia syndrome, can be accompanied by severe clinical manifestations, including pancreatitis.

Clinical Features

LPL deficiency is usually recognized in infancy or childhood as a chylomicronemia syndrome, which consists of marked hypertriglyceridemia associated with recurrent abdominal pain or pancreatitis, which can be life-threatening. A syndrome of recurrent abdominal pain and severe hypertriglyceridemia but without overt pancreatitis or elevated serum amylase concentrations also exists. The pain syndrome is associated with triglyceride levels of more than 22.6 mmol/L (2000 mg/dL) and abates with triglyceride lowering. When plasma triglyceride levels exceed 22.6 mmol/L (2000 mg/dL), the findings include eruptive xanthomas and lipemia retinalis. The plasma may be visibly lipemic; plasma that has been refrigerated overnight can demonstrate chylomicrons (type I pattern) with a cream-like layer on the top, VLDL (type IV pattern) with a turbid plasma infranatant, or both (type V pattern).

Chylomicrons can be assumed to be present if the fasting triglyceride concentration is greater than 11.3 mmol/L (1000 mg/dL). The degree of fasting chylomicronemia is, in large part, determined by the dietary fat intake. Accumulation of chylomicrons in tissue reticuloendothelial cells can lead to hepatomegaly and splenomegaly. Chylomicronemia can also cause neurologic manifestations and dyspnea. CHD is not a prominent feature. Massive triglyceride elevations can occupy significant plasma volume and lead to artificially low measurements of serum electrolytes, such as sodium (pseudohyponatremia), if the electrolyte measurements are not corrected for serum water.

Subjects who are heterozygous for LPL deficiency may have reduced LPL activity and often have milder hypertriglyceridemia, increased VLDL cholesterol levels, and decreased HDL-C levels. This phenotype, which may constitute a subset of familial combined hyperlipidemia, is exacerbated by age and obesity.

Origin and Pathogenesis

Complete LPL deficiency results from homozygosity or compound heterozygosity for mutations in the LPL gene that lead to the absence or inactivation of the LPL protein. A variety of mutations have been described. The combined frequency of homozygosity and compound heterozygosity is approximately 1 in 10^5 persons. Patients who are heterozygous for LPL deficiency have half-normal LPL activities and hypertriglyceridemia may develop in the presence of secondary factors. Heterozygous LPL deficiency occurs at a frequency of 1 in 500 persons in the general population but as high as 1 in 40 in areas of Quebec. In the absence of functional LPL, triglyceride-rich lipoproteins accumulate in the plasma. Because of the impaired clearance of these lipoproteins, plasma triglyceride levels are especially sensitive to dietary fat intake. Whereas chylomicrons are normally cleared from the plasma within 8 hours after eating, clearance in patients with this disorder may take days. The triglyceride-rich particles infiltrate organs, where they are taken up by reticuloendothelial cells. The accumulation of these lipoproteins in the pancreas can cause pancreatitis, presumably resulting from chemical irritation by fatty acids and lyssolecithin liberated by the action of pancreatic lipases.

Diagnosis

LPL deficiency should be suspected in children with pancreatitis or recurrent abdominal pain. Eruptive xanthomas are often present. LPL deficiency should also be suspected in any person who has lipemic serum after a 12-hour fast. Plasma triglyceride levels are usually higher than 11.3 mmol/L (1000 mg/dL) and may be considerably higher, depending on the dietary fat intake. Eruptive xanthomas, lipemia retinalis, and pancreatitis are usually apparent only when triglyceride levels are greater than 22.6 mmol/L (2000 mg/dL). Because LPL deficiency is a recessive disease, family history is usually uninformative except in siblings.

The definitive diagnosis is established by demonstrating the absence of lipase activity in plasma after heparin administration. Heparin, when infused intravenously, displaces LPL from its binding sites on HSPI in capillary endothelium and releases it into plasma, which can be assayed for lipase activity. LPL deficiency must be distinguished from deficiency of apo-CII, a cofactor for LPL, which is another cause of chylomicronemia. Because multiple mutations can result in LPL deficiency, specific mutations can be identified only by specialized laboratories.

Treatment

During the initial stages of treatment, a fat-free diet is provided until plasma triglycerides reach a safe level (e.g., <11.3 mmol/L [1000 mg/dL]). After the acute chylomicronemia syndrome has resolved, the mainstay of treatment is a diet that contains very small amounts of fat (e.g., <10% of calories or 20 to 25 g/day). Because medium-chain triglycerides, in contrast to long-chain triglycerides, are absorbed directly into the portal circulation and do not rely on chylomicron formation for hepatic uptake, they can provide a source of fat in the diet; however, these agents may cause hepatic toxicity. Fatsoluble vitamins should be supplemented. The goal of therapy is to maintain the plasma triglyceride level at less than 11.3 mmol/L (1000 mg/dL), which prevents further episodes of pancreatitis.

Drug therapy for primary LPL deficiency is largely ineffective; however, clofibrate, gemfibrozil, or niacin may lower VLDL production and prevent severe hypertriglyceridemia. Secondary causes of hypertriglyceridemia, such as diabetes mellitus or hypothyroidism, should be screened for and treated if present.
Apolipoprotein CII Deficiency

Apo-CII deficiency is a rare autosomal recessive disorder that occurs in fewer than 1 in 10^6 persons and causes a chylomicronemia syndrome similar to that in LPL deficiency. As in LPL deficiency, the features include pancreatitis or recurrent bouts of abdominal pain in children or young adults and lipemic serum after a 12-hour fast. Plasma triglyceride levels are usually severely elevated (>11.3 mmol/L [1000 mg/dL]), reflecting the accumulation of chylomicrons (type I pattern), VLDL (type IV pattern), or both lipoproteins (type V pattern) in the plasma. This hyperlipoproteinemia results from the lack of apo-CII, an activating cofactor for LPL, which causes a functional LPL deficiency. More than 10 mutations are known to cause apo-CII deficiency. The accumulation of triglyceride-rich particles results in a pathophysiologic process that is nearly identical to that described for LPL deficiency. Heterozygotes may have slightly elevated triglyceride concentrations, but they do not develop pancreatitis.

The diagnosis requires specialized testing to demonstrate that apo-CII is absent on electrophoresis of the plasma apolipoproteins or that the plasma is incapable of activating LPL in vitro. The treatment of apo-CII deficiency is identical to that of primary LPL deficiency with the exception that severe hypertriglyceridemia and pancreatitis in apo-CII-deficient patients can be treated with transfusions of plasma, which contains apo-CII.
Familial Hypertriglyceridemia

Familial hypertriglyceridemia is characterized by increased plasma concentrations of triglyceride-rich VLDLs, which cause elevations of plasma triglycerides but not plasma cholesterol levels.

Clinical Features

Subjects with familial hypertriglyceridemia typically have plasma triglyceride levels in the range of 2.3 to 5.6 mmol/L (200 to 500 mg/dL) and normal LDL-C levels. As the level of LDL-C considered to be normal has declined, the prevalence of individuals with isolated elevations of plasma triglycerides has also diminished. The hypertriglyceridemia is often associated with low plasma HDL-C levels. The elevated triglyceride levels are usually not evident until adulthood and may be exacerbated by secondary factors, including hypothyroidism, estrogen therapy, or alcohol ingestion. Such exacerbations can be associated with severe elevations of triglycerides (>11.3 mmol/L [1000 mg/dL]), placing subjects at risk for eruptive xanthomas and pancreatitis. However, xanthomas are usually not present. Obesity and insulin resistance are common. Although familial hypertriglyceridemia was originally thought to be associated with increased risk for CHD, this relation has been questioned. The association of decreased HDL-C levels with hypertriglyceridemia may contribute to increased CHD risk.

Origin and Pathogenesis

Familial hypertriglyceridemia appears to be caused by overproduction of VLDL triglycerides in the presence of near-normal apo-B production, which leads to the secretion of large, triglyceride-rich VLDL. Secondary disorders (e.g., insulin resistance) that lead to VLDL overproduction can exacerbate the syndrome. The low plasma HDL levels commonly found in hypertriglyceridemia are associated with enhanced fractional catabolism of apo-AI. The genetic defect in familial hypertriglyceridemia is unknown. It is likely that the genetic loci involved in causing hypertriglyceridemia in familial combined hyperlipidemia are also involved in causing isolated hypertriglyceridemia. Whether the large, triglyceride-rich VLDLs are atherogenic is unclear; an enhanced risk for premature CHD may be related to whether concomitant decreases in HDL-C occur.

Diagnosis

Familial hypertriglyceridemia should be suspected in individuals with increased plasma triglyceride levels and normal plasma cholesterol. The disorder can be diagnosed only if hypertriglyceridemia is found in half of the first-degree relatives at risk, and it can be difficult to distinguish from familial combined hyperlipidemia, which may also occur as isolated hypertriglyceridemia related to increased plasma VLDL (type IV pattern). Measurement of plasma lipid levels in children does not help to distinguish these disorders because lipid abnormalities are usually not present until after puberty in either disorder. The elevated VLDL levels can be observed as a cloudy appearance of the plasma after overnight refrigeration.

Treatment

In addition to dietary fat restriction, secondary disorders such as diabetes mellitus, estrogen administration, or alcohol intake should be screened for and treated. Drugs that lower triglyceride levels (e.g., niacin, gemfibrozil) may be useful. Because niacin can impair glucose tolerance, it should be used cautiously in patients with underlying insulin resistance.
Elevated Plasma Lp(a)

This disorder consists of elevations of modified LDL particles in the plasma, in which the apo-B protein of LDL is covalently bonded to apo(a). Apo(a) is a protein of unknown function that shares high sequence homology with plasminogen but is not catalytically active. Some studies suggest that elevated plasma Lp(a) concentrations are associated with an increased risk for CHD.

Clinical Features

There are no characteristic physical findings or lipoprotein patterns to suggest elevated plasma Lp(a) levels. Elevated Lp(a) (>30 mg/dL) may be suspected, however, in patients with premature CHD. Plasma Lp(a) concentrations are influenced by heredity and vary among different ethnic populations. For example, African populations have higher levels. Some data suggest that elevations of plasma Lp(a) may be atherogenic only in the presence of high concentrations of LDL. In support of this, elevated Lp(a) concentrations are a risk factor for development of CHD in subjects with FH. Thus, high Lp(a) levels may be viewed as a potent risk factor for CHD in predisposed individuals.

Origin and Pathogenesis

Apo(a) is found only in humans, nonhuman primates, and hedgehogs. Its function from an evolutionary perspective is unclear. Nevertheless, the presence of high plasma levels of apo(a) appears to have been selected for in certain populations. Apo(a) is attached by disulfide bonds to the apo-B protein of LDL. Plasma Lp(a) levels are in large part determined by heredity and appear to be related to the number of repeats of a so-called kringle motif in the protein. The larger isoforms that contain more kringle repeats are found in lower concentrations in the plasma, possibly related to impaired processing of these large forms for secretion by hepatocytes. The factors that control the production and clearance of Lp(a) are largely unknown. An increased susceptibility to atherosclerosis associated with high plasma levels of Lp(a) could be related to impaired fibrinolysis caused by competition for plasminogen receptors, to effects on smooth muscle proliferation, or to unknown factors.

Diagnosis

The diagnosis of high plasma Lp(a) levels is made by specific assays of plasma for apo(a) or intact Lp(a) particles; care should be taken to ensure that samples are collected and stored appropriately and that the assay does not detect plasminogen. Lp(a) levels greater than 30 mg/dL are considered high.

Treatment

Of the hypolipidemic drugs currently available, only niacin appears to lower plasma Lp(a) levels. Treatment with niacin (4 g/day) lowers Lp(a) levels by 35% to 40%. In postmenopausal women, estrogen therapy lowers Lp(a) levels by about 20%.
Polygenic Hypercholesterolemia

Hypercholesterolemia is defined as a cholesterol value that exceeds the 95th percentile for the population. A study by Goldstein and co-workers suggested that about 10% of patients with hypercholesterolemia have familial combined hyperlipidemia and about 5% have FH. In a large proportion of the remaining 85%, the cause of the hypercholesterolemia is unknown but probably largely involves combinations of genetic and environmental factors. Other genetic factors that contribute to hypercholesterolemia may involve physiologic processes that influence cholesterol absorption, bile acid metabolism, or intracellular cholesterol metabolism. Polygenic hypercholesterolemia is diagnosed by excluding other primary genetic causes, by the absence of tendon xanthomas, and by demonstrating that hypercholesterolemia is present in no more than 10% of first-degree relatives. The hypercholesterolemia is treated according to NCEP guidelines (described later).
Sporadic Hypertriglyceridemia

As with high plasma cholesterol, unknown genetic and environmental factors can result in elevated plasma triglyceride levels. This so-called sporadic hypertriglyceridemia can be distinguished from familial syndromes by the absence of hypertriglyceridemia in relatives. The condition is treated with dietary fat restriction, treatment of secondary conditions that exacerbate hypertriglyceridemia, and drugs that lower triglyceride levels, such as niacin or gemfibrozil.
PRIMARY DISORDERS OF HIGH-DENSITY LIPOPROTEIN METABOLISM

Several genetic disorders can result in decreased or increased plasma levels of HDL-C (Table 34-9).

**Familial Hypoalphalipoproteinemia**

The autosomal dominant disorder familial hypoalphalipoproteinemia is manifested by low plasma HDL-C levels and an increased risk for premature CHD. The diagnosis is suggested by HDL-C levels that are less than the 10th percentile (<0.8 mmol/L [30 mg/dL] in men or 1.0 mmol/L [40 mg/dL] in premenopausal women). There are no characteristic physical findings, but there is often a family history of low HDL-C levels and premature CHD. The genetic and metabolic defects that lead to low plasma HDL levels are unknown, but it appears that up to 50% of low HDL-C levels can be linked to the hepatic lipase or apo-AI/apo-CIII/apo-AIV gene locus. The lack of HDL in the plasma accelerates the development of atherosclerosis, presumably because reverse cholesterol transport or other protective effects of HDL are impaired.

**Drug therapy** should be aimed at raising the plasma HDL concentration or lowering the plasma LDL concentration, or both. Treatments to raise HDL levels include estrogen therapy for postmenopausal women, aerobic exercise, and the drug niacin or gemfibrozil. Because increasing HDL-C is difficult with current therapies, lowering the LDL-C concentration may be the most effective approach, with the premise that the lowering of atherogenic LDL particles can overcome some, if not all, of the effects of low HDL.

Low HDL levels have been found in specific ethnic groups. For example, Southeast Asians, specifically people from the Indian subcontinent, have very low HDL levels in the context of insulin resistance. In Turks, however, low HDL levels are associated with elevated hepatic lipase activity without insulin resistance. Apo-AI Mutations

Mutations in the apo-AI gene can decrease HDL formation and result in low plasma HDL-C levels. Apo-AI deficiency can be caused by point mutations in the apo-AI gene or by deletions or gene rearrangements at the apo-AI/apo-CIII/apo-AIV gene locus. Apo-AI deficiency typically results in plasma HDL-C levels less than 0.3 mmol/L (10 mg/dL). Manifestations include a predisposition to premature CHD, xanthomas, and corneal opacities. The molecular diagnosis can be made only by specialized analysis, including electrophoresis of the plasma apolipoproteins and DNA analysis to identify the mutation. Inasmuch as it is difficult to raise the plasma apo-AI or HDL-C levels in these disorders, the treatment should be directed toward lowering the levels of plasma nonHDL-C.

Other rare variants of apo-AI exist, including apo-AI insufficiency, which is caused by a substitution of cysteine for arginine at amino acid 173 and results in lower plasma HDL-C levels. This mutation is inherited as an autosomal dominant trait and has not been associated with premature CHD. Whether the mutation protects against the development of atherosclerosis or whether this kindred has mitigating genetic or environmental factors is not known. Other apo-AI variants are associated with amyloidosis.

### Table 34-9: Genetic Disorders of High-Density Lipoprotein Metabolism

<table>
<thead>
<tr>
<th>Disorder</th>
<th>Mutant Gene</th>
<th>Mode of Inheritance</th>
<th>Population Frequency</th>
<th>Typical Plasma HDL-C (mmol/L [mg/dL])</th>
<th>Corneal Opacifications</th>
<th>Premature Vascular Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Familial hypoalphalipoproteinemia</td>
<td>Unknown</td>
<td>Autosomal dominant</td>
<td>1/400</td>
<td>0.50 (20)</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Familial apo-AI and apo-CIII deficiency</td>
<td>Apo-AI or apo-AI/apo-CIII</td>
<td>Autosomal recessive</td>
<td>Rare</td>
<td>&lt;0.1 (5)</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Apo-AI insufficiency</td>
<td>Apo-AI</td>
<td>Autosomal dominant</td>
<td>Rare</td>
<td>0.3 (10)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>LCAT deficiency</td>
<td>LCAT</td>
<td>Autosomal recessive</td>
<td>Rare</td>
<td>&lt;0.3 (10)</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Fish-eye disease</td>
<td>LCAT</td>
<td>Autosomal recessive</td>
<td>Rare</td>
<td>&lt;0.3 (10)</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Tangier disease</td>
<td>ABCA1</td>
<td>Autosomal recessive</td>
<td>Rare</td>
<td>&lt;0.1 (5)</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>CETP deficiency</td>
<td>CETP</td>
<td>Autosomal recessive</td>
<td>Rare</td>
<td>&gt;2.6 (100)</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

ABCA1, adenosine triphosphate binding cassette transporter A1; apo-AI, apolipoprotein AI; CETP, cholesteryl ester transfer protein; LCAT, lecithin:cholesterol acyltransferase.

*Clinical manifestations also include orange tonsils.*
Cholesteryl Ester Transfer Protein Deficiency

CETP deficiency is a hereditary syndrome in which plasma HDL-C levels are increased because of diminished activity of plasma CETP. Once thought to be rare, the disorder is not uncommon in the Japanese population. Its features include marked elevations of plasma HDL-C in homozygotes (usually >2.6 mmol/L [100 mg/dL]) and possible protection against development of CHD. Heterozygotes have moderately elevated HDL-C levels. The diminished activity of CETP results in diminished transport of cholesteryl esters from HDL to the apo-B-containing lipoproteins. As a result, more cholesteryl esters are found in HDL, and the ratio of total cholesterol to HDL-C is markedly reduced. Presumably, the combination of reduced amounts of atherogenic LDL and increased antiatherogenic HDL decreases susceptibility to atherosclerosis.

Studies in transgenic mice have confirmed this relation between CETP and atherosclerosis. Although mice normally do not have significant plasma CETP activity and have high plasma HDL-C levels, transgenic mice that express CETP have increased plasma LDL-C, decreased HDL-C, and increased susceptibility to atherosclerosis. Nevertheless, as discussed previously, subjects who are heterozygous for CETP deficiency experience CHD despite high plasma HDL, and it remains to be determined whether lowering CETP activity in humans would have therapeutic value. The molecular diagnosis of CETP deficiency requires the measurement of plasma CETP activity in vitro or identification of the DNA mutation. At present, there is no specific treatment.
Lecithin:Cholesterol Acyltransferase Deficiency

LCAT deficiency is a rare autosomal recessive disorder that causes corneal opacities, normochromic anemia, and renal failure in young adults. About 30 kindreds of LCAT deficiency and a number of mutations have been described. LCAT deficiency results in decreased esterification of cholesterol to cholesteryl esters on HDL particles. As a result, free cholesterol accumulates on lipoprotein particles and in peripheral tissues such as the cornea, red blood cell membranes, and renal glomeruli, presumably because reverse cholesterol transport is impaired. Plasma cholesterol levels in LCAT deficiency are variable, HDL-C levels are reduced, and the ratio of free cholesterol to esterified cholesterol in the plasma is increased. Normally, free cholesterol accounts for about one third of the total cholesterol in the plasma; in LCAT deficiency, free cholesterol accounts for most of the plasma cholesterol. The accumulation of free cholesterol in vascular tissues can lead to premature CHD. At present, there is no means to increase the plasma activity of LCAT; therefore, the treatment is preventive (by dietary fat restriction) and symptomatic (e.g., renal transplantation).

A variant of LCAT deficiency is called fish-eye disease. Although this disorder is also caused by mutations of the LCAT gene, the phenotype is less severe than that seen in complete LCAT deficiency. Fish-eye disease is characterized by low plasma HDL-C levels and corneal opacities; anemia, renal disease, and premature atherosclerosis are not present. The phenotypic differences between LCAT deficiency and fish-eye disease have been attributed to whether LCAT activity is absent from both HDL and apo-B-containing lipoproteins (LCAT deficiency) or from HDL only (fish-eye disease); however, one subject with phenotypic fish-eye disease had normal HDL-associated LCAT activity.
Tangier disease is a rare autosomal recessive disorder associated with hypolipidemia, including decreases in both plasma HDL and LDL-C levels, and the presence of orange tonsils. Other features include corneal opacities, hepatosplenomegaly, peripheral neuropathy, and premature CHD. Metabolic studies have demonstrated that the disorder is associated with enhanced catabolism of plasma HDL. Mutations in ABCA1 have been causally linked to Tangier disease. ABCA1 appears to promote cholesterol efflux from cells such as macrophages, the loss of this function apparently accounts for the impaired efflux of cholesterol from Tangier cells. As a result, massive amounts of cholesteryl esters accumulate in macrophages of the reticuloendothelial system. The orange tonsils observed in this disorder are caused by cholesterol deposits. There is currently no specific treatment. The degree to which heterozygosity for ABCA1 mutations contributes to inherited low-HDL syndromes remains to be determined.
PRIMARY GENETIC HYPOLIPIDEMIAS

Familial Hypobetalipoproteinemia

Familial hypobetalipoproteinemia, an autosomal dominant disorder of apo-B metabolism, is associated with plasma cholesterol and LDL-C levels that are less than one half of normal in heterozygotes and with marked hypocholesterolemia (<1.3 mmol/L [50 mg/dL]) in homozygotes.\(^{12,13}\)

Clinical Features

Heterozygous subjects, about 1 in 500 persons, are usually asymptomatic but come to attention because of the detection of low plasma cholesterol levels. Typically, the total plasma cholesterol level is less than the fifth percentile, and it may be less than 2.6 mmol/L (100 mg/dL). Plasma LDL-C levels are also reduced by one half or more, and HDL-C levels are normal or slightly increased.\(^{14}\) Plasma triglyceride levels are reduced in some kindreds. Although heterozygotes are usually asymptomatic, fat malabsorption has been reported.\(^{15}\) Plasma triglyceride levels are reduced in some kindreds. Although heterozygotes are usually asymptomatic, fat malabsorption has been reported.\(^{15}\)

Subjects who are homozygotes or compound heterozygotes for these apo-B mutations are rare, about 1 in 10\(^5\) persons. Homozygotes may be detected at a young age because of fat malabsorption and decreased plasma cholesterol levels. Fat malabsorption is caused by inability to form chylomicrons in the intestine and subsequent failure to absorb fats and fat-soluble vitamins. Fat malabsorption may be accompanied by retinitis pigmentosa, acanthocytosis, and a progressive neurologic degenerative disease resulting from vitamin E deficiency. The acanthocytosis is caused by alterations in red blood cell membrane lipids. Despite the low plasma cholesterol levels, steroidogenesis appears to be normal except when demands are quite high.\(^{16,17}\) Homozygous subjects who produce enough of a truncated isoform of apo-B to facilitate some fat absorption may have a milder phenotype.

Origin and Pathogenesis

Most cases of known origin result from mutations in the apo-B gene.\(^{18}\) More than 30 mutations have been described;

most are either nonsense or frameshift mutations that lead to the formation of truncated apo-B proteins.\(^{19}\) Metabolic turnover studies indicate that these apo-B gene mutations impair the synthesis of apo-B-containing lipoproteins in some cases\(^{20}\) and enhance the clearance of apo-B-containing lipoproteins from the plasma in others.\(^{21}\) The decreased levels of apo-B-containing lipoproteins in the plasma cause low plasma cholesterol and triglyceride levels. Although most known cases of hypobetalipoproteinemia involve apo-B mutations, additional undefined genetic factors can result in low cholesterol levels.\(^{22}\)

In homozygous subjects, the absence of apo-B leads to impaired intestinal chylomicron formation, which in turn leads to impaired absorption of fats and fat-soluble vitamins. Cholesterol absorption is probably also impaired, as demonstrated in transgenic mice lacking intestinal apo-B expression and chylomicron formation.\(^{23}\) As noted, vitamin E malabsorption results in low tissue stores of tocopherol and a degenerative neurologic disease. Retinal degeneration may also be related to deficiencies of fat-soluble vitamins.\(^{24}\)

Diagnosis

The diagnosis of familial hypobetalipoproteinemia is suggested by low plasma total and LDL cholesterol levels inherited as an autosomal dominant trait. The homozygous condition is suggested by extremely low plasma cholesterol and triglyceride levels in an infant or child with fat malabsorption. The differential diagnosis of the homozygous state includes abetalipoproteinemia (see later discussion) and Anderson’s disease (chylomicron retention disease).\(^{25,26}\) The molecular diagnosis of hypobetalipoproteinemia can be performed only in specialized laboratories by gel electrophoresis of plasma apo-B or by DNA analysis to identify specific mutations.

Treatment and Prognosis

Because heterozygous subjects are almost always asymptomatic, no specific treatment is indicated, but dietary supplementation of fat-soluble vitamins (especially vitamin E) is reasonable. Heterozygotes should be informed that if their spouse also has a very low plasma cholesterol level, their children could have homozygous or compound heterozygous hypobetalipoproteinemia; in this scenario, subjects should be referred to a lipid clinic for genetic counseling.

Subjects with homozygous hypobetalipoproteinemia (phenotypic abetalipoproteinemia) should be treated with large doses of vitamin E orally (100 to 300 mg/kg/day), which can raise the tissue vitamin E concentrations and prevent the neurologic complications.\(^{27}\) It is imperative to make the diagnosis and begin treatment at an early age to prevent nutritional deficiencies. Fat should be provided in the diet up to a level that symptoms allow (usually 15% to 20% of calories). Supplementation with medium-chain triglycerides is probably contraindicated because of reports of liver toxicity.\(^{28}\)
Abetalipoproteinemia

Abetalipoproteinemia is a rare autosomal recessive disorder caused by a deficiency in MTP, which results in a virtual absence of apo-B-containing lipoproteins in the plasma. The clinical features include malabsorption of fat and fat-soluble vitamins from the intestine, which can lead to neurologic disease related to vitamin E deficiency. The disorder is frequently detected in infancy because of fat malabsorption associated with marked decreases in plasma cholesterol and triglyceride levels.

Clinical Features

Abetalipoproteinemia occurs in fewer than 1 in 10^6 persons and has the same phenotype as homozygous hypobetalipoproteinemia (described previously), including malabsorption of fat and fat-soluble vitamins from the intestine, which can lead to neurologic disease related to vitamin E deficiency. The disorder is frequently detected in infancy because of fat malabsorption associated with marked decreases in plasma cholesterol and triglyceride levels.

Origin and Pathogenesis

Abetalipoproteinemia is caused by a deficiency of MTP, a protein that transfers triglycerides or phospholipids onto nascent apo-B-containing lipoproteins during their formation in the ER. Insufficient lipidation of nascent particles impairs the synthesis and secretion of these particles by the intestine and the liver, and little if any apo-B is found in the plasma. Several mutations in the MTP gene have been described. The lack of MTP in the intestine leads to impaired chylomicron formation and malabsorption of fats and fat-soluble vitamins.

Diagnosis

The diagnosis is suggested by fat malabsorption associated with extremely low levels of plasma cholesterol (usually <1.3 mmol/L [50 mg/dL]) and triglyceride in an infant or young child. Cholesterol levels in the parents, who are obligate heterozygotes, are normal. The demonstration of the molecular defect requires a specialized laboratory for detection of low or absent MTP in intestinal biopsy specimens or DNA analysis to identify specific mutations. The differential diagnosis of abetalipoproteinemia includes homozygous hypobetalipoproteinemia, in which the obligate heterozygote parents have low plasma lipid levels, and Anderson's disease. Also called chylomicron retention syndrome, the latter disorder is a rare condition that is phenotypically similar to abetalipoproteinemia. Subjects with Anderson's disease cannot secrete chylomicrons from the intestine owing to an as yet undefined recessive mutation.

Treatment

Abetalipoproteinemia is treated in the same way as homozygous hypobetalipoproteinemia. Large doses of vitamin E are given by mouth to prevent the neurologic sequelae of vitamin E deficiency.
Hepatic Lipase Deficiency

Hepatic lipase deficiency is a disorder associated with lack of hepatic lipase activity in the plasma. Its features include combined hyperlipidemia, with elevated levels of plasma cholesterol (6.5 to 38.8 mmol/L [250 to 1500 mg/dL]) and triglyceride (4.5 to 92.6 mmol/L [395 to 8200 mg/dL]); palmar and tuberous xanthomas; and premature arcus corneae. VLDL levels are increased (however, the VLDL cholesterol/triglyceride ratio is <0.3, in contrast to type III hyperlipoproteinemia), and there is a threefold to fivefold enrichment of triglyceride in the LDL and HDL fractions. HDL-C levels are normal or slightly increased. Susceptibility to atherosclerosis is thought to be increased. The demonstration of hepatic lipase deficiency requires specialized in vitro assays of hepatic lipase activity in plasma or DNA analysis to identify mutations. Dietary restriction of fat and cholesterol can lower the plasma lipid levels.
Sitosterolemia

In this rare disorder, dietary sitosterol and other plant sterols, which are not normally absorbed in significant quantities in the intestine, are absorbed in large amounts, resulting in their accumulation in the plasma and in peripheral tissues. Premature atherosclerosis can occur. The molecular defect was mapped to chromosome 2p21 and was identified as mutations in the genes encoding ABCG8 and ABCG5. Clinically, affected children have tendon xanthomas and normal to high plasma levels of LDL-C; the differential diagnosis includes FH and cerebrotendinous xanthomatosis. The diagnosis can be confirmed by gas-liquid chromatography of plasma lipids to demonstrate the abnormal sterols. Treatment consists of restriction of plant sterols in the diet.
Cerebrotendinous Xanthomatosis

Cerebrotendinous xanthomatosis is a rare disorder of sterol metabolism associated with neurologic disease, tendon xanthomas, and cataracts in young adults. Neurologic manifestations include cerebellar ataxia, dementia, spinal cord paresis, and subnormal intelligence. Premature atherosclerosis is common. Osteoporosis has been reported and is presumably caused by alterations in vitamin D metabolism. The disorder results from mutations that cause deficiencies of 27-hydroxylase, a key enzyme in cholesterol oxidation and bile acid synthesis. As a result, high levels of cholesterol and cholestanol, a 5-dihydro derivative of cholesterol, accumulate in the plasma, tendons, and tissues of the nervous system. Treatment is with chenodeoxycholic acid, often in combination with an HMG-CoA reductase inhibitor.
Acid Cholesteryl Ester Hydrolase Deficiency

Acid cholesteryl ester hydrolase deficiency is an autosomal recessive disorder caused by deficiency of a lysosomal esterase, which results in massive accumulation of cholesteryl esters and triglycerides in lysosomes. In the variant called Wolman's disease, which is usually fatal in the first year of life, there is a complete deficiency of the lysosomal esterase. Cholesteryl ester storage disease is a milder variant in which there may be some residual esterase activity; affected subjects can survive past childhood but may develop premature CHD.
Familial Isolated Vitamin E Deficiency

Familial isolated vitamin E deficiency is a rare disorder characterized by low plasma levels of vitamin E in association with progressive neurologic degenerative disease.\cite{490,491} It is caused by lack of hepatic -tocopherol transfer protein, \cite{492} which is thought to facilitate the incorporation of -tocopherol onto nascent VLDLs during their formation in the liver. In the absence of the protein, there is a lack of vitamin E on VLDL, which is a major transport mechanism for delivery of vitamin E to peripheral tissues. Treatment consists of daily supplementation with high doses of oral vitamin E.
SECONDARY DISORDERS OF LIPID METABOLISM

A number of metabolic diseases and drug therapies influence plasma lipids. The secondary disorders of hyperlipidemia

<table>
<thead>
<tr>
<th>TABLE 34-10 – Factors Affecting Plasma High-Density Lipoprotein Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factors That Increase HDL</td>
</tr>
<tr>
<td>Estrogens</td>
</tr>
<tr>
<td>Exercise</td>
</tr>
<tr>
<td>Alcohol</td>
</tr>
<tr>
<td>Drugs: nicotinic acid, fibrates, HMG-CoA reductase inhibitors</td>
</tr>
<tr>
<td>Factors That Decrease HDL</td>
</tr>
<tr>
<td>Androgens</td>
</tr>
<tr>
<td>Progestins</td>
</tr>
<tr>
<td>Cigarette smoking</td>
</tr>
<tr>
<td>Obesity</td>
</tr>
<tr>
<td>Low-fat diet</td>
</tr>
<tr>
<td>Drugs: propranolol, -blockers</td>
</tr>
</tbody>
</table>

HDL, high-density lipoproteins; HMG-CoA, 3-hydroxy-3-methylglutaryl coenzyme A.

are listed in Table 34-8 . Factors that affect HDL levels are listed in Table 34-10 .

Diabetes Mellitus

Of the common diseases, diabetes mellitus exerts some of the most profound effects on plasma lipid metabolism. Hypertriglyceridemia is found in up to one third of all diabetic patients and is related to the critical role of insulin in the production and clearance of triglyceride-rich lipoproteins from the plasma. In addition, diabetic patients frequently have high plasma levels of atherogenic lipoproteins and low plasma HDL, predisposing them to premature CHD, a leading cause of death in diabetes.

In type 1 diabetes, insulin deficiency and poor glycemic control are associated with increases in the plasma levels of triglycerides and of apo-B-containing lipoproteins because of effects on plasma lipid metabolism in peripheral tissues and the liver. In peripheral tissues, insulin deficiency results in impaired LPL activity and diminished clearance of triglyceride-rich particles. Insulin deficiency also causes enhanced lipolysis, which results in increased free fatty acid flux to the liver. In the liver, increased free fatty acid flux drives triglyceride synthesis and VLDL triglyceride synthesis and secretion. Plasma LDL-C levels may also be increased, possibly because insulin stimulates LDL receptor-mediated degradation of LDL and this is diminished in type 1 diabetes.

In its most severe form, insulin deficiency can cause a chylomicronemia syndrome known as diabetic lipemia. In this disorder, massive increases in plasma triglyceride levels (>22.6 mmol/L [2000 mg/dL]) can result in lipemia retinalis, eruptive xanthomas, fatty liver, and pancreatitis. This disorder arises from an acquired LPL deficiency and is relatively rare in the modern era of insulin therapy. The acquired lack of LPL activity results in accumulation of chylomicrons in the plasma (type I pattern), similar to that seen in primary genetic LPL deficiency. The disorder may result from the occurrence of diabetes melitus in combination with some underlying disorder of triglyceride metabolism. The hyperlipidemia related to insulin deficiency and type 1 diabetes is reversible with intensive insulin therapy. Persistent lipid abnormalities in patients with type 1 diabetes with excellent glycemic control suggest that another disorder of lipid metabolism is present.

In type 2 diabetes, which accounts for more than 90% of cases, the metabolic defect is related to insulin resistance and relative insulin deficiency. The insulin resistance appears to be caused by both genetic and acquired factors; metabolic abnormalities that accompany the insulin resistance include obesity, hyperglycemia, hypertension, plasma lipid abnormalities, and hyperuricemia, which are referred to as syndrome X or the metabolic syndrome. One of the most common lipid abnormalities in type 2 diabetes is a moderate hyperlipidemia characterized by an increase in VLDL (type IV pattern), which can be accompanied by various degrees of chylomicronemia (type V pattern), depending on the dietary fat intake. This disorder is characterized by the accumulation of apo-B-containing lipoproteins, which are likely to be proatherogenic, in the plasma. The plasma triglyceride and cholesterol levels are often moderately elevated, the HDL-C concentration may be low, and IDLs, or remnants, which are probably atherogenic, are also often increased. Plasma levels of LDL are increased in some but not all subjects. However, the hypertriglyceridemia in type 2 diabetes is often characterized by an increase in the small, dense LDLs (LDL subclass pattern B), which are particularly atherogenic; this increase occurs even in the absence of hyperglycemia. In addition, a portion of the plasma LDL undergoes glycosylation, which may increase susceptibility to oxidation. Xanthomas are usually absent in this disorder.

Factors that contribute to the lipoprotein abnormalities in type 2 diabetes include decreased LPL activity in muscle and adipose tissue and increased free fatty acid flux to the liver from peripheral adipose tissue stores. Combined with hepatic overproduction of apo-B, which occurs in insulin resistance, the free fatty acid flux drives triglyceride synthesis and VLDL production in the liver. The lipid abnormalities associated with VLDL overproduction may be exacerbated by a primary genetic disorder of lipid metabolism.

The mainstay of therapy for patients with type 2 diabetes with hyperlipidemia is glycemic control through diet, oral hypoglycemic agents, or insulin therapy. Periodic monitoring of the glycosylated hemoglobin is helpful in assessing the glycemic control. However, in contrast to the situation in type 1 diabetes, the hyperlipidemia in type 2 diabetes may be difficult to eradicate even with excellent glycemic control because these subjects have accompanying genetic and acquired metabolic abnormalities that are not resolved by returning blood sugar levels to normal. Decreasing insulin resistance through weight loss can, however, have dramatic effects on both the hyperglycemia and the hyperlipidemia. Metformin, a hypoglycemic agent, may lower plasma glucose levels and produce a modest lowering of plasma lipid levels. In addition to glycemic control, drugs for diabetic hyperlipidemia include gemfibrozil and HMG-CoA reductase inhibitors. Niacin should be used with caution as it can impair or worsen glucose tolerance. Treatment with insulin can lower plasma LDL-C levels in both type 1 and type 2 diabetes mellitus.
Hypothyroidism

Alterations in thyroid function can have profound effects on plasma lipids. and all patients with significant hyperlipidemia should be screened for hypothyroidism. The classical manifestation of hypothyroidism is an elevation of the plasma LDL-C level (6.5 to 15.5 mmol/L [250 to 600 mg/dL]), but this disorder can also be associated with high plasma triglyceride levels. Levels of HDL-C are usually unchanged or slightly lower in hypothyroidism and may be reduced in hyperthyroidism, the latter effect may be related to alterations in hepatic lipase activity. The elevations of plasma LDL-C in hypothyroidism are associated with impaired clearance of LDL, probably reflecting decreased LDL receptor expression. The high LDL-C levels in hypothyroidism are associated with an increased risk for atherosclerosis, but risk for myocardial infarction is not necessarily increased, perhaps because hypothyroidism decreases myocardial oxygen demand. Subclinical hypothyroidism, in which metabolic abnormalities are present without symptoms, can also cause hypercholesterolemia that responds to treatment with thyroid hormone. Hypothyroidism is also associated with low LPL activity, predisposing to increased plasma triglyceride levels. Hyperlipidemia with hypothyroidism may be more marked in those with an underlying genetic susceptibility, but it responds dramatically to thyroid hormone replacement. In elderly patients with CHD or significant risk factors, thyroid hormone should be replaced cautiously because rapid replacement can exacerbate underlying ischemic heart disease.
Estrogen Therapy

Estrogen therapy increases plasma triglyceride levels and can occasionally cause marked hypertriglyceridemia, especially in predisposed individuals. The hypertriglyceridemia appears to be dose-related. Although most women taking either oral contraceptives or postmenopausal estrogens maintain triglyceride levels in the normal range, massive hypertriglyceridemia can on occasion cause pancreatitis. For this reason, triglyceride levels should be measured in women before estrogen therapy is initiated. Estrogens appear to cause hypertriglyceridemia through increased production of VLDL. LPL activity in adipose tissue is not altered.

Estrogen therapy also enhances the clearance of LDL from the circulation and lowers plasma LDL-C. Enhanced LDL clearance probably results from increased hepatic LDL receptor expression. Treatment of postmenopausal women with estrogen can lower LDL-C by 15%. Estrogens can also reduce the hyperlipidemia of type III hyperlipoproteinemia.

Estrogens have significant effects on HDL metabolism, increasing HDL-C in postmenopausal women by more than 15%, primarily by increasing the HDL subfraction. Women have higher HDL-C levels than men at all ages after puberty. This is presumably a result of the effects of androgens because HDL-C levels decrease at puberty in men but remain constant in women. The mechanism by which gonadal hormones alter the HDL-C levels is uncertain but may involve differences in hepatic lipase activity. The effects of estrogen, which raises HDL and lowers LDL-C levels, can be offset if estrogens are combined with progestational agents, which lower HDL and raise LDL cholesterol levels. Because estrogens tend to lower LDL and raise HDL-C, estrogen therapy in postmenopausal women appears to decrease CHD risk significantly. However, more recent studies have brought this into question.
Alcohol Consumption

The regular consumption of large amounts of alcohol can significantly affect plasma triglyceride metabolism. Alcohol metabolism results in increased NADH levels, which inhibit fatty acid oxidation in the liver. This inhibition leads to increased triglyceride synthesis, fatty liver, and enhanced VLDL production. The enhanced VLDL production raises plasma triglycerides and occasionally causes massive hypertriglyceridemia and pancreatitis, especially in persons with an underlying genetic susceptibility. Plasma electrophoresis reveals an accumulation of VLDL in the plasma (type IV pattern) and occasionally a superimposed chylomicronemia (type V pattern) as these particles compete for saturable clearance mechanisms. Subjects with type III hyperlipoproteinemia are particularly sensitive to the effects of alcohol consumption because the alcohol-induced overproduction of VLDL and associated remnant particles occurs in the setting of impaired remnant clearance. Hypertriglyceridemia can be an important contributor to pancreatitis associated with alcohol consumption.

Alcohol consumption is also associated with higher plasma levels of HDL-C, which in turn may explain why moderate alcohol consumption may protect against CHD. Indeed, moderate consumption of alcohol in the form of wine may be inversely correlated with CHD mortality.
Nephrotic Syndrome

Hyperlipidemia, which almost always accompanies the nephrotic syndrome, is caused predominantly by elevation of LDL (type IIa pattern) but can also be caused by high VLDL levels (type IV pattern). Total cholesterol, VLDL, LDL-C, total triglycerides, and plasma apo-B are all elevated. The ratio of total cholesterol to HDL-C is increased, consistent with an atherogenic phenotype. Plasma Lp(a) levels can also be elevated. The pathogenesis of the hyperlipidemia appears to be related to increased rates of production of LDL or VLDL or both. The cause of VLDL overproduction is unclear, but it may be related to a generalized hypersecretion phenomenon in the liver. Because myocardial infarction ranks second only to renal failure as the cause of death in subjects with nephrotic syndrome, the hyperlipidemia should be treated vigorously. HMG-CoA reductase inhibitors appear to be particularly effective.
Protease Inhibitor Use in Human Immunodeficiency Virus Infection

Combination therapy with protease inhibitors for human immunodeficiency virus infection has been associated with metabolic changes, including hyperlipidemia, lipodystrophy, and insulin resistance, in many patients. The cause of this syndrome is currently unknown. With the improved life expectancy in patients receiving protease inhibitor regimens, there is increased concern about the risk of CHD related to the metabolic side effects. Although data concerning CHD risk and treatment guidelines are lacking, many affected subjects are currently treated with regimens similar to those used in patients with insulin resistance or diabetes not associated with human immunodeficiency virus infection.
Other Drugs

In addition to estrogens, other therapeutic agents can cause hyperlipidemia. These agents include glucocorticoids and antihypertensive agents such as thiazide diuretics and α-adrenergic blockers. Exogenous androgens can reduce HDL-C levels.
TREATMENT OF LIPID DISORDERS

Clinical Trials Providing Rationale for Treating Hyperlipidemia

In addition to proving the cholesterol-diet-heart hypothesis, clinical trials in humans demonstrated that lowering plasma lipid levels reduces the incidence of symptomatic CHD in patients at risk. The Lipid Research Clinics Coronary Primary Prevention Trial was the first of the trials and followed approximately 3800 men (ages 35 to 59) with plasma cholesterol concentrations higher than 7.7 mmol/L (290 mg/dL) and with no prior evidence of CHD. The men were divided into two groups. The control group was given a placebo and minimal dietary advice, while the experimental group was given the lipid-lowering drug cholestyramine and dietary management. Over the 7-year study, reductions of 19%, 19%, and about 25% were seen in total and LDL cholesterol concentrations, in the incidence of fatal and nonfatal myocardial infarction, and in the incidence of angina, respectively. At high doses of cholestyramine, the results were even more dramatic, with 35% and 50% reductions in LDL-C levels and in the incidence of symptomatic CHD, respectively.

The Familial Atherosclerosis Treatment Study took the cholesterol-diet-heart hypothesis a step further. In this study, 120 high-risk men were treated aggressively with lipid-lowering protocols for 2.5 years. All subjects had plasma cholesterol and LDL concentrations higher than 7.0 mmol/L (270 mg/dL) and 4.7 mmol/L (180 mg/dL), respectively, and had all angiographically documented CHD. Digitized quantitative coronary arteriography was used for documentation. Men receiving either drug treatment (lovastatin and colestipol or niacin and colestipol) displayed dramatic reductions in plasma LDL concentrations and in the progression of coronary atherosclerosis. More importantly, established atherosclerotic lesions actually regressed.

Despite the impressive results of these intervention trials, critics continued to question recommendations for diet and drug therapy to reduce the risk of CHD on the basis of the fact that the trials, although they clearly demonstrated a marked reduction in CHD-related events, did not conclusively prove that lowering cholesterol prolonged life. Concern was expressed that an increased risk of death from other causes could result from the treatment. Two important prospective studies have addressed these concerns: the Scandinavian Simvastatin Survival Study (4S), conducted in high-risk patients with preexisting coronary disease (a so-called secondary prevention study), and the West of Scotland Coronary Prevention Study (WOSCOPS), conducted in men without preexisting coronary disease (a so-called primary prevention study).

The 4S trial included 4444 patients, of whom 19% were women, with angina pectoris or a previously documented myocardial infarction and with elevated plasma cholesterol concentrations (5.5 to 8.0 mmol/L [214 to 311 mg/dL]). The subjects, who previously received a lipid-lowering diet, were randomly divided into placebo and treatment groups; the drug tested was simvastatin, an HMG-CoA reductase inhibitor. The groups were observed for an average of 5.4 years (the time interval at which 10% of the study participants died). The mean changes in total plasma cholesterol, LDL-C, and HDL-C in the drug group compared with the placebo group were -25%, -35%, and 8%, respectively. The placebo group had 256 (12%) deaths, and the simvastatin group had 182 (8%), for a relative risk of death of 0.70. Deaths from CHD-related events were 189 in the placebo group and 111 in the simvastatin group, for a relative risk of death of 0.58 in the drug group. There were no differences in noncardiovascular deaths between the two groups (49 versus 46, respectively). In addition, the relative risk for one or more coronary events was 0.66 for the drug group, and there was a 37% reduction in the need for myocardial procedures. In this study, cholesterol lowering prolonged life in persons with established CHD.

The results from the WOSCOPS trial were equally impressive. This trial was designed to test the effect of another HMG-CoA reductase inhibitor (pravastatin) in men (6595 subjects, 45 to 65 years old) with elevated plasma cholesterol concentrations (average 7.0 ± 0.6 mmol/L [272 ± 23 mg/dL]) but with no prior history of cardiac disease. The dosage of pravastatin was 40 mg/day, and the subjects were followed for an average of 4.9 years. Drug treatment lowered plasma cholesterol and LDL-C concentrations by 20% and 26%, respectively, whereas the placebo group had no changes in these levels. The placebo group had 248 documented coronary events (either nonfatal myocardial infarction or death from CHD) compared with 174 for the drug group, representing a 31% reduction in risk with treatment. The drug group had a 32% reduction in death from all cardiovascular causes and a 22% reduction in the risk of death from any cause. Furthermore, reduction in clinical cardiac events was evident within 6 to 12 months. Again, cholesterol lowering in men at high risk for CHD reduced death from cardiovascular events without increasing the risk of noncardiovascular death.

The efficacy and safety of treatment with statins were further demonstrated in subsequent trials. In the Cholesterol and Recurrent Events (CARE) trial and the Long-term Intervention with Pravastatin in Ischemic Disease (LIPID) study, subjects with established CHD at baseline were studied for 5 years. In CARE the average baseline plasma cholesterol was 209 mg/dL (139 and 39 mg/dL for LDL-C and HDL-C, respectively), and in LIPID it was 218 mg/dL (150 and 36 mg/dL for LDL-C and HDL-C, respectively). Treatment with pravastatin (40 mg/day) reduced LDL-C by 25% in CARE and by 28% in LIPID. These decreases were associated with a 24% reduction in CHD in LDL and 29% and 28% reductions in nonfatal myocardial infarctions in LIPID and CARE, respectively.

The CARE and LIPID studies along with 4S demonstrate that patients with CHD and LDL-C levels above 130 mg/dL benefit from lipid-lowering therapy. Although patients in these studies with LDL-C levels between 100 and 130 mg/dL did not display a consistent benefit from treatment, the Veterans Affairs High Density Lipoprotein Intervention Trial demonstrated that CHD patients with LDL-C of 104 mg/dL did benefit from gemfibrozil therapy. Thus, the results from these four trials clearly suggest that treatment of CHD patients with LDL-C higher than 100 mg/dL with hypolipidemic drugs is of benefit and that the target of treatment of CHD patients should be to reduce LDL-C to less than 100 mg/dL.

Like the WOSCOPS trial, the Air Force/Texas Coronary Atherosclerosis Prevention Study (AFCAPS/TexCAPS) targeted men and women without CHD but with moderately elevated LDL-C (average, 156 mg/dL) who were also at risk because of age (men older than 45 years; women older than 55 years) coupled with low HDL-C levels (average, 37 mg/dL). In this trial, lovastatin reduced LDL-C by 26% and primary end-point events, including fatal and nonfatal myocardial infarction and unstable angina pectoris, by 37%.

In addition to proving the diet-cholesterol-heart disease hypothesis, the clinical trials confirm and extend the recommended guidelines for treatment of dyslipidemic patients established by the NCEP, which are discussed in a later section.
Approach to the Hyperlipidemic Patient

Individuals come to attention for evaluation for a lipid disorder because of the presence of atherosclerotic vascular disease, pancreatitis, xanthomas, or xanthelasma or because of detection of a high plasma cholesterol or triglyceride level. The initial evaluation of these patients includes a history and physical examination, including assessment of CHD risk factors, and measurement of plasma lipids.

Risk Factors

The initial examination should include an assessment of risk factors for atherosclerotic CHD. Table 34-11 lists the risk factors, as specified by the NCEP in 2001. Diabetes mellitus is considered to be a disorder equivalent to established CHD with regard to risk. In addition, data indicate that obesity is an independent risk factor for CHD. Particular emphasis should be placed on obtaining a detailed history of all first-degree relatives to identify cholesterol disorders or premature CHD.

### TABLE 34-11 — Major Risk Factors for Coronary Heart Disease

<table>
<thead>
<tr>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Age (men 45 yr; women 55 yr)</td>
<td>1. High HDL-C (1.6 mmol/L [60 mg/dL])</td>
</tr>
<tr>
<td>2. Family history of premature CHD (male parent or sibling &lt;55 yr, female parent or sibling &lt;65 yr)</td>
<td>2. Family history of premature CHD (male parent or sibling &lt;55 yr, female parent or sibling &lt;65 yr)</td>
</tr>
<tr>
<td>3. Current cigarette smoking</td>
<td>3. No history of cigarette smoking</td>
</tr>
<tr>
<td>4. Hypertension (blood pressure 140/90 mm Hg or receiving antihypertensive medication)</td>
<td>4. Blood pressure &lt;140/90 mm Hg on no antihypertensive medication</td>
</tr>
<tr>
<td>5. Low HDL-C (&lt;1.0 mmol/L [&lt;40 mg/dL])</td>
<td>5. HDL-C ≥1.0 mmol/L (&gt;40 mg/dL)</td>
</tr>
<tr>
<td>6. Diabetes mellitus.</td>
<td>6. No history of diabetes mellitus</td>
</tr>
</tbody>
</table>


CHD, coronary heart disease; HDL-C, high-density lipoprotein cholesterol.

*Diabetes is regarded as a CHD equivalent, that is, >20% risk of a CHD event within 10 years.

### Physical Examination

A thorough physical examination should be performed with emphasis on the cardiovascular system and the manifestations of hyperlipidemia. Elevated plasma lipids (cholesterol or triglycerides) can accumulate in macrophage reticuloendothelial cells in certain tissues, particularly skin, tendons, eye, liver, and spleen. Deposits in the skin or tendons are manifest as xanthomas or xanthelasmas. In almost all cases, these tissue lipid deposits are reversible with lipid-lowering therapy. Several of the clinical findings are illustrated in Figure 34-26.

Xanthelasmas (see Fig. 34-26A) are small, raised, yellowish macules that typically appear on or near the eyelids above and around the medial canthus. They are seen in FH, familial defective apo-B100, and type III hyperlipoproteinemia. Xanthelasmas occasionally occur in patients with normal plasma cholesterol levels, possibly as the result of enhanced uptake of oxidized or modified lipoproteins by tissue macrophages. Xanthelasmas typically regress with cholesterol lowering and can often be treated effectively in the setting of normal cholesterol levels with low doses of probucol, an antioxidant drug. Lipemia retinalis (see Fig. 34-26B), a condition in which lipemic plasma can be visualized by routine ophthalmologic examination of the fundi, is typically seen only when the triglyceride levels are 22.6 mmol/L (2000 mg/dL) or higher.

Tendon xanthomas (see Fig. 34-26C and D) are nodular deposits of cholesterol that accumulate in tissue macrophages in the Achilles and other tendons, including the extensor tendons in the hands, knees, and elbows. Tendon xanthomas are often present in FH (approximately 75% of subjects), in familial defective apo-B100, and sometimes in type III hyperlipoproteinemia. Small tendon xanthomas can be overlooked if not specifically sought. The examination of the Achilles tendon should include an assessment for thickness and for irregularities of contour (see Fig. 34-26C). Achilles tendon xanthomas can also be detected by xeroradiography.

Tuberous or tuberoeruptive xanthomas (see Fig. 34-26F) are subcutaneous nodules that develop in the skin over areas susceptible to trauma such as the elbows and knees. They may be singular or multiple and may range from pea-sized to lemon-sized. Tuberous xanthomas are most often seen in type III hyperlipoproteinemia and also occur in FH. Palmar xanthomas (see Fig. 34-26F) are cutaneous deposits in the palmar and digital creases of the hands. This type of xanthoma is almost pathognomonic for high plasma levels of VLDL and type III hyperlipoproteinemia.

Eruptive xanthomas (see Fig. 34-26G) are cutaneous xanthomas that appear as small, yellowish, round papules that contain a pale center and an erythematous base. They can be mistaken for acne. The distribution of eruptive xanthomas includes the abdominal wall, the back, the buttocks, and other pressure contact areas. They are caused by accumulation of triglyceride in dermal histiocytes and generally occur when the plasma triglyceride level is 11.3 to 22.6 mmol/L (1000 to 2000 mg/dL) or more. They can disappear rapidly with lowering of the plasma triglyceride level.

### Screening for Secondary Disorders

The history and physical examination should be directed toward uncovering secondary disorders of lipid metabolism (e.g., diabetes mellitus, hypothyroidism, or the nephrotic syndrome) and toward identifying agents that could cause hyperlipidemia (e.g., estrogens, alcohol, or -adrenergic blockers). In addition, laboratory studies should be performed to measure fasting blood sugar or glycosylated hemoglobin or both and to assess renal and hepatic function and urinary protein. To screen for hypothyroidism, the plasma thyroid-stimulating hormone (thyrotropin) level should be assessed because the prevalence of hypothyroidism is increased in dyslipidemic subjects.

### Measurement of Plasma Lipids

Ideally, plasma lipids should be measured at least twice under fasting steady-state conditions before therapeutic decisions are made. Although plasma lipids are usually measured after a 12-hour fast, the plasma lipid response in the postabsorptive state, which may include significant elevations in atherogenic remnant lipoproteins, may be important in atherogenesis. Because cholesterol is a minor component of chylomicrons, plasma cholesterol can be measured in either the fasting or nonfasting state. Of note, plasma lipids can be decreased in the setting of acute myocardial infarction, and follow-up measurements are essential in these cases.
patients.

Most clinical laboratories measure plasma levels of total triglycerides, total cholesterol, and HDL-C; the last analysis is performed after apo-B-containing lipoproteins are precipitated from the plasma with an agent such as heparin. The plasma LDL-C concentration is then calculated from these measurements by the Friedewald formula:\[\text{LDL cholesterol} = \text{total cholesterol} - \text{HDL cholesterol} - \frac{\text{triglycerides}}{5}\]

This formula relies on an estimate of the VLDL cholesterol that is about 20% of the plasma triglyceride level and is reliable for triglyceride levels of 4.5 mmol/L (400 mg/dL) or less. Plasma LDL concentrations calculated by this formula may be inaccurate in the setting of severe hypertriglyceridemia. Specialized lipid laboratories separate the plasma into different density fractions (e.g., VLDL, LDL, and HDL) by sequential ultracentrifugation of the plasma and then measure the lipid concentrations in each fraction. The main advantage of the latter technique is that VLDL cholesterol, which can reflect atherogenic remnant lipoproteins, is measured directly.

Because the plasma lipids can be divided roughly into the proatherogenic apo-B-containing lipoproteins and the antiatherogenic HDL, assessment of the relative proportions of cholesterol in these two fractions can be valuable in the individual lipid profile. One method is to assess absolute levels of HDL and non-HDL cholesterol.\[\text{Another method is to determine the ratio of total cholesterol to HDL-C; it is desirable for the ratio to be about 4.5 or lower (i.e., to have at least 25% of the plasma cholesterol in the HDL fraction). Both methods allow incorporation of the potentially atherogenic apo-B-containing lipoproteins in the assessment of cardiovascular risk from hyperlipidemia by including VLDL cholesterol levels in the assessment.}\]

Several caveats for interpretation of the plasma triglyceride level deserve mention. First, a triglyceride level higher than 11.3 mmol/L (1000 mg/dL) usually signifies the presence of two or more abnormalities of lipid metabolism (e.g., estrogen therapy in the presence of underlying familial hypertriglyceridemia).\[\text{Second, elevated plasma triglyceride levels can fluctuate markedly in a single person over short periods. The fluctuation occurs because the LPL-mediated clearance mechanisms for triglyceride-rich particles become saturated at plasma triglyceride concentrations of approximately 5.6 mmol/L (500 mg/dL), and above this level plasma triglyceride levels largely reflect dietary influences. In this range, the plasma triglyceride levels may rise precipitously with high dietary fat intake and fall rapidly with dietary fat restriction.}\]

In some instances, visual inspection of plasma after it has been refrigerated overnight can be helpful in understanding a disorder of lipoprotein metabolism. To accomplish this, plasma should be collected in a tube containing ethylenediaminetetraacetic acid (EDTA) and refrigerated overnight. A cream-like layer on the top signifies the presence of chylomicrons (type I hyperlipoproteinemia), which are less dense than plasma and float to the surface. A turbid plasma infranatant signifies high levels of VLDL (type IV hyperlipoproteinemia). Plasma can also be analyzed by electrophoresis on paper or agarose gels and stained for neutral lipids to analyze the amount of lipids in the various lipoprotein classes.\[\text{This type of analysis is now rarely used because it has little utility in the classification or treatment of lipid disorders.}\]

Nevertheless, much of the terminology used to describe lipoprotein disorders is derived from the originally described patterns in Table 34-12.

Specialized tests used to assess plasma lipid disorders include measurements of plasma Lp(a) levels and plasma apolipoproteins and screening of genomic DNA for mutations. Plasma Lp(a) levels can be determined by enzyme-linked immunosorbent assays, if warranted. The Lp(a) assay must distinguish the apo(a) protein from plasminogen, which is highly homologous. Plasma Lp(a) measurement may be helpful in assessing CHD risk, and high levels of Lp(a) may suggest the use of niacin as a therapeutic agent.\[\text{Plasma apo-B and apo-AI levels may be of great value in predicting CHD risk; however, because the assays for these apolipoproteins are difficult to perform and add minimal information to that obtained from plasma cholesterol measurements, they are not routinely used. Specialized tests such as in vitro assays for enzyme activities (e.g., LPL, CETP) or DNA screening for mutations (e.g., the familial defective apo-B100 mutation) are currently performed in specialized laboratories.}\]

Selection of Patients for Plasma Lipid Measurements

Guidelines published by the NCEP in 2001 recommend that a complete plasma lipid profile (total cholesterol, LDL-C, HDL-C, and triglycerides) be measured in all adults 20 years of age and older at least once every 5 years.\[\text{The classification of plasma lipid levels is shown in Table 34-12. An analysis of large, long-term cohort studies has provided evidence that young men with hypercholesterolemia are at substantially}

### TABLE 34-12 – Classification of Plasma Lipid Levels

<table>
<thead>
<tr>
<th>Level</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total Cholesterol</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;200 mg/dL</td>
<td>Desirable</td>
</tr>
<tr>
<td>200-239 mg/dL</td>
<td>Borderline high</td>
</tr>
<tr>
<td>240 mg/dL</td>
<td>High</td>
</tr>
<tr>
<td><strong>HDL-C</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;40 mg/dL</td>
<td>Low (consider &lt; 50 mg/dL as low for women)</td>
</tr>
<tr>
<td>&gt;60 mg/dL</td>
<td>High</td>
</tr>
<tr>
<td><strong>LDL-C</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;100 mg/dL</td>
<td>Optimal</td>
</tr>
<tr>
<td>100-129 mg/dL</td>
<td>Near optimal</td>
</tr>
<tr>
<td>130-159 mg/dL</td>
<td>Borderline high</td>
</tr>
<tr>
<td>160-189 mg/dL</td>
<td>High</td>
</tr>
<tr>
<td>190 mg/dL</td>
<td>Very high</td>
</tr>
<tr>
<td><strong>Triglycerides</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;150 mg/dL</td>
<td>Normal</td>
</tr>
<tr>
<td>150-199 mg/dL</td>
<td>Borderline high</td>
</tr>
<tr>
<td>200-499 mg/dL</td>
<td>High</td>
</tr>
<tr>
<td>500 mg/dL</td>
<td>Very high</td>
</tr>
</tbody>
</table>

HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol.


Increased risk for morbidity and mortality from CHD.\[\text{Others have suggested that cost-benefit analyses of the impact of such screening programs on health care and the potential reversibility of atherosclerosis make it questionable whether to screen young, healthy people without significant risk factors for development of premature coronary artery disease; these researchers recommend delaying plasma lipid screening until an older age.}\]

Plasma triglycerides should be measured in all patients with pancreatitis. The lipemic plasma seen in hypertriglyceridemia can interfere with some assays for serum
Selection of Patients for Treatment

The 2001 NCEP treatment guidelines are based on measurements of plasma cholesterol levels and risk factor assessment. The first step in risk assessment is to determine whether CHD or a CHD equivalent (e.g., diabetes mellitus) is present, as these disorders confer more than 20% risk of a CHD event within 10 years. Framingham risk scores (Table 34-13) are used to identify additional subjects with multiple risk factors conferring a 10-year risk of more than 20%. Treatment options, based on risk stratification, are discussed below. Other criteria for assessing risk in primary prevention have been proposed, including an assessment that places more emphasis on age. Guidelines for managing dyslipidemia based on risk assessment were also published by the European Atherosclerosis Society in 1998.

Treatment decisions can be divided into two major categories: treatment of hyperlipidemia in patients with established CHD and treatment of patients for the primary prevention of CHD. Subjects with hyperlipidemia and established CHD should be aggressively treated to lower plasma cholesterol to NCEP treatment guideline levels. The rationale for this recommendation is that lowering of plasma cholesterol in patients with established CHD decreases subsequent risk for cardiac events and decreases mortality from cardiac events. One study demonstrated a reduction in CHD-related and overall mortality in treated patients: this benefit extended to elderly patients as well. In subjects with established CHD, lipid-lowering therapy causes atherosclerotic lesions to stabilize or regress. However, despite the overwhelming evidence demonstrating the benefits of lipid-lowering treatment, many people with hypercholesterolemia and CHD remain untreated.

The use of lipid-lowering therapies in subjects with hypercholesterolemia and no known CHD (i.e., therapy for primary prevention) has been more controversial. This issue arises in part because there is no easy noninvasive test for assessing the degree of atherosclerosis in human coronary arteries. Therefore, treatment recommendations for asymptomatic persons have been based largely on studies of hypercholesterolemic populations. Primary prevention trials showed that lipid lowering in such patients decreased cardiac events and CHD-related mortality but not overall mortality. However, the WOSCOPS trial (discussed previously) showed that lipid-lowering treatment for 5 years decreased both CHD-related mortality and overall mortality in middle-aged hypercholesterolemic men. This study also established that lipid-lowering therapy improves survival in asymptomatic hypercholesterolemic men. Because the costs of such treatment are high, a noninvasive test to evaluate the coronary arteries and identify which hypercholesterolemic patients are at the highest risk would be useful.

The treatment of hypercholesterolemia in persons older than 65 years is controversial, and few prospective studies have addressed the benefits of treatment in this population. Nevertheless, CHD accounts for a high proportion of deaths in this age group, and the 4S trial demonstrated the survival benefits of treatment in elderly individuals who have established CHD.

Patients with type 2 diabetes mellitus have at least a twofold increase in the risk for CHD, and diabetics without diagnosed CHD have the same risk as nondiabetics with established CHD. In post hoc analyses, clinical trials with HMG-CoA reductase inhibitors (e.g., 4S, CARE, APEX/TexCAPS, and LIPID) demonstrated 20% to 55% reductions in CHD events. For these reasons, the American Heart Association and American Diabetes Association have recommended that guidelines for treatment of hyperlipidemia in diabetics be the same as for patients with CHD. Accordingly, the latest NCEP guidelines recommend considering diabetes mellitus as a CHD equivalent, with identical treatment goals. HMG-CoA reductase inhibitors have been recommended as a first-line treatment. Severe hypertriglyceridemia (>11.3 mmol/L [1000 mg/dL]) should be treated aggressively because it is associated with a high risk for pancreatitis, a potentially fatal disease.

Treatment Goals

The treatment goals of the 2001 NCEP guidelines, based on LDL-C and risk assessment, are summarized in Table 34-14. For patients with clinical CHD, the treatment goal for LDL-C should be a level less than 2.6 mmol/L (100 mg/dL). A useful rule is to lower total cholesterol to roughly 4.1 mmol/L (160 mg/dL) and LDL-C to 2.6 mmol/L (100 mg/dL). These values are similar to the average cholesterol levels in the Japanese population, in which CHD is much less prevalent. In general, LDL-C levels should be lowered to less than 4.1 mmol/L (160 mg/dL) for patients with minimal risk (10-year risk <10%) and to less than 3.4 mmol/L (130 mg/dL) in those with moderate risk (10-year risk 10-20%). It is also useful to monitor the ratio of total cholesterol to HDL-C, which should be 4.5 or less.

### Table 34-13 – Framingham Risk Scoring Tables

<table>
<thead>
<tr>
<th>Age (yr)</th>
<th>Points</th>
<th>Age (yr)</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>2034</td>
<td>-9</td>
<td>2034</td>
<td>-7</td>
</tr>
<tr>
<td>3539</td>
<td>-4</td>
<td>3539</td>
<td>-3</td>
</tr>
<tr>
<td>4044</td>
<td>0</td>
<td>4044</td>
<td>0</td>
</tr>
<tr>
<td>4549</td>
<td>3</td>
<td>4549</td>
<td>3</td>
</tr>
<tr>
<td>5054</td>
<td>6</td>
<td>5054</td>
<td>6</td>
</tr>
<tr>
<td>5559</td>
<td>8</td>
<td>5559</td>
<td>8</td>
</tr>
<tr>
<td>6064</td>
<td>10</td>
<td>6064</td>
<td>10</td>
</tr>
<tr>
<td>6569</td>
<td>11</td>
<td>6569</td>
<td>12</td>
</tr>
<tr>
<td>7074</td>
<td>12</td>
<td>7074</td>
<td>14</td>
</tr>
<tr>
<td>7579</td>
<td>13</td>
<td>7579</td>
<td>16</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Total Cholesterol (mg/dL)</th>
<th>Points</th>
<th>Total Cholesterol (mg/dL)</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;160</td>
<td>0</td>
<td>&lt;160</td>
<td>0</td>
</tr>
<tr>
<td>160-199</td>
<td>4</td>
<td>160-199</td>
<td>4</td>
</tr>
<tr>
<td>200-239</td>
<td>7</td>
<td>200-239</td>
<td>8</td>
</tr>
<tr>
<td>240-279</td>
<td>9</td>
<td>240-279</td>
<td>11</td>
</tr>
<tr>
<td>280</td>
<td>11</td>
<td>280</td>
<td>13</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age (yr)</th>
<th>Points</th>
<th>Age (yr)</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>2039</td>
<td>0</td>
<td>4049</td>
<td>0</td>
</tr>
<tr>
<td>5059</td>
<td>1</td>
<td>6069</td>
<td>1</td>
</tr>
<tr>
<td>7079</td>
<td>3</td>
<td>2039</td>
<td>0</td>
</tr>
<tr>
<td>4049</td>
<td>1</td>
<td>5059</td>
<td>0</td>
</tr>
<tr>
<td>6069</td>
<td>3</td>
<td>7079</td>
<td>3</td>
</tr>
<tr>
<td>2039</td>
<td>0</td>
<td>4049</td>
<td>0</td>
</tr>
<tr>
<td>5059</td>
<td>1</td>
<td>6069</td>
<td>1</td>
</tr>
<tr>
<td>7079</td>
<td>3</td>
<td>2039</td>
<td>0</td>
</tr>
<tr>
<td>4049</td>
<td>1</td>
<td>5059</td>
<td>0</td>
</tr>
<tr>
<td>6069</td>
<td>3</td>
<td>7079</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HDL (mg/dL)</th>
<th>Points</th>
<th>HDL (mg/dL)</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>-1</td>
<td>60</td>
<td>-1</td>
</tr>
</tbody>
</table>
TABLE 34-14 -- National Cholesterol Education Program: Treatment Recommendations Based on Risk Assessment

<table>
<thead>
<tr>
<th>Risk Category</th>
<th>LDL-C Goal, mmol/L (mg/dL)</th>
<th>LDL-C Level at Which to Initiate Therapeutic Lifestyle Change, mmol/L (mg/dL)</th>
<th>LDL-C Level at Which to Consider Drug Therapy, mmol/L (mg/dL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHD or CHD risk equivalent</td>
<td>&lt;2.6 (100)</td>
<td>2.6 (100)</td>
<td>3.4 (130)</td>
</tr>
<tr>
<td>(10-yr risk &gt;20%)</td>
<td>2.63.3 (100129): drug optional]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2+ Risk factors</td>
<td>&lt;3.4 (130)</td>
<td>3.4 (130)</td>
<td>10-yr risk 10%20%; 3.4 (130)</td>
</tr>
<tr>
<td>(10-yr risk &lt;20%)</td>
<td>10-yr risk &lt;10%; 4.1 (160)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>01 Risk factor</td>
<td>&lt;4.1 (160)</td>
<td>4.1 (160)</td>
<td>4.9 (160) [4.14.8 (160189): drug optional]</td>
</tr>
</tbody>
</table>


CHD, coronary heart disease; LDL-C, low-density lipoprotein cholesterol.

If severe hypertriglyceridemia is present, the goals for triglyceride lowering are to lower the plasma triglyceride level to less than 4.5 mmol/L (400 mg/dL), which markedly reduces the risk for development of pancreatitis. The 2001 NCEP guidelines reflect the fact that even moderate hypertriglyceridemia (>1.7 mmol/L, or 150 mg/dL) is associated with increased CHD risk. If plasma triglycerides remain above 2.3 mmol/L (200 mg/dL) after the LDL-C goal is reached, further reduction may be achieved by increasing the drug therapy.
Treatment of Hyperlipidemia

The treatment of hyperlipidemia is directed primarily at lowering plasma cholesterol levels to prevent morbidity and mortality from CHD. The rationale and the use of diet or drugs for this purpose have been reviewed. Treatment modalities recommended by the NCEP include therapeutic lifestyle changes, which include diet, weight management, and physical activity recommendations, and drug therapy. A major goal in the treatment of hypertriglyceridemia is to prevent pancreatitis. Patients with CHD or CHD equivalent should immediately start appropriate lipid-lowering therapy. Patients

| TABLE 34-15 – Clinical Identification of the Metabolic Syndrome |
|------------------|------------------|
| **Risk Factor**  | **Defining Level** |
| Abdominal obesity| Waist circumference |
| Men              | >102 cm (>40 in)  |
| Women            | >88 cm (>35 in)   |
| Triglycerides    | 150 mg/dL         |
| HDL-C            |                   |
| Men              | <40 mg/dL         |
| Women            | <50 mg/dL         |
| Blood pressure   | 130/85 mm Hg      |
| Fasting glucose  | 110 mg/dL         |


HDL-C, high-density lipoprotein cholesterol.

*Overweight and obesity are associated with insulin resistance and the metabolic syndrome. However, the presence of abdominal obesity is more highly correlated with the metabolic risk factors than is an elevated body mass index. Therefore, the simple measurement of waist circumference is recommended to identify the body weight component of the metabolic syndrome.

Some male patients can develop multiple metabolic risk factors when the waist circumference is only marginally increased, for example, 94-102 cm (37-40 in). Such patients may have a strong genetic contribution to insulin resistance and, like men with categorical increases in waist circumference, they should benefit from changes in life habits.

Dietary Treatment

All patients should receive instruction about restriction of dietary saturated fat and cholesterol. On average, the Western diet contains 35% to 40% of calories as fat (about 15% to 20% saturated fat, 10% polyunsaturated fat, and 10% monounsaturated fat) and approximately 380 mg of cholesterol per day. In contrast, the average diet in many underdeveloped countries is closer to 10% of calories as fat. The American Heart Association has recommended a single diet, termed Step 1 Dietary Therapy by the NCEP, for all persons older than 2 years. The diet consists of 50% of calories in the form of carbohydrate (complex carbohydrate preferred), 20% as protein, and no more than 30% as fat. Saturated fat should constitute less than 10% of total calories, mono- and polyunsaturated fat up to 10% of total calories. Cholesterol intake should be limited to 250 mg/day or less. A more stringent restriction of saturated fat (<7% of calories) and cholesterol (200 mg/day) intake (Step 2 diet) is recommended for patients with established coronary artery disease. This diet is similar to the therapeutic lifestyle changes diet proposed by the NCEP. In patients with established CHD, drug therapy should be instituted concomitantly with dietary therapy. Changing from a typical Western diet to a Step 1 diet lowers plasma cholesterol levels by only 5% to 10%. The adoption of a Step 2 diet (25% of calories as fat) has produced mixed results in two noteworthy studies. One study, restricting fat intake to 25% of calories lowered cholesterol levels by only 5%; however, in this outpatient study, it was not clear that the diets were strictly adhered to, as evidenced by comparison of the expected and the actual weight loss in participants. In an inpatient (metabolic ward) study, reducing fat intake from 43% to 25% of calories lowered total cholesterol levels by 17% and LDL-C levels by 23%. Even more stringent limitation of fat intake (e.g., 10% of calories) can further lower plasma lipids. In conjunction with modifications of other lifestyle factors, this diet achieved a reduction of about 25% in total plasma cholesterol levels, an average weight loss of 10 kg, and angiographic regression of coronary artery disease.

TABLE 34-16 – National Cholesterol Education Program: Therapeutic Lifestyle Changes for Reducing Coronary Heart Disease Risk

1. Diet:
   - Saturated fat (and trans-esterified fatty acids) less than 7% of total calories
   - Polyunsaturated fat up to 10% of total calories
   - Monounsaturated fat up to 20% of total calories
   - Total fat 25%-35% of total calories
   - Carbohydrates (predominantly complex) 50%-60% of total calories
   - Fiber 20 g/d
   - Protein 15% of total calories
   - Cholesterol <5.2 mmol/L (200 mg/dL)
   - Consider plant stanols/sterols (2 g/d) to enhance LDL-C lowering

2. Weight reduction
3. Increased physical activity

LDL-C, low-density lipoprotein cholesterol.

latter studies suggest that dietary restriction of fat intake can lower plasma cholesterol levels provided the diet is followed. The typical saturated fat and cholesterol contents of common foods are indicated in Table 34.17.

The effects of various types of fat in the diet have been studied extensively. Current recommendations are that dietary fat intake be lowered, primarily by restricting saturated fat intake because saturated fats appear to have the greatest propensity to elevate plasma cholesterol levels. The mechanism of this effect appears to involve decreased receptor-mediated clearance of LDL from the plasma. Similarly, high levels of cholesterol intake raise plasma cholesterol by reducing receptor-mediated catabolism of LDL and by increasing LDL synthesis.

Polyunsaturated fats, such as linoleic acid, which are chiefly found in vegetable oils, have less deleterious effects on plasma cholesterol levels than saturated fats. However, because the long-term effects of consumption of large amounts of polyunsaturated fats are unknown, current recommendations are that polyunsaturated fat constitute no more than 10% of total calories. Polyunsaturated fats that have been hardened by hydrogenation (e.g., in margarines) result in the conversion of some double bonds in the fatty acid from the cis to the trans configuration; these trans fatty acids appear to exert effects on plasma cholesterol that are comparable to those of saturated fatty acids.

Fish oils are rich in a particular subset of long-chain polyunsaturated fats containing double bonds at the n-3 or omega (3) position, such as eicosapentaenoic acid (C-20) or docosahexaenoic acid (C-22). Because the incidence of CHD is decreased in populations that consume relatively large amounts of fish oils, these polyunsaturated fatty acids have been evaluated for the treatment of hyperlipidemia and the prevention of CHD. Although they lower VLDL levels and are effective for treating hypertriglyceridemia, fish oils appear to have little effect on total cholesterol levels and increase LDL levels in many patients with hypertriglyceridemia. Fish oils may have beneficial effects on cardiovascular disease by means other than their effects on lipoproteins, such as inhibition of platelet aggregation, but a prospective study of male health professionals found no protective effects of increased intake of dietary fish oil on the development of CHD. In addition, fish oil supplementation can worsen glycemic control in persons with type 2 diabetes.

Of all the types of fatty acids, monounsaturated fats may have the least deleterious effects on plasma lipoprotein metabolism. Monounsaturated fats, such as oleic acid, are found in high quantities in olive oil and canola oil. When substituted for saturated fats, monounsaturated fats lowered total plasma cholesterol levels without lowering plasma HDL-C levels.

Other dietary components may influence plasma lipid levels. For example, soluble fibers such as pectin or oat bran, which may bind bile acids in the gut and promote net cholesterol excretion, can result in modest (~10%) decreases in LDL-C levels. Margarine containing sitostanol, a nonabsorbed plant stanol that inhibits cholesterol absorption, reduces serum cholesterol by about 10%. Garlic and walnuts are reported to result in modest decreases in plasma cholesterol levels.

### Drug Treatment

Categories of drugs for treatment of lipid abnormalities are listed in Table 34.18. These include drugs that interfere with bile acid absorption from the gut, such as bile acid sequestrants, or with cholesterol biosynthesis in cells, such as HMG-CoA reductase inhibitors. These agents reduce cholesterol levels and increase LDL receptor expression in cells, thereby lowering plasma LDL concentrations. Other agents, including niacin and gemfibrozil, either inhibit VLDL synthesis and secretion or enhance the clearance of triglyceride-rich particles by enhancing VLDL-mediated catabolism. Finally, drugs that work primarily as antioxidants are being evaluated for treating hypertriglyceridemia, fish oils appear to have little effect on total cholesterol levels and increase LDL levels in many patients with hypertriglyceridemia. Fish oils may have beneficial effects on cardiovascular disease by means other than their effects on lipoproteins, such as inhibition of platelet aggregation, but a prospective study of male health professionals found no protective effects of increased intake of dietary fish oil on the development of CHD. In addition, fish oil supplementation can worsen glycemic control in persons with type 2 diabetes.

Of all the types of fatty acids, monounsaturated fats may have the least deleterious effects on plasma lipoprotein metabolism. Monounsaturated fats, such as oleic acid, are found in high quantities in olive oil and canola oil. When substituted for saturated fats, monounsaturated fats lowered total plasma cholesterol levels without lowering plasma HDL-C levels.

### HMG-CoA Reductase Inhibitors

Several potent inhibitors of HMG-CoA reductase, the enzyme that catalyzes the rate-limiting step in cholesterol biosynthesis, are available. Inhibition of cholesterol biosynthesis up-regulates cellular LDL receptors and enhances clearance of LDL from the plasma into cells. The inhibitors differ in their side chains, which can affect their relative hydrophobicity. For example, lovastatin and simvastatin are relatively hydrophobic and lipophilic compared with others.

### Table 34.17 -- Cholesterol and Saturated Fat Content in Some Common Foods

<table>
<thead>
<tr>
<th>Food</th>
<th>Cholesterol (mg/100 g)</th>
<th>Saturated Fat (g/100 g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eggs</td>
<td>500</td>
<td>3</td>
</tr>
<tr>
<td>Organ meats (liver, kidney)</td>
<td>&gt;300</td>
<td>2</td>
</tr>
<tr>
<td>Butter</td>
<td>230</td>
<td>50</td>
</tr>
<tr>
<td>Shrimp, crab, lobster</td>
<td>110</td>
<td>1</td>
</tr>
<tr>
<td>Cheese</td>
<td>110</td>
<td>21</td>
</tr>
<tr>
<td>Meat (beef, pork, lamb)</td>
<td>90100</td>
<td>513</td>
</tr>
<tr>
<td>Poultry (no skin)</td>
<td>90</td>
<td>1</td>
</tr>
<tr>
<td>Fish</td>
<td>70</td>
<td>1</td>
</tr>
<tr>
<td>Ice cream (10% fat)</td>
<td>40</td>
<td>7</td>
</tr>
<tr>
<td>Sherbet; frozen yogurt</td>
<td>4</td>
<td>&lt;1</td>
</tr>
<tr>
<td>Milk, whole (3.5%)</td>
<td>14</td>
<td>2</td>
</tr>
<tr>
<td>Milk, skim</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Cottage cheese</td>
<td>6</td>
<td>&lt;1</td>
</tr>
<tr>
<td>Margarine, soft</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>Vegetable oil</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>Coconut oil, cocoa butter</td>
<td>0</td>
<td>75</td>
</tr>
</tbody>
</table>


### Table 34.18 -- Drugs Commonly Used for Treating Hyperlipidemia

<table>
<thead>
<tr>
<th>Class</th>
<th>Drugs Available</th>
<th>Dosage</th>
<th>Major Lipoprotein Decreased</th>
<th>Mechanism</th>
<th>Side Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bile acid sequestrants</td>
<td>Cholestyramine</td>
<td>412 g bid</td>
<td>LDL</td>
<td>Increase sterol excretion; increase LDL receptor-mediated removal of LDL</td>
<td>Gastrointestinal symptoms; can increase triglycerides; binds other drugs</td>
</tr>
<tr>
<td></td>
<td>Colestipol</td>
<td>515 g bid</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Niacin, which is more hydrophilic, these properties appear to have little effect on the ability to lower LDL levels. Therapeutic doses of these agents reduce total cholesterol and LDL-C levels by 20% to 55%. Plasma triglyceride levels greater than 250 mg/dL are reduced by amounts that are comparable to the reductions in LDL-C. High doses of the most potent statins reduce triglyceride levels by 35% to 40%. In patients with triglyceride levels less than 250 mg/dL, statins reduce triglyceride levels by less than 25%. Statins increase plasma HDL-C levels by 5% to 10%. In addition to lipid-lowering effects, a multitude of potentially cardioprotective effects are being ascribed to statins, including improved endothelial function, increased plaque stability, decreased inflammation, decreased lipidoprotein oxidation, and improved circulation. The reductase inhibitors are well tolerated and cause few side effects. The most serious potential side effect is myopathy, which occurs in less than 1% of persons taking lovastatin and can cause myoglobinuria and renal failure. Patients taking HMG-CoA reductase inhibitors in whom myalgias develop should have serum creatine phosphokinase measurements, and the drug should be stopped immediately if evidence of myositis is found. The risk of myopathy is increased if an HMG-CoA reductase inhibitor is used in combination with niacin or gemfibrozil which can cause myopathy by themselves, or with drugs that are metabolized by the 3A4 isozyme of cytochrome P450 (CYP3A4), such as erythromycin, cyclosporine, nefazodone, or protease inhibitors.

Bile acid sequestrants are anion-exchange resins that exchange chloride for negatively charged molecules in the intestine, these agents can interfere with the absorption of other medications, including levothyroxine, digoxin, warfarin, and thiazide diuretics. Therefore, resins are given at least 4 hours before or 1 hour after other medications.

The most inexpensive drug for treating hyperlipidemia is the B vitamin niacin. Therapeutic doses of niacin (typically 2.0 to 4.5 g/day) lower both total and LDL cholesterol by 15% to 30%, lower triglyceride levels by 30% to 40%, and raise HDL-C levels by 15% to 25%. Maximal HDL increases usually occur with therapeutic doses of 1.5 to 2.0 g/day. Niacin

<table>
<thead>
<tr>
<th>Drug</th>
<th>20%25%</th>
<th>26%30%</th>
<th>31%35%</th>
<th>36%40%</th>
<th>41%50%</th>
<th>51%55%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atorvastatin</td>
<td>10</td>
<td>20</td>
<td>40</td>
<td>60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fluvastatin</td>
<td>20</td>
<td>40</td>
<td>80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lovastatin</td>
<td>10</td>
<td>20</td>
<td>40</td>
<td>80</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### TABLE 34-19 – Drug Selection Based on Major Lipid Abnormality

<table>
<thead>
<tr>
<th>Major Elevated Plasma Lipid(s)</th>
<th>Drugs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cholesterol</td>
<td>HMG-CoA reductase inhibitor</td>
</tr>
<tr>
<td>Cholesterol and triglyceride</td>
<td>Niacin</td>
</tr>
<tr>
<td>Triglyceride</td>
<td>Niacin</td>
</tr>
<tr>
<td>Lp(a)</td>
<td>Niacin</td>
</tr>
</tbody>
</table>

HMG-CoA, 3-hydroxy-3-methylglutaryl coenzyme A.

Statins, only pravastatin and fluvastatin are not extensively metabolized by CYP3A4. Although these drugs may be less likely to cause myopathy when used with other CYP3A4-metabolized drugs, myopathy has been reported with their use.

To minimize the risk of myopathy, drug combinations should be avoided. However, statins may be used with a predisposing drug without increasing myopathy risk if the statin is administered at no more than 25% of its maximal dose. The plasma creatine kinase level should probably be monitored if the combinations are used. Serum transaminase elevations (greater than three times normal) occur in 2% to 3% of patients. The long-term side effects of HMG-CoA reductase inhibitor therapy are not known, but no significant long-term toxicities have been observed with lovastatin, which has now been in use for more than 15 years.

**Bile Acid Sequestrants**

Bile acid sequestrants are anion-exchange resins that exchange chloride for negatively charged bile acids. The bound bile acids are then excreted in the feces. The increased excretion of bile acids causes increased oxidation of cholesterol to form bile acids in hepatocytes, and the resultant up-regulation of hepatic LDL receptors in turn lowers plasma LDL concentrations. Because bile acid sequestrants act in the intestine, the side effects are limited to local effects in the gastrointestinal system (e.g., bloating, gas, constipation). At therapeutic doses, these agents can lower plasma cholesterol levels by 15% to 25%. However, they can increase plasma triglyceride levels and must be used with caution in patients predisposed to hypertriglyceridemia. In addition, because they bind negatively charged molecules in the intestine, these agents can interfere with the absorption of other medications, including levothyroxine, digoxin, warfarin, and thiazide diuretics. Therefore, resins are given at least 4 hours before or 1 hour after other medications.

**Niacin**

The most inexpensive drug for treating hyperlipidemia is the B vitamin niacin. Therapeutic doses of niacin (typically 2.0 to 4.5 g/day) lower both total and LDL cholesterol by 15% to 30%, lower triglyceride levels by 30% to 40%, and raise HDL-C levels by 15% to 25%. Maximal HDL increases usually occur with therapeutic doses of 1.5 to 2.0 g/day. Niacin

---

### TABLE 34-20 – Doses (mg) of Statins Required to Achieve Various Reductions in Low-Density Lipoprotein Cholesterol from Baseline

<table>
<thead>
<tr>
<th>Drug</th>
<th>20%25%</th>
<th>26%30%</th>
<th>31%35%</th>
<th>36%40%</th>
<th>41%50%</th>
<th>51%55%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atorvastatin</td>
<td>10</td>
<td>20</td>
<td>40</td>
<td>60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fluvastatin</td>
<td>20</td>
<td>40</td>
<td>80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lovastatin</td>
<td>10</td>
<td>20</td>
<td>40</td>
<td>80</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
also lowers plasma Lp(a) concentrations by up to 40%. The preparation must be niacin and not niacinamide, which has no efficacy. The mechanism whereby niacin affects plasma lipids is unclear but seems to be associated with decreased hepatic VLDL production.

The most troublesome side effect of niacin therapy is a flushing syndrome that occurs shortly after taking the medicine. Flushing can be minimized by initiating therapy with small doses (e.g., 100 mg) and gradually increasing the dosage to the therapeutic range over weeks to months. Repeated dosing is associated with a gradual tolerance to the flushing syndrome. In addition, taking an aspirin about 1 hour before the niacin can diminish the flushing, possibly by inhibiting prostaglandin-mediated side effects.

The most serious complication of niacin therapy is hepatotoxicity, and therapy should be accompanied by monitoring of serum liver function tests. Mild increases in serum transaminases are common when doses are increased rapidly; however, therapy should be discontinued if transaminases reach highly elevated levels (e.g., <10 times normal). Because hepatotoxicity appears to be more common with sustained-release preparations of niacin, the immediate-release crystalline form is preferred. Other side effects of niacin therapy include impairment or worsening of glucose tolerance and hyperuricemia. Data suggest that niacin can be used safely in patients with glucose intolerance or diabetes mellitus, but the drug should be used with great caution in patients with a history of gout and is contraindicated in patients with active peptic ulcer disease.

Fibric Acid Derivatives

The fibric acid derivatives, clofibrate and gemfibrozil, lower plasma triglycerides by about 40% and increase HDL-C levels by about 10% but have only minor effects on LDL-C. These agents act by activating the peroxisome proliferator-activated receptor, a nuclear hormone receptor that is expressed in the liver and other tissues. This results in increased fatty acid oxidation, increased LPL synthesis, and reduced expression of apo-CIII, all of which contribute to lowering plasma triglycerides. The physiologic results are a decrease in VLDL triglyceride production and an increase in LPL-mediated catabolism of triglyceride-rich lipoproteins. This receptor also stimulates the expression of apo-AI and apo-AII, leading to increased HDL levels. These agents are given twice a day and are well tolerated. Side effects include gastrointestinal discomfort and possibly an increased incidence of cholesterol gallstones. Fibric acid derivatives should be used with great caution in the setting of renal insufficiency because patients with this condition have an increased risk of myopathy.

Clofibrate received adverse publicity because of a large clinical trial in which slightly more cancer deaths were noted in the clofibrate-treated group. However, a later analysis did not substantiate this finding, and there is no firm evidence that the drug is carcinogenic in humans. In two subsequent trials, the Helsinki Heart Study (primary prevention) and the Veterans Affairs High Density Lipoprotein Intervention Trial (secondary prevention), gemfibrozil treatment reduced fatal and nonfatal CHD events without changes in mortality rates.

Combination Therapy

For patients with severe elevations of plasma cholesterol (e.g., >7.8 mmol/L [300 mg/dL]), in whom treatment goals are to reduce the plasma and LDL cholesterol levels by 50% or more, combination drug therapy is usually required. Often this can be achieved with combinations that employ lower doses than needed when these agents are used as single agents. For example, combined therapy with an HMG-CoA reductase inhibitor and either niacin or a bile acid sequestrant can lower plasma cholesterol levels by more than 50%. Occasionally, in patients with severe hypercholesterolemia, combinations that include an HMG-CoA reductase inhibitor, niacin, and a bile acid sequestrant can achieve the desired lipid-lowering effect. Patients with diabetes mellitus who have high plasma levels of triglyceride and VLDL cholesterol may benefit from combination therapy with a low dose of an HMG-CoA reductase inhibitor and gemfibrozil. Because the use of HMG-CoA reductase inhibitors with niacin or gemfibrozil is associated with a higher risk of myopathy, such combinations must be used with caution. Monitoring of the serum creatine kinase level may be helpful, and only low doses of the reductase inhibitors (20 mg/day of lovastatin or the equivalent) should be used in most patients.

Drugs on the Horizon

A number of strategies to improve the treatment of dyslipidemia are under development. For example, more potent statins capable of lowering LDL-C levels by more than 65% are being developed. Inhibitors of MTP represent another strategy. These agents have the potential to lower both plasma triglycerides and cholesterol levels by inhibiting hepatic VLDL production. Agents that impair cholesterol absorption, such as ezetimibe or an ACAT2 inhibitor, may offer alternative strategies for lowering plasma cholesterol levels. Inhibitors of CETP have the potential to raise HDL levels and inhibit atherosclerosis.

There has been intense interest in the use of antioxidants to treat or prevent atherosclerosis. Probucol, a potent antioxidant transported on lipoproteins, protected against the development of atherosclerotic lesions in animals. However, in a trial in humans, probucol failed to benefit patients with established peripheral vascular disease. A number of new antioxidants are under development. The antioxidant vitamins (the fat-soluble vitamins E and A and the water-soluble vitamin C) may also be antiatherogenic. Evidence to support this comes from population studies in which high intake of antioxidants was associated with decreased coronary events. However, a prospective trial examining carotene supplementation failed to show a benefit. In a prospective trial of vitamin E supplementation, 400 IU/day reduced the risk of nonfatal infarction in patients with established CHD. More prospective data demonstrating that supplemental antioxidant vitamin therapy can prevent atherosclerosis or CHD events in humans are needed before a recommendation can be made.

Estrogen Replacement Therapy and Coronary Heart Disease

Because estrogen replacement therapy in postmenopausal women reduces LDL-C and Lp(a) levels and raises HDL-C levels and because CHD risk increases substantially in postmenopausal women, there has been intense interest in using estrogens to prevent CHD. However, two clinical trials have shown no benefit of estrogen replacement therapy in women with established CHD. It remains to be determined whether estrogen replacement will be effective in the primary prevention of CHD. Estrogen replacement has been shown to have multiple noncardiovascular benefits.

Surgical Treatment

Partial ileal bypass surgery has been used to reduce lipid levels in patients with severe hypercholesterolemia who cannot tolerate lipid-lowering drugs. This surgical therapy can reduce total cholesterol levels by 20% to 25% and cause regression of angiographically measured atherosclerotic lesions. In addition, liver transplantation and portacaval shunting have been used as experimental therapies for homozygous FH (see previous discussion).
Treatment to Raise High-Density Lipoprotein Cholesterol

Patients with familial hypoproteinemia may have normal or modestly increased plasma cholesterol levels but have very low HDL-C levels, resulting in a predisposition to CHD. Such patients may have high ratios of total cholesterol to HDL-C (e.g., >10) despite having a normal plasma cholesterol level. At present, there are no highly effective therapies to raise HDL-C levels. Estrogens can be given to postmenopausal women and exercise in sustained amounts may modestly increase HDL-C levels. HDL-C levels increased by approximately 2 mg/dL for every 10 miles run per week in one study of recreational runners. Alcohol, when consumed in modest quantities, can also increase HDL-C levels. Of the available drug therapies, niacin results in the largest increase in HDL-C levels (about 15% to 25%); gemfibrozil and HMG-CoA reductase inhibitors increase HDL-C by about 10%. However, the potent ability of HMG-CoA reductase inhibitors to lower total cholesterol can improve the ratio of total cholesterol to HDL-C.
Treatment of the Chylomicronemia Syndrome

Patients with the chylomicronemia syndrome often present with acute pancreatitis and severe hypertriglyceridemia (triglycerides >22.6 mmol/L [2000 mg/dL]). These patients should be treated with total fat restriction until the triglyceride level falls to a safe range (e.g., <11.3 mmol/L [1000 mg/dL]), at which time a fat-restricted diet (e.g., <10% of calories) can be instituted and the plasma triglyceride level further monitored. The goal is to maintain the triglyceride level at less than 11.3 mmol/L (1000 mg/dL) and preferably less than 4.5 mmol/L (400 mg/dL). Often this can be accomplished by modifying the diet and eliminating or modifying secondary causes of hyperlipidemia such as drugs, glucose intolerance, or alcohol consumption. However, such patients often require a triglyceride-lowering drug, such as gemfibrozil or niacin, to maintain the plasma triglyceride level in a range that should prevent subsequent episodes of pancreatitis. Recommendations for the management of the hypertriglyceridemia associated with pregnancy have been described.
References


125. Edwards PA, Tabor D, Kast HR, Venkateswaran A. Regulation of gene expression by SREBP and SCAP. Biochim Biophys Acta 2000; 1529:103113.
128. Strickland DK, Ashcom JD, Williams S, et al. Sequence identity between the \( \gamma \) -macroglobulin receptor and low density lipoprotein receptor-related protein suggests that this molecule is a multifunctional receptor. J Biol Chem 1990; 265:1740117404.
References


1696


246. Fredrickson DS. If's time to be practical. Circulation 1975; 51:209211.


11. Myant NB. Familial defective apolipoprotein B-100: a review, including some comparisons with familial hypercholesterolemia. Atherosclerosis 1993; 104:118.


24. Carew TE, Schwenke DC, Steinberg D. Antithrombotic effect of probucol unrelated to its hypercholesterolemic effect: evidence that antioxidants in vivo can selectively inhibit low density lipoprotein degradation in macrophage-rich fatty streaks and slow the progression of atherosclerosis in the Watanabe heritable hyperlipidemic rabbit. Proc Natl Acad Sci USA 1987; 84:7723729.


References


389. Riedler PM, Hennekens CH, Stampfer MJ. A prospective study of lipoprotein(a) and the risk of myocardial infarction. JAMA 1993; 270:21952199.


400. Galvish D, Azrani N, Breslow JL. Plasma Lp(a) concentration is inversely correlated with the ratio of Kringle IV/Kringle V encoding domains in the apo(a) gene. J Clin Invest 1989; 84:20212027.


1700
References


1702
557. Webb OL, Laskarzewski PM, Glueck CJ. Severe depression of high-density lipoprotein levels in weight lifters and body builders by self-administered exogenous testosterone and anabolic-androgenic steroids. Metabolism 1984; 33:971975.
References

References


Cobb MM, Teitelbaum HS, Breslow JL. Lovastatin efficacy in reducing low-density lipoprotein cholesterol levels on high- vs low-fat diets. JAMA 1991; 265:9971001.


A great deal is now known about the mechanisms underlying human tumorigenesis in general and about neoplasia of the endocrine glands in particular. The application of general principles of neoplasia to endocrine tumors has been a productive area of research and is now being translated into clinical applications. Endocrine tumorigenesis also involves some special, if not unique, features that must be considered in understanding the pathogenesis of endocrine tumors. The aim of this chapter is to present the general principles of neoplasia as a framework through which current knowledge about and future advances in endocrine tumorigenesis can be understood.
MOLECULAR TUMOR BIOLOGY

Both inherited tumor predisposition syndromes and the more common noninherited (sporadic) forms of neoplasia are genetic diseases in the sense that tumors develop when specific damage to a gene leads to deregulated cell growth. As a general rule, damage to one such gene controlling cell growth is not sufficient to confer a neoplastic phenotype on a cell; instead, multiple mutations accumulate over time. Inherited tumor syndromes constitute a special case in which mutation of one key gene is already present in each somatic cell at birth.

Clonality

Concepts of Clonality and Clonal Evolution

All cancers and many benign hypercellular expansions are monoclonal; that is, they are composed of cells that are the descendants of a single clonal progenitor cell in which the accumulation of a sufficient number of deoxyribonucleic acid (DNA) alterations (and, perhaps, other epigenetic damage that does not actually change the DNA) caused a selective advantage. Over time, this selective advantage, manifested as an increase in proliferative capacity, a decrease in the normal cell death rate, or both, leads to the development of a neoplasm.

The monoclonality of tumors implies that the necessary accumulation of mutations occurs only rarely in the large population of cells in a tissue. Viewed in another way, the identification of a specific monoclonal (also called clonal) change in DNA, found in all or most of the cells of a neoplasm but not in that individual’s constitutional DNA (e.g., obtained from adjacent normal tissue or leukocytes), indicates that this DNA alteration had been advantageous in the accelerated evolutionary process that is oncogenesis. Such a DNA lesion, especially if it recurs in other tumors of the same type, is therefore likely to contribute to tumor development. In fact, one of the strongest types of evidence to implicate a given gene as a driving force in tumorigenesis is the demonstration of clonal DNA changes within or near that gene.

This situation contrasts with the one in which an increase or a decrease in the expression (ribonucleic acid [RNA] or protein levels) of a particular gene occurs in a tumor; such changes can be secondary consequences of tumor-related processes and may or may not themselves drive or contribute to the neoplastic phenotype. This caveat is especially important to bear in mind, given that the Human Genome Project and modern microarray technology have made it possible to examine simultaneously expression changes in a huge number of genes in a tissue or tumor.

An original clonal progenitor or transformed cell does not necessarily contain all the genetic lesions that are ultimately present in the mature, clinically apparent tumor. A continuing process of clonal evolution may result in the development of additional DNA damage that provides an incremental selective advantage to the single tumor cell in which it occurs. Over time, the progeny of this cell may become the dominant clonal population. The percentage of neoplastic cells in the final tumor that contain such a later mutation may vary markedly and depends on factors such as the duration of the mutation’s existence and the relative rates of proliferation and death of the various cell populations.

Endocrine tumors are often sufficiently differentiated to express the hormonal activity characteristic of the corresponding normal cell type, but the hormonal function of the tumor is typically regulated in an abnormal fashion. It is important to understand that the genes that cause tumors of endocrine tissues do so only because of their effects on cell proliferation and accumulation. Such genes need not influence hormonal function. Furthermore, a mutant gene that alters hormonal function but confers no selective growth advantage is not tumorigenic. Nevertheless, the frequent coexistence of growth deregulation and hormonal hyperfunction in endocrine tumors does indicate that the tumor-causing genes may directly or indirectly alter hormone control pathways.

In certain instances, such as a mutation affecting the subunit of stimulatory G protein (Gαs) in growth hormoneproducing thyroid adenomas, a single mutant gene can directly contribute to both cell proliferation and hormonal hyperfunction. In general, however, the relationship between hypercellularity caused by clonally selected mutant genes and hormonal hyperfunction is poorly understood.

Hyperplasia versus Neoplasia

Not all hypercellular expansions are monoclonal. For instance, the generalized proliferative response of all cells of a tissue to an extrinsic stimulus yields a polyclonal expansion, examples of which include the hyperthyroidism of Graves’ disease and the early, reversible secondary hyperparathyroidism in states of chronic hypocalcemia. Such polyclonal expansions represent biologic hyperplasia, whereas any monoclonal growth (benign or malignant) is a true neoplasm. Analyses of tumor clonality have been used to distinguish between these types of tumorigenic mechanisms. Nevertheless, the genesis of some tumors may involve both types of processes. For example, a generalized stimulus to polyclonal hyperplasia can, by increasing the chances of mitosis-related DNA damage in one cell, foster the emergence of a monoclonal population capable of eventually overwhelming or replacing its hyperplastic neighbors.

The clinical and histopathologic use of the term hyperplasia does not necessarily correspond to the biologic meaning described earlier, a situation that has engendered much confusion. For example, the usual primary hyperparathyroidism has been clinicopathologically classified as adenomas when a single gland is abnormal and as hyperplasia when the individual patient has multiple hypercellular glands. No histopathologic criteria can reliably predict whether a single or multiple glands are involved on the basis of analysis of only one such gland. Not only are clinical adenomas monoclonal neoplasms, but many parathyroid glands from patients with multigland “hyperplasia” are also monoclonal. It is therefore important to ensure that the terms used in the description of endocrine tumorigenesis are clearly defined.

Insights into Tumor Pathogenesis

The clonal status of a cellular proliferation is of fundamental importance in deciphering its pathogenesis; thus, endocrine tumors have been studied to determine whether the expansion is monoclonal or polyclonal. One way of determining that a tumor is monoclonal is to identify a DNA or chromosomal lesion that, because of its tumor specificity and presence in all or most of the neoplastic cells, directly defines the expansion as monoclonal. Examples of cytogenetically defined clonal abnormalities are chromosome translocations such as the t(9;22) Philadelphia chromosome in chronic myelogenous leukemia and the t(14;18) translocation in follicular lymphoma.

The use of classical cytogenetics is technically more difficult in solid tumors than in hematopoietic tumors because hematopoietic cells divide in culture much more readily and yield excellent metaphase chromosomal spreads. Specimens of endocrine tumors are difficult to obtain for culture and to analyze cytogenetically. Fortunately, improved methods for the cytogenetic and molecular cytogenetic study of tumors, including fluorescence in situ hybridization, comparative genomic hybridization, and chromosome painting, promise to open up new avenues for the detection of clonal chromosomal lesions in endocrine tumors.

Examples of monoclonal abnormalities defined by molecular methods in endocrine neoplasia include (1) Gαs gene mutations in growth hormoneproducing pituitary tumors; (2) thyrotropin (TSH) receptor gene mutations in thyroid tumors, and (3) cyclin D1 or PRAD1 gene rearrangements in parathyroid adenomas.

Identification of tumor-specific changes, such as deletions of DNA markers in particular regions of the tumor genome, also serves as evidence of monoclonality, even though the specific genes affected by such deletions may not be known.

Indirect methods can determine the clonal status of tumors without the necessity of identifying the specific genes or chromosomal regions that are clonally mutated and involved in tumorigenesis. These methods have generally exploited the phenomenon of random X chromosome inactivation (the Lyon phenomenon) in women.


Random X chromosome inactivation occurs early in female embryonic development in all somatic cells. In any cell, the choice of which X chromosome is inactivated is random; once that choice is made, however, the decision is faithfully transmitted to all progeny of that cell (Fig. 35-1A). Usually, therefore, the maternally derived X chromosome is inactive in about 50% of the cells in a normal tissue, and the paternally derived X chromosome is inactive in about 50%.

Polychromatization, representing a generalized expansion of many or all original cells within a tissue, maintains the relatively even mix of active maternal and paternal X chromosomes characteristic of the normal tissues. In contrast, the neoplastic cells within a monoclonal tumor are derived from a single progenitor and all should reflect an identical X chromosome pattern, with either the maternal or the paternal X chromosome uniformly inactivated (see Fig. 35-1A).

A unifying feature of methods based on the analysis of X chromosome inactivation to determine the clonal status of a tumor is the use of a normally occurring variant, or polymorphism, at the genetic or protein level to distinguish between a woman’s two X chromosomes. The other step involves assaying some property that reflects the state of X chromosome inactivation imposed on the tumor cell chromosomes at the polymorphic site. Assays to reflect X chromosome inactivation status include assessment of gene expression (RNA or protein levels) or regional DNA methylation.

A polymorphism in glucose-6-phosphate dehydrogenase (G6PD) was the first to be used in X chromosome inactivational analyses of tumor clonality. Although only a small minority of women are heterozygous for electrophoretically distinguishable isoforms of this X chromosome-encoded enzyme, thus most tumors have been unsuitable for clonal analysis. Furthermore, the method fails to detect the monoclonality of certain tumors, perhaps because of differences in the level of G6PD expression in tumor cells compared with contaminating admixed normal cells within the analyzed samples.

DNA polymorphisms are now preferred for clonal analyses to distinguish between the two X chromosomes. High rates of heterozygosity make it possible to analyze most tumors. Some multiallelic polymorphisms, based on differences in the number of highly repeated sequence units in a genomic location, are heterozygous in more than 90% of women, and a large number of DNA polymorphisms have been described on the X chromosome (and on all chromosomes). Most, however, cannot be used in clonality studies because they have not been characterized for detectable changes that correlate with the state of activity of the X chromosome on which an allele resides. Some of the X-linked polymorphisms that have been valuable in clonal analyses of human tumors include restriction fragment length polymorphisms in the HPRT and PGK gene regions, a minisatellite repeat (>10 nucleotide core repeated unit) region called M27 or DXS255 and a microsatellite repeat (<10 nucleotide core repeated unit) locus within the androgen receptor gene. Changes in DNA methylation in the vicinity of certain polymorphic sites on the X chromosome correlate with the activity of that X chromosome and are useful in clonal analyses. Methylation of specific cytotoxic nucleotides is an epigenetic process.

that is faithfully replicated from a given cell to its progeny, and this process may have a role in maintaining the activity status of the particular X chromosome. Methylation at certain specific sites may be easily detected through the action of methylation-sensitive restriction endonucleases, which cleave when their target sites are unmethylated but cannot do so when a methyl group is present.

One cannot predict how or whether X chromosome inactivation will affect the methylation of a particular nucleotide in or near a gene; correlations must be established separately for individual genomic sites. For example, a useful restriction site in the HPRT region is consistently methylated when on the active X chromosome and unmethylated when on the inactive X chromosome; the opposite pattern is observed for an informative site in the PGK region. Despite the need for such empirical validation, the use of DNA methylation as a surrogate marker for the status of X chromosome activity has major advantages in clonality studies and eliminates the vulnerability of the assay to the vagaries of gene expression in tumor cells.

The analysis of clonality in endocrine tumors has yielded important insights. For example, the fact that most corticotropin-producing pituitary tumors are monoclonal has shown that Cushings’s disease is not explained solely by a generalized hypothalamic stimulation of corticocortrophs. Typical parathyroid adenomas are also monoclonal outgrowths and are not, as previously suggested, asymmetric forms of multiglandular hyperplasia. Most solitary thyroid nodules are monoclonal, which might have been expected, but some are nodules within multinodular goiters.

One often overlooked pitfall in the interpretation of any X inactivation analysis must be emphasized: Polyclonal patterns cannot be interpreted definitively. For example, admixed normal cells or tumor-specific aberrancies in DNA methylation may obscure the detection of a monoclonal cell population. Thus, the true extent of monoclonality in severe secondary or tertiary hyperparathyroidism might be even higher (but not lower) than demonstrated. These unexpected results indicate that monoclonal parathyroid neoplasms are common in uremic refractory hyperparathyroidism and suggest that autonomous parathyroid function in this disorder is due to the outgrowth of true neoplasms, presumably on a background of preexisting (and more reversible) polyclonal parathyroid hyperplasia.

Figure 35-1X chromosome inactivation analysis of tumor clonality with the use of the M27 polymorphism. A, Diagrammatic illustration of general principles. Left. Lightly shaded and dark rectangles represent the maternally and paternally inherited copies of the two X chromosomes of somatic cells early in the development of a female embryo. As embryogenesis proceeds, one of the X chromosomes in each somatic cell is randomly chosen for inactivation (lyonization); the inactivated chromosome is represented as a small oval with shading corresponding to its origin. Subsequently, daughter cells (third column) faithfully maintain the same selection of inactivated X chromosome as found in their parent cells. Accordingly, an adult tissue typically contains a mixture of approximately 50% cells with the maternally inherited X chromosome inactive and 50% with the paternally inherited X chromosome inactive. Right, A polyclonal tumor arising from a large number of cells in a tissue, maintains this relatively even mixture of cells with different X inactivation patterns. B, Partial restriction endonuclease map of the M27 locus (DXS255) and an example of two distinguishable M27 alleles. The variable number of tandem repeat (VNTR, minisatellite) region, which is highly variable in its length from person to person, is shown in stripes. A 2.5-kb DNA fragment used as the hybridization probe is shown as a solid rectangle. Cleavage sites for restriction enzyme Pst I flank the locus. The enzyme Msp I cleaves the sequence CCGG whether or not the internal cytosine is methylated. In contrast, the enzyme Hpa II cleaves this sequence only if the internal cytosine is unmethylated. The diagrammed Msp I-Hpa II site actually represents a 270-base-pair region containing three such sites, two of which vary in their methylation status in accord with location on the active versus the inactive X chromosome. In the example to the two distinguishable alleles, variation in size of the minisatellite repeat region (VNTR) causes a difference in the size of the Pst I fragment restriction detectable by hybridization to the labeled M27 probe. If an individual with these two alleles had a monoclonal tumor, in which the larger allele was uniformly associated with the active X chromosome, the resulting Southern blot pattern would correspond to monoclonal pattern 1 in C. C, Schematic diagram of prototypical Southern blot hybridization patterns for X inactivation analysis using M27. A monoclonal tumor can exhibit only one of the two monochromatic patterns shown. The Pat + Msp I control digestion is useful for marking the sizes of fully cleaved alleles. (A to C, From Arnold A, Brown MF, Urena P, et al. Monoclonality of parathyroid tumors in chronic renal failure and in primary hyperparathyroidism. J Clin Invest 1995; 95:20472053. Copyright, The American Society for Clinical Investigation.)

Figure 35-1B and C illustrates the methodology used in a study of X chromosome inactivation in the pathogenesis of refractory secondary (or tertiary) hyperparathyroidism in patients with uremia. Because of multigland involvement, it has been assumed that this disease predominantly involves polyclonal (non-neoplastic) cellular proliferations, but the clonality of these “hyperplastic” tumors had not been comprehensively assessed. Clonality was examined with the M27 (DXS255) DNA polymorphism; 64% of the informative patients with renal failure undergoing hemodialysis and with refractory hyperparathyroidism harbored at least one monoclonal parathyroid tumor, and 63% of all tumors examined had monoclonal X inactivation patterns.

One often overlooked pitfall in the interpretation of any X inactivation analysis must be emphasized: Polyclonal patterns cannot be interpreted definitively. For example, admixed normal cells or tumor-specific aberrancies in DNA methylation may obscure the detection of a monoclonal cell population. Thus, the true extent of monoclonality in severe secondary or tertiary hyperparathyroidism might be even higher (but not lower) than demonstrated. These unexpected results indicate that monoclonal parathyroid neoplasms are common in uremic refractory hyperparathyroidism and suggest that autonomous parathyroid function in this disorder is due to the outgrowth of true neoplasms, presumably on a background of preexisting (and more reversible) polyclonal parathyroid hyperplasia.
Predisposing Influences

Both environmental and genetic factors can contribute to the risk of a particular type of tumor development over a lifetime. In highly penetrant inherited syndromes, the genetic predisposition is overriding; in most instances, however, the intimate relation between environmental and genetic factors makes the traditional “nature versus nurture” question difficult to assess.

A commonly overlooked genetic component of tumor predisposition is sex; however, the specific mechanisms by which sex influences the risk of endocrine tumors such as thyroid cancer, parathyroid adenoma, and adrenal cancer are not well understood. Additional genetic variables, which may be common in the population, also influence the chance of tumor development. For example, some independent studies have shown that one allele of a common nucleotide polymorphism within the vitamin D receptor gene is overrepresented in women with primary hyperparathyroidism. It is possible that this hyperparathyroidism-associated allele may yield fewer vitamin D receptors than the other common allele and thereby confer less of an antiproliferative effect on parathyroid cells, subtly but significantly increasing the lifetime risk of parathyroid neoplasia.

An important environmental factor in endocrine tumorigenesis is ionizing radiation. Both thyroid and parathyroid tumors have been linked to prior head and neck irradiation in a dose-dependent fashion. Whereas the latency period for tumor development after radiation exposure in the United States is quite long, childhood thyroid cancer has been observed with markedly increased frequency in the aftermath of the 1986 Chernobyl nuclear accident. Whether such heightened susceptibility is solely a function of age and amount of exposure or is related in part to other factors, such as iodine deficiency in the region, has not been established but highlights the point that tumor predisposition and development can be influenced by complex interacting factors.
Oncogenes and Tumor Suppressor Genes

Two broad categories of genes are implicated in the excessive cell proliferation and other properties that result in the outgrowth and evolution of a neoplastic clone. An oncogene carries a gain-of-function mutation in its regulatory or coding region that results in dysregulation of its normal product or in the formation of an abnormal protein product. Typically, only one mutated allele need be present for an oncogene to exert its tumorigenic action. The normal, unmutated version of an oncogene is called a proto-oncogene. Proto-oncogenes may be converted to oncogenes by various molecular genetic mechanisms, such as fusion of part of its coding region with that of another gene, chromosomal translocations or inversions that alter its regulatory environment, and point mutations.

A tumor suppressor gene, in contrast, normally acts to restrain cell proliferation or other potential aspects of the malignant phenotype. This restraint can directly control proliferation, for example, by regulating the cell cycle, or affect proliferation indirectly, for example, by maintaining genomic stability. Thus, the definition of a tumor suppressor gene is not restrictive concerning its specific cellular function. Tumors are provoked by inactivating or loss-of-function mutations in such genes; the most common inactivating mechanisms are gene deletion and point mutations. Both alleles of a “classical” tumor suppressor gene must be inactivated to eliminate the functional protein product and to contribute to tumorigenesis.

The existence of a critical tumor suppressor gene is often inferred from the frequent finding that a particular subchromosomal stretch of DNA is clonally lost in a particular tumor type. The deletion typically involves only one of a gene's two alleles. Because such deletions are often large, the case that the correct tumor suppressor gene has been found can be made most convincingly when the nondeleted allele of the gene is shown to harbor another clonal inactivating lesion, such as a coding region mutation or microdeletion. It is also possible that certain “nonclassic” tumor suppressors may contribute to neoplasia through haploinsufficiency, in which inactivation of only one allele provides the relevant functional alteration.

Although there has been some evolution in, and even controversy about, the best definition of a tumor suppressor gene, “the simplest, most inclusive, and cleanest genetic definition” is that favored by Haber and Harlow, namely “genes that sustain loss-of-function mutations in the development of neoplasia.” This definition, although unrestricted concerning the function of the tumor suppressor, “does have an essential component—the unequivocal demonstration of inactivating mutations.” Thus, evidence for reduced or absent expression of a gene in a particular tumor, or even the ability of a gene to inhibit cell proliferation, should not be accepted as sufficient to award it designation as a tumor suppressor if not accompanied by clear evidence of clonal mutation.

Some oncogenes or tumor suppressor genes contribute to tumors of only one or a few cell types, whereas other genes are involved in many different types of tumors. It is rare for any one particular gene to be invariably involved in the development of a given type of tumor, and different combinations of mutated genes may have similar cellular and clinical consequences (genetic heterogeneity).

Finally, genetic “hits” affecting multiple oncogenes and tumor suppressor genes within a single cell appear to be required if that cell is to become neoplastic. Thus, in most instances, no single activated oncogene or inactivated tumor suppressor gene is necessary or sufficient for the development of tumors.
Tumors may acquire mutations that in themselves increase the rate at which new mutations develop at other sites in the genome. Many tumor biologists have commented on the apparently excessive number of genetic changes observed in individual cancers compared with measured rates of mutation in normal cells. Genetic changes are usually considered to occur or to be fixed during DNA replication or mitosis. However, some mutations may be time-dependent rather than replication-dependent, meaning that they can arise even in the absence of cell proliferation. One mechanism by which somatic mutations may occur would be a defect in postreplication mismatch repair, the pathway responsible for removing and correcting mismatched base pairs in DNA. At least one form of inherited genetic instability can result from inactivation of "mutator" genes in the mismatch repair system, hMSH2 or hMLH1, for example. Although such mutational mechanisms might in theory be quite important in benign or slow-growing endocrine tumors that derive from slowly proliferating normal tissues, to date little direct evidence supports this hypothesis. Furthermore, some investigators have drawn conclusions about the extent of genomic instability in endocrine tumors based on the number of observed chromosomal abnormalities. Such conclusions cloud the distinction between "state" and "rate." One must note that the presence of genetic alterations in a tumor, even at high prevalence, does not necessarily imply that the tumor is genetically unstable. True instability is defined as an abnormal rate, and the snapshot of a mature tumor's complement of mutations (state) does not provide information about their rate of occurrence.

As noted by Lengauer and colleagues, several factors in addition to true instability may explain the greater frequency of observed mutations in tumors than in non-neoplastic tissues. These include potentially crucial differences in the selective conditions of the tumor versus normal cell environment, involving humoral and intercellular interactions. Thus, a given mutation might confer an important selective advantage and lead to clonal expansion only when it occurs in a tumor cell environment, whereas the same mutation (occurring at the same rate) in a normal cell environment would not lead to clonal expansion, might even cause apoptosis, and would thus avoid detection. It is not necessary to invoke a raised mutation rate in explaining the evolutionary process of tumor outgrowth, and often an increased mutation rate might well confer a disadvantage on the cell.

In summary, selection is the driving force for tumor growth; this key principle is as applicable to slow-growing and benign endocrine tumors as it is to aggressive cancers.
SOMATIC GENETIC ALTERATIONS IN ENDOCRINE TUMORS: EXAMPLES

Neoplasia is, in large part, a genetic disease in which most of the critical DNA damage occurs somatically rather than through inheritance (germ line mutations). Somatic alterations of both oncogenes and tumor suppressor genes are implicated in endocrine tumorigenesis. A few illustrative examples are presented here.

Thyrotropin Receptor and $G_s$ in Toxic Thyroid Adenomas

The fact that solitary toxic thyroid adenomas often contain somatic mutations in different genes whose products are functionally interrelated highlights the point that heterogeneous genetic lesions can converge on common pathways to predispose to neoplasia. Solitary toxic thyroid adenomas are characterized by autonomous (TSH-independent) hyperfunction and growth. Because TSH is normally a stimulus to both functional activity and growth of thyroid cells, toxic adenomas behave as though they are stimulated by TSH chronically and inappropriately. Somatic mutations within the TSH receptor itself can lead to TSH-independent (constitutive) activation of the receptor and cause both clonal expansion and hyperthyroidism in a subset of toxic adenomas.\(^7\)\(^43\) Analysis of how these point mutations mimic the effects of TSH binding has provided insight into the mechanism of activation of $G$ protein-coupled receptors.

The TSH receptor is a member of the large family of $G$ protein-coupled receptors, and its major actions on thyrocyte proliferation and differentiated function are mediated through the cyclic adenosine monophosphate (cAMP) signaling pathway. The predominant $G$ protein involved in transducing the stimulatory effect of the receptor on adenylyl cyclase is $G_s$. $G_s$ is also a proto-oncogene and has undergone clonally selected, activating point mutation in a subset of toxic thyroid adenomas.\(^7\)\(^9\)\(^44\) $G_s$ genes bearing such gain-of-function mutations have been termed $gsp$ oncogenes and are also present in some growth hormone-secreting pituitary tumors. Not surprisingly, these pituitary tumors demonstrate constitutive activation of the cAMP pathway, a prime mediator of proliferation and hormonal function in the somatotroph.

In this context, it is instructive to raise the example of McCune-Albright syndrome (see Chapter 24 and see Chapter 27), in which activating $G_s$ mutations are present in multiple tissues of a given patient owing to mutation early in embryonic development and subsequent genetic mosaicism. Hyperthyroidism and acromegaly are among the characteristic components of this syndrome. Another gene encoding an element in the cAMP signaling pathway, the protein kinase A type I- regulatory subunit gene $PPKAR1A$, has been implicated in the familial syndrome termed Carney’s complex, which includes corticotropin-independent Cushing’s syndrome related to autonomously functioning adrenocortical nodules, growth hormone-secreting pituitary tumors, and male precocious puberty caused by hormonally active testicular tumors.\(^45\)
Cyclin D1 (PRAD1) in Hyperparathyroidism and General Oncology

Another illustrative example is that of the cyclin D1, or PRAD1, oncogene. Unlike many endocrine tumor-associated oncogenes, cyclin D1 is also commonly involved in nonendocrine tumors. Interestingly, this gene, now appreciated to be of central importance to molecular oncology and to normal cellular physiology, was discovered in the molecular dissection of an endocrine tumor.

Many human oncogenes were discovered because they are adjacent to nonrandom chromosome breakpoints in tumors. Chromosome breaks and rearrangements probably occur often in normal cells but are recognized only when they result in deregulation of the expression of a growth-related gene and confer a selective advantage on the cell. Cyclin D1 or PRAD1 was identified as the putative oncogene adjacent to one such breakpoint on chromosome 11 in a subset of parathyroid adenomas (Fig. 35-2). On the 11q13 side of the inversion breakpoint, the promoter and coding exons of the cyclin D1 gene remain in contiguity with each other. Across the breakpoint are regulatory sequences from the upstream region of the parathyroid hormone (PTH) gene on 11p15 that normally function to enhance PTH gene transcription in the presence of parathyroid tissuespecific signals (likely to be DNA binding proteins found in the nucleus). Such transcriptional enhancer sequences can act over distances of many kilobases to enhance transcription, and cyclin D1 transcription is increased in these tumors. Although the variability in potential breakpoint sites makes it difficult to determine the true incidence of such rearrangements, cyclin D1 protein levels are elevated in 20% to 40% of parathyroid adenomas.

In a broader context, the tissue-specific enhancer-driven expression of cyclin D1 in parathyroid tumors is analogous to the activation of oncogenes such as BCL2 or MYC in B-cell lymphomas. In these tumors, chromosomal rearrangements lead to a juxtaposition of immunoglobulin gene enhancer elements and the oncogene, which is thereby inappropriately activated.

One cautionary note: This functional assignment for cyclin D1 is overwhelmingly derived from cultured cell systems, which might not fully reflect the true in vivo roles of this key oncoprotein. Thus, the detailed mechanism of action of cyclin D1, both normally and when dysregulated in tumorigenesis, should be further explored.

Cyclins are regulatory subunits of holoenzymes whose catalytic subunits are cyclin-dependent kinases (CDKs). The major kinase partner for cyclin D1 appears to be CDK4 or, in some cell types, CDK6. The protein product of the retinoblastoma tumor suppressor gene, pRB, has been recognized as one substrate for cyclin DCDK4/6 complexes. Whether all such relevant phosphorylations occur in G1 phase progression or might in part push quiescent G0 cells into an active G1 cycling mode, for example, requires further study.

Natural inhibitors of CDK function also exist, and p16INK4a is recognized as a key inhibitor of cyclin DCDK4/6 complexes. Thus, inactivation of p16 might be expected to be as oncogenic as cyclin D1 overexpression, and p16 is, indeed, a tumor suppressor gene in familial melanoma and several types of sporadic human tumors. Interestingly, inactivating mutations of p16 are uncommon, if they occur at all, in parathyroid adenomas. Hence, the cellular consequences of p16 loss and cyclin D1 activation may not precisely overlap. Interestingly, some early evidence exists for possible noncdk-dependent actions of cyclin D1.

The significance of cyclin D1 in human neoplasia extends far beyond its involvement in endocrine tumors. It is the long-sought BCL1 oncogene that is deregulated by the characteristic t(11;14) translocation in mantle cell or centrocytic B-cell lymphomas. Thus, assessment of cyclin D1 gene rearrangement or expression is clinically useful in the molecular diagnosis of B-cell neoplasia. In addition, cyclin D1 is a key oncogene in breast cancer, squamous cell cancer of the head and neck, esophageal cancer, and a variety of other tumors. Cyclin D1 and other members of its oncogenic pathway or pathways may serve as targets for development of antineoplastic therapies.
RET Gene Rearrangements in Papillary Thyroid Cancer

Activating mutations of the RET proto-oncogene are responsible for multiple endocrine neoplasia type 2 (see Chapter 36), but this gene is also involved in human disease through another mechanism: Clonal somatic rearrangements of this gene are found in about 25% of papillary thyroid cancers but not in other types of thyroid (or nonthyroidal) cancer. Inactivating germ line mutations of RET are a cause of Hirschsprung's disease, in which parasympathetic ganglia fail to migrate properly in the gastrointestinal tract (see Chapter 36).

RET encodes a member of the receptor tyrosine kinase superfamily, and one endogenous ligand that binds to and activates RET is the glial cell line-derived neurotrophic factor (GDNF). Normally, RET acts mainly during embryogenesis.

The RET protein, like other receptor tyrosine kinases, has extracellular, transmembrane, and intracellular (tyrosine kinase) domains (Fig. 35-3). The papillary cancerspecific RET rearrangements described to date involve a chromosomal break that fuses the intracellular tyrosine kinase domainencoding portion of the RET gene to one of several alternative partners. The RET fusion partner gene segment provides a new promoter that is constitutively active in thyroid cells and encodes a new in-frame N-terminus for the oncoprotein (see Fig. 35-3).

Three major classes of RET fusion oncogenes are called RET-PTC1, RET-PTC2, and RET-PTC3; the corresponding fusion partner genes are H4, R1, and ELE1.

The N-termini of these fusion oncogenes allow dimerization and activation of the RET tyrosine kinase, bypassing the usual requirement for ligand binding. The specificity of these RET fusion oncogenes for papillary-type thyroid cancer is not well understood. It appears that only this thyroid cell type tends to produce such fusions or that only this cell has the molecular machinery that can respond to this form of RET kinase activation. Perhaps development of a papillary cancer may be an inevitable response to the occurrence of the RET rearrangement early in the life of a thyroidal neoplasm.

Finally, the RET proto-oncogene and its fusion partners, especially ELE1, may be highly susceptible to breakage and fusion as a consequence of ionizing radiation. More than 60% of papillary cancers that have developed in young people exposed to the nuclear fallout from the Chernobyl reactor contain RET-PTC fusion oncogenes. The spatial contiguity of RET and its rearrangement partners in the interphase thyroid cell nucleus may provide a structural basis for the occurrence of such rearrangements by enabling a single radiation particle track to cause a double-stranded DNA break that is misrepaired.
GERM LINE MUTATIONS PREDISPOSING TO ENDOCRINE NEOPLASIA: EXAMPLES

Patients in whom a particular tumor develops on the basis of a strong inherited predisposition typically constitute only a minority of patients with that tumor. Nonetheless, the lessons learned from identifying the molecular basis of inherited tumor predisposition are important, both clinically and from a fundamental biologic perspective. In addition, some genes in which germ line mutations cause rare genetic syndromes have subsequently been found to be somatically mutated in the more common, sporadic occurrences of the same tumors.

Germ line mutations that predispose to neoplasia can occur in either proto-oncogenes or tumor suppressor genes; in other words, mutations that cause inherited tumor syndromes can be of either the gain-of-function or the loss-of-function type. However, most germ line mutations identified to date are inactivating mutations, thus identifying the affected genes as tumor suppressors by definition. A few genes responsible for heritable endocrine neoplasia syndromes have been discovered, and others are still being sought.

RET Mutations in Multiple Endocrine Neoplasia Type 2

Multiple endocrine neoplasia type 2 (MEN 2) is a collection of three syndromes:

- MEN 2A
- Familial medullary thyroid cancer
- MEN 2B

These syndromes are detailed in Chapter 36.

All are autosomal dominant syndromes in which the most consistent feature is predisposition to medullary thyroid cancer. Genetic linkage analysis demonstrated that the MEN 2 gene or genes must lie near the centromere on chromosome 10, and an evaluation of candidate genes known to map to this region led to the identification of germ line RET point mutations that co-segregated with the disease. The existence of de novo RET mutations in patients with an MEN 2 phenotype, but a negative family history and the demonstration that the RET mutations activate RET as a transforming oncogene further helped establish that germ line RET mutations are the predominant cause of MEN 2.

Only a few codons (for cysteine residues in the extracellular domain) harbor the RET mutations in almost all patients with MEN 2A and most patients with familial medullary thyroid cancer, and only two codons (the predominant one encodes a conserved threonine within the tyrosine kinase domain's catalytic core) are mutated in virtually all patients with MEN 2B. This molecular specificity contrasts with the many ways in which tumor suppressor genes are typically found to be inactivated and provided the first clue that RET mutations in MEN 2 are activating mutations. In fact, MEN 2 was the first inherited human tumor syndrome found to be due to an activated oncogene in the germ line. The limited number of locations for RET mutations in MEN 2 makes it relatively easy to detect the mutations and make the molecular diagnosis. This methodology is available commercially.

At a time when the general concept of germ line DNA testing for cancer predisposition has been enveloped in controversy, MEN 2 has provided the best example in oncology of the efficacy of molecular diagnosis in mainstream clinical management. Specific features not often present in other cancer predispositions that have contributed to the impact of the molecular diagnosis of MEN 2 include:

1. The high penetrance of a life-threatening cancer (medullary thyroid carcinoma).
2. The ease and accuracy of DNA diagnosis (better than with previously available pentagastrin testing).
3. The availability of a lifesaving intervention (thyroidectomy) without major risk.

As alluded to earlier, somatic RET mutations (much more often of the MEN 2B type than the MEN 2A type) are also present in a minority of sporadic medullary thyroid carcinomas and pheochromocytomas, but not in primary hyperparathyroidism.

In MEN 2, although the germ line RET mutation is present in all somatic cells of the affected individual, the MEN 2-related tumors are not, strictly speaking, inherited but develop from a hyperplastic endocrine cell mass into clonal neoplasms over time. The monoclonality of the medullary thyroid carcinomas and pheochromocytomas in MEN 2 indicates that somatic molecular mutations, probably affecting several oncogenes or tumor suppressor genes, or both, must occur in a single cell and act in concert with the inherited germ line RET mutation for the neoplasms to emerge. In fact, several such clonal somatic molecular lesions, including allelic deletions on chromosomes 1, 3, 11, 17, and 22, have been found in MEN 2-associated tumors, pointing to the genomic locations of important but unidentified putative tumor suppressor genes.
MEN1 Tumor Suppressor Gene

Multiple endocrine neoplasia type 1 (MEN 1) is a familial predisposition to tumors of the parathyroid glands, anterior pituitary, pancreatic islets, and duodenum. Other tumors, including carcinoid tumors, adrenal adenomas, lipomas, and angiofibromas, also occur with increased frequency in this disorder (see Chapter 36).

MEN 1 is inherited in an autosomal dominant pattern, indicating that a single mutant gene is responsible for transmitting the tumor predisposition. Genetic linkage analysis implicated a region on the long arm of chromosome 11 (11q13) as the site of the MEN1 gene (i.e., the site of the normal gene whose mutant form causes MEN 1). However, this target region on 11q13 contains a large number of genes, resulting in a long and arduous positional cloning effort before the successful identification of MEN1.

Even before it was identified, considerable evidence indicated that MEN1 is a tumor suppressor gene. One of the two cellular copies of the MEN1-linked chromosomal region was frequently found to be somatically and clonally deleted in parathyroid and pancreatic tumor tissues from patients with MEN 1. When the parental origin of the deleted allele could be identified, it always derived from the unaffected parent. Thus, patients with MEN 1 were believed to have inherited one inactivated copy of the critical gene from the affected parent, with the remaining normal copy becoming somatically inactivated, often by a large regional deletion, in a clonal progenitor cell from which the tumor would emerge. Because the germ line aﬀected MEN1 chromosomal region appeared to be grossly intact in most instances, it seemed likely that most germ line mutations in the MEN1 gene would be small deletions or point mutations that would require sequencing of DNA for detection. Direct testing made possible by the cloning of the MEN1 gene has conﬁrmed these hypotheses.

MEN1 germ line mutations have been detected in most MEN 1 families and in so-called sporadic MEN 1, that is, where no positive family history exists. Many of these mutations would clearly be expected to truncate the translated product severely and can safely be categorized as inactivating mutations. It can be reasonably anticipated that when deﬁnitive functional testing is available, other reported mutations, especially in the missense category, will prove to be inactivating lesions as well. The identiﬁcation of MEN1 has also opened up the potential for direct and presymptomatic molecular diagnosis in established or suspected MEN 1 kindreds or individuals. However, the beneﬁts of such testing over existing methods with respect to prophylactic interventions and improvements in morbidity or mortality are not well established, in strong contrast to the situation for RET testing in MEN 2.

The MEN1 gene encodes a 610-amino-acid protein termed menin. The structure of menin has yielded few clues to menin’s normal function, except for its nuclear localization signals. That menin may play a role in regulation of gene transcription has been suggested by the finding that it can bind a member of the activator protein 1 (AP-1) transcription factor family, JunD, and inhibit the ability of JunD to activate transcription. The interaction of menin with JunD carries the potential to link its function with the regulation of cell senescence and apoptosis, given insights into the role of JunD in p53-dependent pathways. Tissue surveys have shown MEN1 expression to be nearly ubiquitous, and the reason for endocrine tissue speciﬁcity in the ultimate development of tumors remains to be deciphered.

As previously stated, the molecular pathology of MEN1 strongly suggests that it functions as a classical tumor suppressor gene, requiring biallelic inactivation in order to drive the emergence of a clinically signiﬁcant tumor. In a patient with MEN 1, after somatic deletion on 11q13 has occurred in a particular parathyroid or islet cell, for example, that cell would be devoid of the normal tumor suppressor function of the MEN1 gene product and would thereby acquire a selective advantage over its neighbors. The high incidence of endocrine tumors in MEN 1 (which demonstrates almost 100% penetrance) implies that somatic inactivation of the remaining normal gene copy is a common development in the context of the patient’s entire endocrine tissue complement. It does not mean, however, that loss of function in both MEN1 alleles is suﬃcient for tumorigenesis, and additional cooperating oncogenic lesions may be needed.

Allelic losses of DNA markers on 11q13, including the MEN1 region, are found in some of the more common, sporadically occurring versions of the tumors associated with familial MEN 1. They include a frequency of 25% to 40% for 11q13 allelic loss in sporadic parathyroid adenomas, for example. Interestingly,
Summary

Genetic paradigms that have originated in the study of malignant neoplasia are useful in the molecular dissection of endocrine tumors, including common benign endocrine tumors. Identification of clonally selected mutations in oncogenes and tumor suppressor genes has opened the door to understanding the control of growth in endocrine tissues. Furthermore, the fact that these genetic alterations can affect endocrine cell function as well as cell number may be exploited in devising medical therapies in the future.
References


42. Tomlinson I, Bodmer W. Selection, the mutation rate and cancer: ensuring that the tail does not wag the dog. Nat Med 1999; 5:1112.


The multiple endocrine neoplasia (MEN) syndromes were described early in this century\(^1\) and subsequently classified into two principal categories: MEN type 1 (MEN1) and MEN type 2 (MEN2). The MEN2 syndrome has been further subcategorized into two principal variants called MEN2A and MEN2B (formerly MEN3). Our understanding of the MEN syndromes has evolved through several phases. The first was the descriptive phase in which the fully developed clinical syndromes and their familial patterns were described.\(^2\)\(^3\)\(^4\)\(^5\)\(^6\)\(^7\)\(^8\)\(^9\)\(^10\)\(^11\)\(^12\)\(^13\)\(^14\)\(^15\)\(^16\)\(^17\) The second phase involved the development of tumor surveillance techniques to identify the syndromes before they became significant clinical problems.\(^9\)\(^10\)\(^11\)\(^12\)\(^13\)\(^14\)\(^15\)\(^16\)\(^17\) Early identification of the manifestations and improvements in management techniques have had a significant impact on the morbidity and mortality associated with these syndromes.\(^18\) We are now in the third phase, the elucidation of the genetic and molecular basis of these syndromes. This chapter summarizes the current understanding of these disorders and provides a framework for understanding future developments.

MEN1 and MEN2 share certain characteristics. First, the usual tumor is composed mainly of cells with the capability to secrete one or more peptide hormones or small amines. This defines the tumor as endocrine (secretory) in nature. Second, the tumors in MEN1 and MEN2 are often benign with major clinical effects caused by hormone hypersecretion. As a corollary, there are often excellent therapies based on medications or surgery. Third, malignant transformation of certain cell types is also a component of each syndrome. Fourth, as with many hereditary tumors, some tumors in MEN1 and MEN2 occur relatively early and some with multiplicity; multiplicity is seen here as multiple foci within a tissue and as tumors in multiple tissues. Last, both of these syndromes have an autosomal dominant pattern of inheritance.

The MEN syndromes provide clear-cut examples of two different genetic mechanisms for tumorigenesis. MEN1 tumors are caused by loss of function, inactivation of a tumor suppressor gene, MEN1. In contrast, MEN2 tumors are caused by gain of function, activation of a proto-oncogene, RET. There are also other important contrasts between MEN1 and MEN2. Most important is that the primary C-cell malignancy of MEN2, medullary thyroid carcinoma (MTC), can be recognized at the earliest stages and prevented or cured. This is an option, in part, because MTC arises within the thyroid gland, so that complete excision is possible. In contrast, malignancies associated with MEN1 are difficult to recognize, and they occur in organs (pancreas, duodenum, bronchi) whose ablation would carry unacceptable morbidity.
MULTIPLE ENDOCRINE NEOPLASIA TYPE 1

The association of parathyroid, enteropancreatic endocrine, and pituitary neoplasia is called multiple endocrine neoplasia type 1. Although there were earlier descriptions, the syndrome was recognized as a clinical and familial syndrome by Moldawer and colleagues in 1964 (thus the eponym Wermer's syndrome). MEN2 was recognized and classified as distinct from it in 1968. In previous years, MEN1 patients presented with advanced manifestations of parathyroid, pancreatic islet, and/or pituitary neoplasia in the third and fourth decades of life. However, improved carrier ascertainment and improved tumor surveillance have now resulted in earlier identification of hormonal syndromes.

The most common mode of presentation for MEN1 is currently within the context of an identified kindred; less frequently, a newly ascertained individual with advanced disease may be the propositus of a new kindred, a previously unidentified member of a known kindred, or an example of a de novo mutation. Despite its earlier discovery, MEN1 remains the most challenging of the MEN syndromes. The many affected tissues cause complexity and expense in diagnosis and treatment. Each affected patient can be expected to undergo at least two surgical procedures. It is important for the clinician to recognize the high probability of recurrent or new neoplasms in all affected organ systems and to balance this likelihood against the potential effects of a deficiency syndrome associated with complete organ removal. Furthermore, even with satisfactory control of symptoms from hormone excess, cases have a high likelihood of eventual MEN1-related cancer.

TABLE 36-1 -- Features of Multiple Endocrine Neoplasia Type 1 with Estimated Average Penetration (in parentheses) among Adults

<table>
<thead>
<tr>
<th>Endocrine Features</th>
<th>Nonendocrine Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parathyroid adenoma (95%)</td>
<td>Facial angiofibroma (85%)</td>
</tr>
<tr>
<td>Enteropancreatic</td>
<td>Collagenoma (70%)</td>
</tr>
<tr>
<td>Gastrinoma (40%)*</td>
<td>Lipoma (30%)</td>
</tr>
<tr>
<td>Insulinoma (10%)</td>
<td>Leiomyoma (5%)</td>
</tr>
<tr>
<td>Nonfunctioning, including pancreatic polypeptide-oma (20%)</td>
<td>Ependymoma (&lt;1%)</td>
</tr>
<tr>
<td>Other: glucagonoma, VIPoma, somatostatinoma, etc. (each &lt;2%)</td>
<td></td>
</tr>
<tr>
<td>Foregut carcinoid</td>
<td></td>
</tr>
<tr>
<td>Thymic carcinoid nonfunctioning (2%)</td>
<td></td>
</tr>
<tr>
<td>Bronchial carcinoid nonfunctioning (4%)</td>
<td></td>
</tr>
<tr>
<td>Gastric enterochromaffin-like tumor nonfunctioning (10%)</td>
<td></td>
</tr>
<tr>
<td>Anterior pituitary</td>
<td></td>
</tr>
<tr>
<td>Prolactinoma (25%)</td>
<td></td>
</tr>
<tr>
<td>Other: nonfunctioning (10%), growth hormone + prolactin, growth hormone (both 5%), ACTH (2%), thyrotropin (rare)</td>
<td></td>
</tr>
<tr>
<td>Adrenal</td>
<td></td>
</tr>
<tr>
<td>Cortex</td>
<td></td>
</tr>
<tr>
<td>Nonfunctioning (30%)</td>
<td></td>
</tr>
<tr>
<td>Functioning or cancer (2%)</td>
<td></td>
</tr>
<tr>
<td>Medulla: pheochromocytoma (&lt;1%)</td>
<td></td>
</tr>
</tbody>
</table>

ACTH, adrenocorticotropic hormone; VIP, vasoactive intestinal peptide.

*Italics indicate tumor type with substantial (>20%) of cases) malignant potential.

Many "nonfunctioning" MEN1 tumors synthesize a peptide hormone or other factors (such as small amine) but do not oversecrete enough to produce a hormonal expression. Omits nearly 100% prevalence of nonfunctioning and clinically silent tumors, some of which are detected incidental to enteropancreatic surgery in MEN1.

Hyperparathyroidism in Multiple Endocrine Neoplasia Type 1

MEN1 is uncommon with a population prevalence of about 1 in 30,000 and accounts for only about 1% to 3% of cases of primary hyperparathyroidism. Hyperparathyroidism is the most common hormonal manifestation of MEN1 (Table 36-1). Successful subtotal parathyroidectomy is also followed within 10 years by recurrent hyperparathyroidism in half of MEN1 cases. In fact, true recurrent hyperparathyroidism in sporadic hyperparathyroidism is unusual, and recurrence should suggest the possibility of MEN1. True recurrent hyperparathyroidism, as with other tumor recurrences in MEN1, could arise theoretically from a small remnant of tumor tissue or from new mutation (second hit) in residual normal tissue. Fourth, hyperparathyroidism in MEN1 almost never progresses to parathyroid cancer, despite the fact that hyperparathyroidism occurs earlier in MEN1 than in sporadic cases.

Expressions of Hyperparathyroidism

Hyperparathyroidism in MEN1 is most frequently asymptomatic; expressions include hypercalcemia, urolithiasis, parathyroid hormone (PTH)-induced bone abnormalities, musculoskeletal complaints, weakness, and alterations of mental status. These features are similar to those associated with other forms of hyperparathyroidism (see Chapter 26).

Hyperparathyroidism in MEN1 differs in some ways from that caused by a sporadic adenoma. The first way is a difference in epidemiology. Hyperparathyroidism in MEN1 has an earlier age of onset (typically 25 years versus 55 years) (see Chapter 26). By age 40 years, about 95% of MEN1 carriers have been hypercalcemic. The second is a different parathyroid pathology; enlargement, albeit highly asymmetric, of multiple parathyroid glands is usually present at the time of parathyroid exploration in MEN1 (Fig. 36-2). Third, the distributions of outcomes of parathyroid surgery differ. The presence of multiglandular disease and the need to examine each gland during a surgical procedure inevitably result in a higher postoperative rate of hyperparathyroidism and a lower rate of euthyroidism. Successful subtotal parathyroidectomy is also followed within 10 years by recurrent hyperparathyroidism in half of MEN1 cases. In fact, true recurrent hyperparathyroidism in sporadic hyperparathyroidism is unusual, and recurrence should suggest the possibility of MEN1. True recurrent hyperparathyroidism, as with other tumor recurrences in MEN1, could arise theoretically from a small remnant of tumor tissue or from new mutation (second hit) in residual normal tissue. Fourth, hyperparathyroidism in MEN1 almost never progresses to parathyroid cancer, despite the fact that hyperparathyroidism occurs earlier in MEN1 than in sporadic cases.

There are several characteristics of hyperfunctioning parathyroid cells in MEN1 that may have mechanistic implications. First, most or all parathyroid glands have been overgrown by one or a few neoplastic clones by the time of surgery in MEN1 (Fig. 36-3, top). Second, a circulating growth factor is specific to the plasma of MEN1 cases and mitogenic toward normal parathyroid cells in vitro (see later). Third, a phenomenon observed in sporadic parathyroid adenomas, a rightward shift in the set-point for calcium suppression of PTH secretion, occurs to a lesser extent in MEN1 parathyroid tumors (see Chapter 26). This set-point abnormality may

Larsen: Williams Textbook of Endocrinology, 10th ed., Copyright © 2003 Elsevier
be caused by secondary decrease in the amount of the calcium-sensing receptor on the parathyroid cell surface.

Hyperparathyroidism Management

Decision for Surgery

Surgery is the treatment of choice for hyperparathyroidism in MEN1, although the timing and the type of operation remain controversial. Parathyroid surgery is definitely indicated in a MEN1 patient with an elevated PTH and other moderately advanced features, such as an albumin-adjusted serum calcium level higher than 3.0 mmol/L (12.0 mg/dL), kidney stones, or PTH-induced bone disease.

An alternative pathogenesis could be stepwise evolution from one clone, that is, third hits to genes other than those encoding parathyroid hormone.

Minutes after inflation of a blood pressure cuff to occlude venous return from the graft, the graft is implicated as the main source of PTH.

Contralateral arm confirms graft function (this does not exclude other parathyroid tissue in the neck or chest). Similarly, if peripheral PTH falls by more than 50% within 45 minutes of the graft, graft function is expected to have multiple parathyroid tumors (as in MEN1). These tests are even more likely to be helpful during parathyroid reoperations in MEN1, as the number of parathyroid tumors is likely to be much greater in these patients.

Diagnosis of Parathyroid Gland Multiplicity

Sensitive ultrasound transducers routinely image parathyroid tumors intraoperatively in difficult locations, such as within the thyroid gland and within scar from prior surgery. Rapid "on-line" assay of PTH can be done at 5-minute intervals with a turnaround time of 10 minutes for each result.

Several intraoperative tools can increase the likelihood of successful parathyroid surgery. Rapid "on-line" assay of PTH can be done at 5-minute intervals with a turnaround time of 10 minutes for each result.

A PTH fall of 50% or more from baseline predicts that no hyperfunctioning parathyroid tissue remains. Sensitive ultrasonic transducers routinely image parathyroid tumors intraoperatively in difficult locations, such as within the thyroid gland and within scar from prior surgery.

Figure 36-2 Parathyroid gland sizes at initial parathyroidectomy for 18 cases with familial multiple endocrine neoplasia type 1. Volumes of all glands at one operation are connected by a vertical line. Dashed horizontal line is upper limit of normal gland volume (0.075 cm³, equivalent to 75 mg mass). (Modified from Marx SJ, Menzel J, Campbell G, et al. Heterogeneous size of the parathyroid glands in familial multiple endocrine neoplasia type 1. Clin Endocrinol (Oxf) 1991; 35:521-526.)

Figure 36-3 Tumor multiplicity within a tissue in multiple endocrine neoplasia type 1 (MEN1). Top, Hypercellular parathyroid gland from patient with MEN1. The gland is totally replaced by diffuse sheets and two discrete nodules of chief cells. This image could reflect three or more second hits to the normal copy of the MEN1 gene in three different clone precursor cells and thus growth of three or more independent clones. An alternative pathogenesis could be stepwise evolution from one clone, that is, third hits to genes other than MEN1. Bottom, Duodenal mucosa from a second MEN1 patient, showing two large, submucosal microgastrinomas. Each tumor was positive for gastrin immunostain and negative for other peptide hormones. Possible development of these tumors could be caused by secondary decrease in the amount of the calcium-sensing receptor on the parathyroid cell surface.
two adjacent tumors could have followed mechanisms suggested for the two parathyroid nodules at the top. (From I. Lubensky, National Institutes of Health, Bethesda, Md.)

peripheral PTH. Lastly, if curative, surgical removal of parathyroid tissue from the forearm graft bed during the inevitable second or third operation is technically easier than a neck reexploration. Cryopreservation of parathyroid tumor fragments is a useful option in MEN1, given the high rate of postoperative hypoparathyroidism in MEN1. Cryopreservation permits a late parathyroid autograft. [46]

At initial parathyroid operation in MEN1, partial thymectomy should be done through the cervical incision. This procedure not only results in removal of intrathymic parathyroid tissue but also can remove thymic carcinoid tissue, an issue discussed later in this chapter. It is important to do this during

![Figure 36-4](image)
**Figure 36-4** Effect of parathyroidectomy in patients with multiple endocrine neoplasia type 1 and Zollinger-Ellison syndrome. Basal acid output and fasting serum gastrin are shown. All patients became normocalcemic except for one (case 4), who remained hypercalcemic. Dashed line is upper limit of normal for serum gastrin. Upper limit of normal for basal acid output is 15 mEq/hour. (From Jensen RT. Management of the Zollinger-Ellison syndrome in patients with multiple endocrine neoplasia type 1. J Intern Med 1998; 243:477-488.)

![Figure 36-5](image)
**Figure 36-5** Intact parathyroid hormone (PTH) by rapid assay during parathyroid surgery. Normal range is indicated by dashed lines. The patient had multiple endocrine neoplasia type 1 and primary hyperparathyroidism without prior parathyroidectomy. Three and one half similarly enlarged parathyroid glands (0.8 to 1.6 g; normal less than 0.08 g) and the accessible portions of the thymus were removed at the times indicated; the thymus contained no parathyroid tumor. A rapid fall of PTH below a cutoff criterion (such as 50% drop within 5 minutes) indicates that little or no hyperfunctioning parathyroid tumor remains. Note that removal of the first two parathyroid tumors was not followed by a fall in PTH. The PTH assay result for each time point was available within several minutes to help establish the time at which no hyperfunctioning parathyroid tumor remained and thereby contribute to serial decisions about extending or ending the operation. (From S. K. Libutti, H. R. Alexander, and A. Remaley, National Institutes of Health, Bethesda, Md.)

![Figure 36-6](image)
**Figure 36-6** Serum ionized calcium and peripheral midregion parathyroid hormone (PTH) levels after total parathyroidectomy and graft of parathyroid tissue in a patient with multiple endocrine neoplasia type 1. After total parathyroidectomy (PTX) and grafting of parathyroid tissue to the nondominant forearm, the ionized calcium and PTH levels fell to subnormal values. Subsequent measurements demonstrated a continuous rise of the ionized calcium and peripheral PTH levels (taken from the arm not containing the transplant) over a 60-month period, which necessitated removal of some of the grafted parathyroid tissue. The numbers in parentheses represent selected PTH values for blood taken from the brachial vein immediately downstream from the grafted parathyroid tissue. These results demonstrate continued secretion of PTH by the graft in the presence of hypercalcemia. The upper, lighter shaded area shows the normal range for ionized calcium; the lower, darker shaded area shows the normal range for serum midregion PTH. (Data from L. E. Mallette, Baylor College of Medicine.)

the primary operation because scar tissue may prevent a simple transcervical thymectomy at parathyroid reoperation.

Parathyroid surgery in the patient with MEN1 requires judgment, familiarity with neck anatomy, and experience. The probability of an excellent outcome is improved substantially when primary or reoperative procedures are performed by an experienced endocrine surgery team.
Enteropancreatic Neuroendocrine Tumors

Neoplasia of the enteropancreatic neuroendocrine cells is the second most common endocrine manifestation of MEN1 and eventually occurs in about 60% or more of MEN1 patients (see Table 36-1). Also, multiple clinically silent enteropancreatic macroadenomas may be recognized at surgery or autopsy in nearly 100% of MEN1 cases older than 40 years. The tumors are multiple, can oversecrete various hormones, and can become malignant. Approximately one third of MEN1 patients die from MEN1-related cancer; among these fatal cases, the ratio of enteropancreatic neuroendocrine cancer to malignant carcinoid is 2:1. It is similarly notable that, in this era of excellent pharmacotherapy for gastric hyperacidity, about one sixth of MEN1 patients with gastrinoma may die from metastatic gastrinoma but rarely from hypergastrinemia-induced metabolic complications.

Although the MEN1 patient may show symptoms or signs caused by one enteropancreatic hormone, there are often several associated and asymptomatic tumors with production of the same or other hormones. The frequency of peptide immunostaining in MEN1 pancreatic islet tumors is glucagon 35%, insulin 25%, pancreatic polypeptide 25%, and no hormone 10%; limited data are available for immunostaining of commonly observed duodenal gastrinomas. Interpretation of pancreatic islet histology in MEN1 has changed over the decades. Early studies emphasized hyperplastic processes and budding of islet cells from ducts (nestsidioblastosis). Such features have now been reinterpreted as non-specific. The overriding and important islet lesion in MEN1, now termed multifocal microadenoma, is a mononuclear or oligonuclear process (see Fig. 36-3, legend). No earlier precursor stage is currently recognized. However, heterozygous knockout of the MEN1 gene in the mouse provides a good model of human MEN1; multiple giant hyperplastic islets are striking and precede insulinaemia in this model, suggesting that subtle islet hyperplasia is an undetected tumor precursor lesion in MEN1 of humans.

Rarely, enteropancreatic neuroendocrine tumors occur in several members of a family without other features of MEN1; these have been insulinomas. Nonfunctioning pancreatic islet tumors and pheochromocytoma can also appear uncommonly in a familial setting as expressions of von Hippel-Lindau (VHL) syndrome.

Gastrina

Gastrinomas are also seen in MEN1. The symptoms and signs reflect two processes, malignancy and gastrin induction of acid secretion by the stomach. Gastrinomas are found in about 40% of adults with MEN1 (see Table 36-1). ZES is defined here as symptoms or signs of gastric acid hypersecretion caused primarily by a gastrin-secreting enteropancreatic neuroendocrine tumor or tumors. Among patients with ZES, MEN1 is found commonly, on the order of 25% in large series. Most of these MEN1 cases are readily recognizable from personal and family history. In contrast, among carefully defined sporadic ZES cases without obvious MEN1, occult MEN1 is uncommon, a conclusion based on lack of family history, long-term follow-up, and mutational analysis of the MEN1 gene.

Symptoms of ZES include diarrhea, esophageal reflux, and those associated with peptic ulceration. Prospective tumor surveillance studies have more clearly defined the range of presenting features. Symptoms can antedate recognition of fasting hypergastrinemia. At the extreme, ulcer perforation can be caused infrequently by hypergastrinemia, even without prior symptoms. The laboratory diagnosis of gastrinoma is established by finding an elevated serum gastrin level. Other causes of elevated gastrin (false-positive results) that must be differentiated from gastrinoma include hypochlorhydria, including that resulting from pharmacologic agents that inhibit peptic acid secretion. Hyperparathyroidism in MEN1 can exacerbate hypergastrinemia (see earlier) (see Fig. 36-4).

Recognition of elevated gastrin should be followed by assessment of the gastric acid secretion rate without acid-blocking drugs; the normal rate is below 15 mEq/hour (or below 5 mEq/hour after acid-reducing surgery). The diagnosis of gastrinoma can also be confirmed by measuring the gastrin response to secretin (2 U/kg). A gastrin increase of more than 114 pmol/L (200 pg/mL) is diagnostic of gastrinoma. This test differentiates gastrinoma from other hypergastrinemic states, such as retained gastric antrum, massive small bowel resection, or gastric outlet obstruction. Secretin was unavailable for several years in the United States but was recently approved for diagnosis of gastrinoma. An alternative gastrin provocation method such as calcium infusion has not been standardized and would be unappealing with coexistent hyperparathyroidism. Gastric endoscopy is advisable at the initial evaluation for ZES and allows assessment of peptic ulcerations and duodenal gastrinomas; it should also be used to search for gastric carcinoids, which are frequent in MEN1 (see later).

Like hyperparathyroidism, gastrinomas have certain features that are relatively specific for MEN1. On the average, gastrinoma begins 10 years earlier with MEN1 than without it, a lesser age difference than 30 years for hyperparathyroidism. The gastrinomas in MEN1 are often small, multiple, and intraduodenal (see Fig. 36-3, bottom). The duodenal predominance differs only modestly from that in sporadic gastrinoma. Because the gastrin cell or D cell is not normally found in the duodenum or pancreas, gastrinoma in these locations may be judged as ectopic and malignant, independent of its histologic grade. Gastrinomas in MEN1 have a high propensity to metastasize to local nodes. High-grade aggressive behavior, including distant spread to the liver and occasionally other tissues, also occurs in about 20% of cases. Diffuse hepatic metastases are particularly ominous, predicting a 5-year survival of only 50%. The prognosis of gastrinoma in MEN1 is similar to that in sporadic cases. No early features have allowed reliable prediction of which gastrinomas behave aggressively.

Therapy of Gastrinoma

Most centers have reported virtually zero success rate for cure of gastrinoma in MEN1 by surgery, even though one third of gastrinoma cases without MEN1 are cured by surgery. Unique characteristics of gastrinoma in MEN1 that contribute to the low rate of curative resection are the multiplicity of small tumors and the frequency of local metastases. The largest tumor was not a gastrinoma in 40% of operations. Extreme approaches including total pancreatectomy have been suggested, but the associated surgical morbidity seems unacceptable. Only one group has reported frequent surgical cure of gastrinoma in MEN1, aided by duodenectomy and endoscopic transillumination of the mucosa. Other groups have not reported similar success rates despite the use of similar approaches. Differences in criteria for cure and in selection of patients, such as age, may contribute.

The development of H2 histamine receptor antagonists (cimetidine and ranitidine) and proton pump inhibitors (omeprazole and pantoprazole) makes it possible to perform a pharmacologic gastrectomy for ZES. The pump inhibitors are even more effective than the H2 receptor antagonists. If compliance is good, the need for total gastrectomy is eliminated. Side effects, including those from achlorhydria, are mild. Gastric carcinoids develop in rats given large doses of acid pump inhibitors. Gastric carcinoids are also seen in MEN1; however, the pump inhibitors do not seem to exacerbate them in MEN1. There remains disagreement about whether the pump inhibitors worsen enterochromaffin-like cell hyperplasia in sporadic ZES. The somatostatin analogue octreotide inhibits the secretion of both gastrin and gastric acid, and it is under evaluation for a role in malignant gastrinoma. In addition, the gastrin-lowering effect of somatostatin analogues may account for their effective suppression of gastric carcinoid mass in MEN1.

Although medical therapy for ZES in MEN1 is preferred and effective, the need for lifetime medical therapy, the recognition that small duodenal gastrinomas cause a high percentage of cases, and the poor outcome (50% 5-year survival) in patients with hepatic metastasis lead to frequent reexamination of treatment choices. The same issues have led to suggestions for earlier and more aggressive surgical intervention, although at present there is no objective evidence that such intervention is more successful.

Insulinoma

Insulinoma is the second most common hormone-secreting enteropancreatic neuroendocrine tumor in MEN1, with an overall frequency of 10% among adults with...
MEN1 (see Table 36-1). By coincidence, MEN1 also accounts for approximately 10% of all cases with insulinoma. The clinical features and diagnostic criteria are the same in MEN1 and sporadic cases gleaned from symptomatic fasting hypoglycemia with high insulin, C peptide, or proinsulin (see Chapter 32). Insulinoma syndrome in MEN1 is usually caused by a single dominant and benign pancreatic islet tumor, although ancillary non-hyposecreting islet tumors that stain for insulin or another gut hormone are common in MEN1. The main insulinoma is generally 2 to 3 cm in diameter and located anywhere in the pancreas. Removal of the main insulinoma is usually curative. Rarely, more than one tumor causes the insulinoma syndrome in MEN1. The postoperative recurrence rate of insulinomas may be higher in MEN1 than in sporadic cases (20% versus 5% at 5 years). As is the case with hyperparathyroidism, recurrent insulinoma in MEN1 may, in theory, have arisen from residual tumor or a new clone.

Preferred treatment is surgical removal of the insulinoma. Other incidental pancreatic islet macroadenomas should also be removed because of the possibility that one may be a hypersecreting insulinoma and also because it might become malignant. Somatostatin receptor scintigraphy (SRS) can give 30% to 60% true positive images. When surgery is done with guidance by intraoperative ultrasonography, the success rate should be satisfactory, although no large series has yet documented this in insulinoma in MEN1. Before intraoperative ultrasonography was available, routine distal pancreatectomy was sometimes recommended. This is now rarely needed but still should be contemplated for prevention of other tumors.

Several techniques, based on insulin radioimmunoassay, can be useful for localization of an insulinoma in patients with multiple islet cell tumors in MEN1. These include infusion of calcium into selectively catheterized pancreatic arteries with measurement of insulin in right or left hepatic venous effluent. Identification of an insulin peak after an intra-artrial calcium infusion localizes the insulinoma to the distribution of the artery. Other tests that have been useful include rapid intraoperative insulin and glucose levels in serum or intraoperative insulin levels in fine-needle aspirates of a pancreatic tumor (S. K. Libutti et al. National Institutes of Health, Bethesda, MD, unpublished).

Metastatic insulinomas causing hypoglycemia should be treated surgically, but operative strategies are less likely to be successful. Hypoglycemia caused by unlocalized or metastatic insulinoma can be controlled with diazoxide and somatostatin analogues are less effective.

**Tumors Secreting Glucagon, Vasoactive Intestinal Peptide, or Other Hormones**

**Glucagonoma Syndrome**

This syndrome consists of hyperglycemia, anorexia, glossitis, anemia, diarrhea, venous thrombosis, and a characteristic skin rash termed necrotic migratory erythema (see Chapter 30, Chapter 32, and Chapter 38). Glucagonoma syndrome is rare in MEN1, although one third of MEN1 enteropancreatic neuroendocrine tumors immunostain for glucagon. Glucagonoma is usually large and metastatic at presentation. Palliation is often possible with surgery or another ablative procedure (see later). Some patients have responded to the somatostatin analogue octreotide, although an initial response has not predicted a long-term response.

**Enteropancreatic neuroendocrine tumors in MEN1 frequently also oversecrete pancreatic polypeptide.**

**Other Endocrine and Enzymatic Hormones**

These are occasionally oversecreted by enteropancreatic neuroendocrine tumors in MEN1. They include adrenocorticotropic hormone (ACTH), PTH-related peptide, somatostatin, and calcitonin. The last can cause confusion with thyroidal C-cell cancer.

**Pancreatic Polypeptide/Secreting and Other "Nonfunctional" Tumors**

One third of enteropancreatic neuroendocrine tumors in MEN1 immunostain mainly for pancreatic polypeptide, similar to the fractions that immunostain mainly for insulin or for glucagon. Enteropancreatic neuroendocrine tumors in MEN1 frequently also oversecrete pancreatic polypeptide. This is not associated with any hormonal syndrome resulting from the peptide. Like other nonfunctional enteropancreatic neuroendocrine tumors in MEN1, these are often large, malignant, and metastatic at presentation.

Nonfunctional tumor is an abused but convenient term. In the context of MEN1, it is applied here to enteropancreatic neuroendocrine tumors, anterior pituitary tumors, or foregut carcinoids that do not immunostain for the common hormones of that tissue or that may immunostain for one or more hormones but do not hypersecrete the hormone and do not cause a hormonal syndrome. Most enteropancreatic tumors in MEN1 fit this definition, and most never become a clinical problem. However, if a nonfunctioning tumor becomes malignant, its lack of symmetric hormone hypersecretion can allow progression to an advanced stage before recognition.

**Staging of Enteropancreatic Neuroendocrine Tumors in Multiple Endocrine Neoplasia Type 1**

Appropriate management of enteropancreatic neuroendocrine tumor or tumors in MEN1 is challenging because of the multicentric nature of the tumors and the need to decide between surgical and other approaches. An enteropancreatic tumor causing a hormone excess state is likely to be accompanied by nonfunctional tumor(s). Management of these tumors is accomplished with imaging of sporadic tumors of the same types cannot be generalized to MEN1 tumors. Accurate localization of tumor and, in particular, identification of metastatic disease are critical for preoperative decision making. The multicentricity and variable size of these tumors stretch the limitations of radiologic techniques that have difficulty imaging tumors smaller than 1 cm in diameter. Their rarity has prevented organization of controlled trials. Despite these challenges, there has been considerable progress over the past decade.

Somatostatin receptor scintigraphy (SRS) is a generally useful method for imaging enteropancreatic neuroendocrine and foregut neuroendocrine tumor. It can image primary tumor and local or distant metastases. It is particularly useful for gastrinomas in MEN1, and it has led to the development of novel therapeutic procedures in MEN1-associated gastrinomas. Although it is the single most powerful imaging test, SRS still fails to image one third of lesions identified at surgery even in sporadic gastrinoma. The yield of SRS with sporadic insulinoma is somewhat lower than with other pancreatic islet tumors, with 30% to 60% true positives. Abdominal imaging by computed assisted tomography (CAT), particularly helical CAT, is a generally useful method for imaging enteropancreatic neuroendocrine and foregut neuroendocrine tumors. It is particularly useful for gastrinomas in MEN1 and it has led to the development of novel therapeutic procedures in MEN1-associated gastrinomas. Although it is the single most powerful imaging test, SRS still fails to image one third of lesions identified at surgery even in sporadic gastrinoma. The yield of SRS with sporadic insulinoma is somewhat lower than with other pancreatic islet tumors, with 30% to 60% true positives. Abdominal imaging by computed assisted tomography (CAT), particularly helical CAT, combined with early phase images after contrast injection or magnetic resonance imaging provides enhanced sensitivity for detection of small lesions and is complementary to SRS.

No imaging technique utilized for evaluation of MEN1 enteropancreatic tumors is completely satisfactory. Endoscopic ultrasonography with or without needle aspiration of a pancreatic mass may have selected applications but will probably be technically demanding, scarce, and expensive option for the foreseeable future.

With the exception of endoscopic ultrasonography, the current preoperative imaging methods are not able to image tumors confined to the pancreas and less than 1.5 cm in diameter. They also fail to identify metastases in 25% of cases and the extent of tumor multiplicity in MEN1 cases. In contrast, intraoperative ultrasonography is a useful tool for localizing small tumors not detectable by the eye or fingers of the surgeon. This technique has become the primary approach for diagnosis of small gastrinomas in most medical centers, although experience has been limited to sporadic tumors.
Functional (i.e., insulin-specific) testing can be useful to assess insulinoma because, unlike other enteropancreatic neuroendocrine tumors in MEN1, insulinoma may be symptomatic when small and solitary (see earlier).

Serum markers in MEN1, mainly chromogranin-A, provide useful diagnostic tools in creating suspicion of the presence of an enteropancreatic tumor. Chromogranin-A has not been helpful in insulinoma, perhaps because of the small tumor mass. An example of the usefulness of hormonal markers is the relationship between glucose and insulin, C peptide, and proinsulin in the diagnosis of quite small insulinomas. In contrast, chromogranin-A and gastrin as possible tumor markers have not been reliable indices of gastrinoma extent or progression.

Treatment of Enteropancreatic Neuroendocrine Tumors in Multiple Endocrine Neoplasia Type 1

Aspects of treatment specific to gastrinomas and insulinoma in MEN1 have already been described. Treatment of these and other enteropancreatic neuroendocrine tumors in MEN1 is controversial and guided in part by staging procedures and local preferences. The main controversies are highlighted.

Is Tumor Size Important?

Metastasis has been associated with gastrinomas more than 3 cm in diameter. This association has led some to recommend resection for all enteropancreatic tumors above a size cutoff of 2.5 to 3 cm. One analysis of this strategy suggested a failure to prevent later emergence of hepatic metastases. Some others have not found a relation of tumor size and metastasis and do not use a size criterion.

Should All Enteropancreatic Neuroendocrine Tumors Be Removed?

There is no consensus on this point. A reality confronted in MEN1 is that for every identifiable pancreatic tumor there are likely to be several smaller unidentified tumors (clones) that coexist or emerge at a later date. Improvements in pancreatic surgical technique, however, have made it possible to excise smaller lesions surgically, although the rationale for doing this is less clear. Certainly there is no compelling evidence to suggest that surgical removal of small tumors, unless they produce a hormonal syndrome, improves overall outcome. Some urge removal of all detectable macroadenomas if removal would not be dangerous. Others urge a large size cutoff (2.5 to 3 cm in diameter) for removal.

Should Metastatic Enteropancreatic Cancer Be "Debunked"?

Total pancreatectomy with a high rate of complications has been used for very large tumors. Many methods are under exploration for resecting or otherwise ablating enteropancreatic neuroendocrine cancer. Results are too preliminary to justify endorsing any of these.

Should Medications Be Used to Control Tumor Progression?

Enteropancreatic neuroendocrine tumors are generally differentiated and quite resistant to chemotherapy. Several regimens have been tried including streptozotocin, doxorubicin, or interferon, but there is no proof of long-term efficacy. Octreotide has been effective in inhibiting hormone secretion by benign and malignant enteropancreatic neuroendocrine tumors; however, it has not been effective by itself in blocking growth of these tumors except for malignant gastrinoma. Its use in multidrug regimens needs further evaluation.
Pituitary Adenoma

Anterior pituitary tumor occurs in about one third of MEN1 cases. The frequency of MEN1 in cases of apparently sporadic pituitary tumor is probably below 5%, although estimates vary widely to as high as 15% with prolactinoma. The overall frequency of hormones hypersecreted is similar to that in non-MEN1 pituitary tumors: oversecretion of prolactin 60%, oversecretion of growth hormone with or without prolactin 15%, nonsecreting 25%, and oversecretion of ACTH 5%. Excessive secretion of thyrotropin or gonadotropins is rare. Pituitary mass effects can be the principal problem. In fact, pituitary tumors in MEN1 have been larger and less responsive to treatment than those without MEN1. Pituitary tumor can occur early in MEN1 and is occasionally the first recognized feature. Rarely, two independent tumors have been suggested. However, the difficulty of recognizing multiple tumors in the anterior pituitary may lead to underestimation of their frequency. Inactivation of both copies of the MEN1 gene in one or a few pituitary cells leads to monoclonal or oligoclonal proliferation.

Prolactinoma

Prolactinoma is the most common pituitary tumor in MEN1 and the third most frequent endocrine tumor in MEN1 after parathyroid tumors and gastrinomas. The general properties are similar to those of sporadic prolactinoma. MEN1 prolactinoma may be large. Dopamine agonists (e.g., cabergoline, bromocriptine, pergolide, quinagolide) are the preferred treatment. A reduction in side effects and greater potency make cabergoline the current treatment of choice and have improved patients’ compliance. In patients who escape from the growth inhibitory effects of these dopamine agonists or who are noncompliant, transphenoidal surgery combined with radiation therapy is usually effective.

Tumors Producing Growth Hormone or Growth Hormone-Releasing Hormone

The clinical features of growth hormone excess are similar in cases with and without MEN1. There are two different etiologic mechanisms with different treatment implications. The majority of MEN1 pituitary adenomas arise clonally from inactivation of both alleles of the MEN1 gene in a tumor precursor cell. Additional genes such as Gsp (another term for the subunit of the stimulatory G protein) may be implicated, in this case by activating mutation of Gsp. The second mechanism of pituitary GH tumorigenesis is overproduction of GHRH by pancreatic islet or carcinoid tumor. The resulting secondary pituitary tumor is a polyclonal or hyperplastic process, which responds poorly to therapy directed only at the pituitary; removal of the primary GHRH-producing tumor is essential. Although acromegaly secondary to GHRH is rare in sporadic or MEN1 cases, a disproportional fraction of such cases have had MEN1. Thus, measurement of serum GHRH in MEN1 acromegalic cases seems worthwhile. GH-producing pituitary tumors also produce GHRH locally, but this has not interfered with the interpretation of serum GHRH levels.

Treatment for acromegaly with MEN1 is the same as without MEN1. Surgery is usually the first choice, but the development of other pharmacologic therapies including long-acting somatostatin receptor antagonists and growth hormone receptor antagonists can provide effective, albeit expensive, control. In patients with large tumors causing mass effects or those in whom GH effects are not controlled by surgery or pharmacologic therapy, radiation using an external beam, gamma knife, or proton beam is an alternative.

Corticotropin Hypersecretion or Primary Adrenocortical Hyperfunction

Cushing’s syndrome in MEN1 can be caused by a pituitary tumor producing corticotropin (ACTH) or uncommonly by ectopic production of ACTH from a carcinoid or an islet tumor, or by ectopic production of ACTH-releasing hormone (CRH). Therapy should be directed initially to treat the ACTH- or CRH-producing primary tumor. When therapy directed toward the primary source is not successful, corticosteroid production can be controlled by bilateral adrenalectomy or medical therapy. One or both adrenal glands are enlarged in up to 40% of MEN1 cases. This enlargement, most commonly discovered during panoramic imaging, is generally clinically silent and rarely requires treatment. The silent enlargement represents a presumably polyclonal or hyperplastic process of unknown etiology, and it rarely behaves as a neoplasm. Rare MEN1 cases have been identified with primary hypercortisolism, hyperaldosteronism, or adrenocortical cancer; these have not been proved to be intrinsic features of MEN1.
Foregut Carcinoid Tumors

Carcinoid tumor is recognized in 5% to 15% of MEN1 cases.\textsuperscript{21} \textsuperscript{154} Although sporadic carcinoid is derived mainly from midgut and hindgut, MEN1 carcinoid is primarily found in derivatives of the foregut (thymus, bronchus, stomach). Certain carcinoid tumors, unlike any other manifestation of MEN1, have a sex-specific distribution. Thymic carcinoid is found mainly in males, and bronchial carcinoid is identified mainly in females.\textsuperscript{1725} The average age of carcinoid recognition in MEN1 is 45 years,\textsuperscript{180} later than that of other MEN1 tumors. This later age may reflect their lack of compression-induced symptoms and the lack of a hormone oversecretion syndrome with most MEN1 carcinoids.

Thymic carcinoid in MEN1 is usually found at an already advanced stage as a large invasive mass. Less commonly it is recognized during chest imaging or during thymectomy adjunctive to parathyroidectomy. Thymic carcinoid is more frequently malignant (about 70%) than bronchial carcinoid (about 20%) in MEN1.\textsuperscript{180} \textsuperscript{181} \textsuperscript{182} \textsuperscript{183} \textsuperscript{184} MEN1 thymic or bronchial carcinoids rarely oversecrete ACTH, calcitonin, or GHRH; similarly, they rarely oversecrete serotonin or histamine and rarely cause the carcinoid syndrome. They may thus be considered as clinically nonfunctioning. Mediastinal or bronchial carcinoids are best imaged by computed tomography; however, SRS is often positive and therefore useful.\textsuperscript{187}

Gastric carcinoid has been recognized more recently and is less well characterized in MEN1. It is a tumor of histamine-secreting enterochromaffin-like cells. Gastric carcinoids do not cause a hormonal syndrome in MEN1. In up to 15% of MEN1 cases, they have been recognized incidentally during endoscopy.\textsuperscript{68} The overall malignancy rate seems low, but there are exceptions.\textsuperscript{188} At early stages they can regress after treatment with somatostatin analogues.\textsuperscript{85}

Carcinoid occurs occasionally in several members of a small family without other manifestations of MEN1; the etiology of these associations is not known.\textsuperscript{189} \textsuperscript{190} \textsuperscript{191}
Miscellaneous Features of Multiple Endocrine Neoplasia Type 1

Miscellaneous Endocrine Tumors in Multiple Endocrine Neoplasia Type 1

Pheochromocytoma

This is a rare feature in MEN1. There have been fewer than 10 reported cases. Most have been unilateral and chemically silent; one was malignant. In two tumors, 11q13 loss of heterozygosity (LOH) was documented, making it likely that all or most of these rare pheochromocytomas are true clonal expressions from biallelic MEN1 gene inactivation. This is supported by more frequent pheochromocytoma in a mouse model of MEN1.

Thyroid Follicular Neoplasms

This has been associated with MEN1 since the earliest reviews. Although the association is probably correct, it is more likely related to the high incidence of thyroid follicular neoplasms in the general population (unrelated to MEN1) that are uncovered during the inevitable neck exploration for parathyroid disease in MEN1 than any causal association. Further support for a coincidental association is the failure to identify MEN1 gene mutations in sporadic thyroid follicular tumors.

Miscellaneous Nonendocrine Tumors

MEN1 has nonendocrine features that vary from rare to common, with some offering possible use in the diagnosis of MEN1.

Lipoma

This has been associated with MEN1 for over 30 years. MEN1 lipomas are generally dermal, small, and sometimes multiple. Their frequency in MEN1 is about 30% versus 5% in control subjects without MEN1. The frequency of lipomas in normal subjects has limited their use for MEN1 carrier ascertainment.

Multiple Facial Angiofibromas

These have been found in 88% of MEN1 cases but not in control subjects. Half of MEN1 cases have five or more. They are acniform papules that do not regress and that may extend across the vermilion border of the lips.

Collagenoma

This was also observed in 72% of MEN1 cases but not in control subjects. Collagenomas are whitish macular lesions about the trunk, sparing the face and neck. The MEN1 lipomas, angiofibromas, and collagenomas show loss of one copy of 11q13. Thus, it is likely that these are clonal neoplasms and caused by inactivation of the (first and then) second copy of the MEN1 gene.

Spinal Cerebellar Ependymoma

This has been seen in four MEN1 cases. There are no studies to determine whether 11q13 LOH or other MEN1 gene abnormalities are causative.

Malignant Melanomas

This has occurred in at least seven MEN1 cases, but direct involvement of the MEN1 gene has not been tested.

Leiomyoma (of Esophagus, Lung, Rectum, or Uterus)

This has been reported in several MEN1 cases. Analyses of 11q13 LOH established that esophageal and uterine leiomyoma are specific to MEN1 cases. Similar MEN1 inactivation was not implicated in sporadic uterine leiomyoma.

Varying Penetrance of Tumors by Tissue or by Age

MEN1 is perhaps the most heterogeneous of all multiple neoplasia syndromes. The many tumors of MEN1 have a wide range of penetrance (see Table 36-1). If the organ is paired and the penetrance is high, the tumors are generally bilateral (i.e., parathyroid); if the tumor is rare in MEN1, its random occurrence is generally unilateral even in a paired organ (i.e., pheochromocytoma). Naturally, the apparent penetrance of any tumor type is heavily dependent upon the scrutiny that the organ is given. Thus, the frequent facial angiofibromas of MEN1 were not recognized until 1997. When symptoms alone are the main basis for disease recognition, the first feature of MEN1 in adolescents is not hyperparathyroidism but rather prolactinoma or insulinoma. For each tumor type, penetrance necessarily increases with age (see Fig. 36-1). Overall, the penetrance for MEN1 (usually parathyroid) reaches nearly 100% by age 50, but occasional obligate MEN1 mutation carriers have not shown any tumor beyond age 70. Earliest penetrance and earliest preventable morbidity must be evaluated in decisions about when to begin carrier ascertainment for tumors in a likely carrier. The earliest ages for identification of specific tumor expression in MEN1 have been as follows: prolactinoma at age 5, insulinoma at age 6, hyperparathyroidism at age 8, and gastrinoma at age 12 (R. Jensen, personal communication). The information about morbidity for most of these cases is incomplete; thus, more information is needed before it is possible to make consensus recommendations regarding the correct age at which to begin tumor surveillance and possibly intervention.

Phenotypes or Varying Tumor Penetrance by Family

Clustering of clinical subvariants of MEN1, similar to that seen for MEN2 (see later in chapter), has been evaluated. Preliminary analyses in small MEN1 families suggested clusters of ACTH-producing pituitary tumors, insulinomas, carcinoid, and aggressive gastrinomas. Identification of a specific mutation that correlates with a specific clinical variant in multiple kindreds would be most meaningful. Although subsequent analysis has failed to identify such a relationship (see
Hyperparathyroidism

Hyperparathyroidism is the most common clinical feature of MEN1 and occurs at a relatively young age. It would therefore not be surprising to identify isolated hyperparathyroidism in small families with early or occult MEN1, particularly those with a disproportionate number of young members. Larger families (four, five, or more affected members) have been identified with familial isolated hyperparathyroidism (FIH) and an identifiable MEN1 mutation but still could represent a random part of the normal spectrum of MEN1 expression. Eventually, most would probably develop other clinical features of MEN1. Two FIH families with MEN1 mutation have been particularly large, with 8 and 13 hyperparathyroid members, raising the likelihood that, in some families, isolated hyperparathyroidism may exist and continue as the only manifestation of MEN1 mutation. MEN1 mutation is rare in families with FIH.

When MEN1 occurs in its typical forms, it is easily diagnosed. Presentation as a single sporadic tumor, as FIH, or as familial isolated pituitary tumor (see earlier) not only is rare but also presents the clinician with a difficult diagnostic challenge.

Sporadic Tumor or Tumors

MEN1 can occur without a recognized or even recognizable family history of MEN1. When sporadic cases present with two or more typical tumors, some cases meet the definition criteria for MEN1 (see later), for others the suspicion of MEN1 is high. The prevalence of MEN1 mutation is 10% to 90%, depending on the specific tumors (see later). When the sporadic case presents in only one tissue, the suspicion and the true frequency of MEN1 mutation are low. The frequency of occult MEN1 with sporadic tumor can be estimated as follows: hyperparathyroidism (2%), gastrinoma (5%), pheochromocytoma (5%), progressing adrenal adenomas (5%). Factors that increase the likelihood of MEN1 in these settings are earlier onset and tumor multiplicity in the same organ.

Familial Isolated Hyperparathyroidism

When hyperparathyroidism is familial and isolated, the main possibilities include occult MEN1 (see earlier), familial hypocalciuric hypercalcemia (FHH), hyperparathyroidism with jaw tumor syndrome (HPT-JT), MEN2A, and so-called true FIH (see Chapter 28). FHH, with a frequency similar to that of MEN1, is an autosomal dominant disorder characterized by lifelong hypercalcemia with normal urine calcium excretion. PTH levels and parathyroid gland mass are normal or minimally increased. After subtotal parathyroidectomy, the residual parathyroid tissue directs persistent hypercalcemia after subtotal parathyroidectomy. The parathyroid dysfunction is not neoplastic but polyclonal. A remarkably high rate of persistence after subtotal parathyroidectomy and a low morbidity without surgery justify efforts to avoid parathyroid surgery in FHH. Useful diagnostic features of FHH are the low ratio of renal calcium clearance to creatinine clearance (in the presence of hypercalcemia) and the onset of hypercalcemia in relatives typically before age 1 year. Two thirds of FHH index cases have an activating mutation of the calcium-sensing receptor gene (CASR). Most of the rest are believed to have an undetected mutation of CASR, suggested by genetic linkage to chromosome 3q; occasional families have the FHH syndrome with mutation in unknown genes at 19p or 19q. One family with a missense mutation of CASR had features intermediate between FHH and typical hyperparathyroidism.

HPT-JT is a syndrome of hyperparathyroidism, jaw tumors, and renal lesions. Transmission is autosomal dominant, and it is caused only by the yet unidentified HRTF2 gene at 1q24–1q32. The commonest and sometimes the only feature is hyperparathyroidism. The hyperparathyroidism typically involves one parathyroid gland at a time and there is a uniquely high malignant potential in the parathyroid tumor; 15% of reported cases have had parathyroid cancer. The associated jaw tumors are osseous or cementifying fibromas. Unlike the jaw tumors of hyperparathyroidism, they are not influenced by the hyperparathyroid status. The associated renal lesions are multiple renal cysts, hamartomas, or Wilms’ tumor. An international consortium is working to identify this gene. Occult MEN2A, theoretically another cause of FIH, has not been identified in the form of FIH.

Many small kindreds with two or three affected members receive a diagnosis of FIH. For years FIH was not pursued as a syndrome because of bland features and the belief that most kindreds had occult MEN1. A detailed analysis of many kindreds with FIH found occult MEN1, FHH, or HPT-JT in the minority. Whether the remaining individuals have true FIH or some other genetic form of hyperparathyroidism will become clear only with long-term follow-up or when robust deoxyribonucleic acid (DNA) tests become available for all of the genes for FHH (at least three genes), HPT-JT, and other hereditary forms of hyperparathyroidism.

Familial Isolated Pituitary Tumor

Familial isolated tumor of the anterior pituitary has been recognized in several small and few large families. The tumors are usually somatotropinomas, occasionally prolactinomas. In theory, familial isolated tumor of the anterior pituitary could be an expression of occult MEN1. To date, however, no family with familial isolated somatotropinoma has had a MEN1 mutation (see later). It is more likely that most of these families harbor mutations of another unknown gene or genes.
The MEN1 Gene: Normal or Mutated

The Normal MEN1 Gene and Normal Menin

Larsson and colleagues showed in 1988 that the MEN1 gene mapped to chromosome 11q13 and that it was probably a tumor suppressor gene (see the following). However, almost a full decade passed before the MEN1 gene was identified by positional cloning. This strategy involved a progressive narrowing of the candidate gene interval, cloning all the DNA in the narrowed interval, and identifying all or most genes therein. The final step required sequence analysis of each of these genes in a panel of DNA from familial MEN1 index cases, a systematic process that led to the identification of the one gene that carried the defining mutations.

The MEN1 gene is 10 kilobases (kb) in size and encodes transcripts of 2.7 and 3.1 kb. The transcripts are expressed in all or most tissues and with little cell cycle dependence. They encode a 610-amino-acid protein termed menin. Rat, mouse, zebra fish, snail, Drosophila, and human menins are highly homologous.

Menin has two nuclear localization signals near the carboxyl terminus that are likely to be responsible for its predominantly nuclear compartmentalization. The first interacting protein partner identified for menin was selectively junD but not other members of the activator protein-1 (AP1) transcription factor family including fos, fra, or other jun proteins. The menin-junD interaction may confer upon junD unique effects by which junD differs from other members of the AP1 transcription family. For example, junD has several actions opposite to those of C-jun, and in the absence of menin binding to it, junD behaves more like C-jun. The importance of the menin-junD interaction for the development of MEN1 is unclear. Homozygous knockout of junD in the mouse resulted in no identifiable abnormality of tissues involved in MEN1. Other studies have identified SMAD3, PEM, NM23, nuclear factor B, and several other proteins that potentially interact with menin. Each interaction has unknown importance.

Tumorigenesis: Sequential Two-Step Inactivation of the MEN1 Gene

The first DNA-based discoveries in MEN1 suggested that the MEN1 gene was a tumor suppressor (see Chapter 35), observations supported by subsequent studies (see later). Complete inactivation of a gene's function, required in addition to the inherited or somatically acquired first hit (inactivating mutation), a second hit at the same genetic loci that finishes the inactivation of both copies of the MEN1 gene. Inactivation of the second allele can be by mutation or another (epigenetic) means such as promoter methylation, though the latter has not been found for MEN1. A two-hit model for tumorigenesis was developed by Alfred Knudson to account for epidemiologic observations in retinoblastoma: that, in comparison with sporadic cases, some hereditary tumors occurred earlier and in multiple sites. This can now be generalized to say that, in a hereditary tumor, the germ line mutation is obligatorily present in every cell. Thus, the earliest step seen in sporadic tumorigenesis is the presence of the second hit, by which the MEN1 gene is bypassed. Multiple independent cells in susceptible organs are thus primed for somatic mutations at the second and still normal copy to cause early and multiple tumors. This model can be extended to stepwise tumorigenesis by an oncogene such as RET (see later).

Somatic Point Mutations (First Hit) of the MEN1 Gene in Sporadic Tumors

MEN1 is one of the most frequently mutated genes in sporadic endocrine tumors. The frequency of MEN1 mutation is 10% to 20% in parathyroid adenomas, 50% to 58% in non-adenomatous parathyroid gland tumors, 10% to 20% in insulinomas, 50% in VIPomas, and 25% to 35% in bronchial carcinoids. Other sporadic endocrine tumors show a lower frequency of MEN1 somatic mutation: 0% to 5% in anterior pituitary tumors, 0% to 1% in thyroid tumors, 0% in benign or malignant adrenocortical neoplasms, 0% in uremic secondary hyperparathyroidism, and 0% in parathyroid cancer. Sporadic nonendocrine tumors have undergone little evaluation; the MEN1 mutation frequency was 2/19 in angiomatoses, 1/6 in lipomas, 0% in lung cancer other than carcinoid, 1% in malignant melanoma, and 0% in leukemia.

The First Step (First Hit) Can Be in Germ Line or in Somatic Tissue

Virtually all germ line or somatic first hits at the MEN1 gene have been small mutations, involving one or several bases. The mutations are broadly distributed across the MEN1 open reading frame, so much so that half of newly ascertained index cases are found to have a "novel" mutation. At the same time, the other half show a recurring mutation. These are equally distributed between cause by common ancestry (founder effect) and cause by a hot spot for new mutation.

Accumulated patterns of germ line and somatic first-hit MEN1 mutations have further supported the two-hit gene inactivation hypothesis for the MEN1 gene. Three fourths of MEN1 first-hit mutations predict premature truncation of the menin protein. Although the biologic functions of menin are not established, such truncation mutations would probably cause menin inactivation or even absence. For example, all truncation-type MEN1 mutations cause loss of the most carboxyl-terminal nuclear localization signal and could thus compromise the nuclear localization of menin. The remainder predict missense or replacement of one to three amino acids. The functional consequences of any one missense mutation are uncertain and even hard to distinguish from a rare benign polymorphism; however, their frequent occurrence specifically in MEN1 establishes that all or most are deleterious mutations. Loss of menin function in the junD-mediated transcription assay occurs even with many missense (amino acid-changing) MEN1 mutations; this result also predicts menin inactivation.

The Second Hit in MEN1 Tumorigenesis

The second hit is usually a large chromosomal or subchromosomal rearrangement (i.e., mutation), causing deletion that includes the remaining normal MEN1 gene. Another mechanism for creating a mutant second copy is deletion of the normal copy and then duplication of the DNA from the mutant chromosome 11, so called gene conversion. In either case, the result is that neither copy of the MEN1 gene remains normal. LOH or loss of alleles at the affected locus is usually inherent in this process and can provide evidence that gene inactivation has occurred in that chromosomal segment. Less common mechanisms for the second hit include small mutations (one to three bases) or promoter methylation. The second hit is delayed after the first hit. It is always in somatic tissue and usually occurs postnatally.

Loss of Heterozygosity about Chromosome 11q13 as a Research Tool

LOH about 11q13 has been used mainly to deduce loss of the normal copy of the MEN1 gene. In MEN1, 11q13 LOH was found for almost 100% of the following tumor types: parathyroid tumor, gastrinoma and other pancreatic islet tumor, gastric carcinoid, anterior pituitary tumor, and mesenchymal tumors (lipoma, angiofibroma, collagenoma, and...
Among sporadic endocrine tumors of the type found in MEN1, some but not all have frequent 11q13 LOH. The frequencies of 11q13 LOH in these tumors have been as follows: sporadic parathyroid 30% to 40%, gastrinoma 3% to 4%, uremic parathyroid 0% to 5%, parathyroid cancer 0%, gastrinoma 25% to 70%, insulinoma 30%, bronchial and other carcinoid 40% to 70%, anterior pituitary 5% to 10%. The subsequent finding of nearly universal 11q13 LOH in MEN1 parathyroids established that clonal growth was predominant, that tumorigenesis was by gene inactivation, and that a MEN1 growth factor for parathyroid glands in MEN1 could not represent the immediate product of the MEN1 gene.

**Germ Line MEN1 Mutations: Multiple Endocrine Neoplasia Type 1 Phenotypes and Phenocopies**

There have been no clear relations of MEN1 genotype with phenotype, unlike the situation in MEN2 (see later). The truncating MEN1 mutations have the same diverse types of tumor expression as the missense mutations, and phenotypic expression does not differ between amino-terminal and carboxyl-terminal mutations. The distribution of somatic mutations about the open reading frame is similar to that of germ line mutations and seems not entirely random. In particular, there appears to be a deficiency of missense mutations near the carboxyl terminus and a cluster of missense mutations between amino acids 100 and 200. Otherwise, there is no clear clustering of missense mutations that could point to a zone of menin protein susceptible to change of function.

The p16INK4a and p27KIP1 genes encode members of the two cyclin-dependent kinase inhibitor families that participate in the cell cycling pathway, which also includes retinoblastoma and cyclin D1. This syndrome resembles to MEN1 and MEN2 raises the possibility that the tumorigenic pathways of MEN1 or MEN2, or both, overlap and interact with the cell cycling pathway.

Among sporadic parathyroid tumors of varying aggressiveness.

**Conclusion**

Inactivation of certain other tumor suppressor genes, alone or in combination, can cause specific endocrine tumors in mice. In particular, mice with homozygous inactivation of both p16INK4a and p27KIP1 develop at least eight types of proliferative tissue, including tumors of parathyroid, pituitary, pancreas islet, and duodenum (as in MEN1); in addition, they develop C-cell cancers and pheochromocytoma (as in MEN2). The knocked out genes encode members of the two cyclin-dependent kinase inhibitor families that participate in the cell cycling pathway, which also includes retinoblastoma and cyclin D1. This syndrome resembles to MEN1 and MEN2 raises the possibility that the tumorigenic pathways of MEN1 or MEN2, or both, overlap and interact with the cell cycling pathway.

**Tumorigenesis: Steps Toward MEN1 Inactivation**

Other unknown genes can be implicated in MEN1 tumor evolution through gene loss of function (tumor suppressor gene) or gene gain of function (oncogene). Inactivation of certain other tumor suppressor genes, alone or in combination, can cause specific endocrine tumors in mice. In particular, mice with homozygous knockout of both p16INK4a and p27KIP1 develop at least eight types of proliferative tissue, including tumors of parathyroid, pituitary, pancreas islet, and duodenum (as in MEN1); in addition, they develop C-cell cancers and pheochromocytoma (as in MEN2). The knocked out genes encode members of the two cyclin-dependent kinase inhibitor families that participate in the cell cycling pathway, which also includes retinoblastoma and cyclin D1. This syndrome resembles to MEN1 and MEN2 raises the possibility that the tumorigenic pathways of MEN1 or MEN2, or both, overlap and interact with the cell cycling pathway.

The earliest tissue-level effects toward tumorigenesis in MEN1 are not well defined. Although a widespread role for hyperplasia prior to neoplasia has been seen in MEN2, hyperplasia has been subtle or absent in MEN1 tissues. However, the mouse Men1 knockout model for MEN1 has striking giant hyperplastic islets as a precursor for insulinoma, suggesting that more subtle hyperplasia may have gone unrecognized in human MEN1. If there is a role for hyperplasia, it would still be uncertain whether this is an expression of inactivation of one MEN1 allele or further processes.

Genome instability has been suggested in studies of MEN1 lymphocytes and fibroblasts. MEN1 leukocytes show a subtle deficiency in repair of DNA damage. Clonal cell proliferation has been identified in mesenchymal perivascular tissues about MEN1 angiofibromas; this could represent a precursor stage of that tumor. The MEN1 plasma contains a growth factor that promotes mitogenesis in normal parathyroid cells. Considering the overriding roles of MEN1 gene inactivation and of clonal growth, it is not certain whether the growth factor is a contributor to or a consequence of oncogenesis in MEN1.
Testing for Carrier State and for Tumor Emergence in Multiple Endocrine Neoplasia Type 1

Screening and Counseling for Multiple Endocrine Neoplasia Type 1

A screening-program for MEN1 patients should routinely meet three main objectives: identify MEN1 carriers, identify MEN1 tumors particularly at a treatable stage, and be cost-effective (Fig. 36-9). Screening tests can be classified in two categories: ascertainment of carriers of MEN1 predisposition and periodic surveillance for MEN1 tumors (see Fig. 36-9). *

*The term screening has been applied to several processes in the setting of MEN. Herein, a distinction is made between testing for carrier ascertainment and testing for periodic surveillance of tumors. Note that when carrier testing with DNA (mutation or haplotype test) is not possible, streamlined and periodic tumor surveillance becomes the preferred method for carrier ascertainment.

Encounters for carrier ascertainment often involve counseling patients. In addition to standard genetics topics, counseling in MEN1 addresses two different faces of MEN1: an endocrinopathy with good but complex management options and a cancer syndrome with limited management options. A MEN1 information Web page can help in orientation: (http://www.niddk.nih.gov/health/endocr/pubs/men1/men1.htm). Experience with tumor surveillance in MEN1 families has shown that compliance with a simple and regular surveillance protocol is high (Table 36-2); complicated, expensive, and erratic efforts are associated with lower compliance.

Benefits and Limitations of Carrier Ascertainment

The benefits of this type of analysis are several. First is the secure proof of the MEN1 carrier state in an individual with a mutation and, equally important, the potential to exclude the

<table>
<thead>
<tr>
<th>Tumor</th>
<th>Age to Begin Testing (yr)</th>
<th>Biochemical Tests Annually</th>
<th>Imaging Tests Every 35 yr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parathyroid adenoma</td>
<td>8</td>
<td>Calcium (especially Ca++)</td>
<td>None</td>
</tr>
<tr>
<td>Gastrinoma</td>
<td>20</td>
<td>Gastrin, gastric acid output; secretin-stimulated gastrin</td>
<td>None</td>
</tr>
<tr>
<td>Insulinoma</td>
<td>5</td>
<td>Fasting glucose</td>
<td>None</td>
</tr>
<tr>
<td>Other enteropancreatic</td>
<td>20</td>
<td>Chromogranin-A</td>
<td>In-DTPA octreotide ; CAT or MRI</td>
</tr>
<tr>
<td>Anterior pituitary</td>
<td>5</td>
<td>Prolactin, IGF-I</td>
<td>MRI</td>
</tr>
<tr>
<td>Foregut carcinoid</td>
<td>20</td>
<td>Chromogranin-A</td>
<td>CAT</td>
</tr>
</tbody>
</table>

CT, computed tomography; DTPA, diethylenetriaminepentaaetic acid; ECL, enterochromaffin-like; IGF-I, insulin-like growth factor I; MRI, magnetic resonance imaging; PTH, parathyroid hormone.


MEN1 carrier state by a normal sequence analysis when an affected member of the kindred has an identified mutant MEN1 gene sequence. This type of information can assist in decisions about family planning, future medical needs, and so forth. Second, the information from an index case, if shared, can be helpful to relatives unaware of their status. In particular, after a germ line MEN1 mutation is first identified, the information may be shared with a laboratory and with relatives, and DNA-based information may be used to develop an accurate shortcut test for that mutation in relatives (see later). Third, the information is useful to the physician. It assists in further plans about counseling and tumor surveillance. Occasionally, it is important in a decision about surgery, as in a case of apparently sporadic gastrinoma.

What the MEN1 carrier ascertainment test does not do is also important. MEN1 carrier ascertainment, unlike similar testing for MEN2, does not routinely lead to a recommendation for medical or surgical intervention. MEN1 cancers, in contrast to MTC in the thyroid, arise in tissues that cannot easily be ablated. This lack of mutation-guided intervention makes mutation testing at early ages in MEN1 less urgent than in MEN2. One approach is to recommend DNA testing in children of gene carriers at age 5 years, the youngest age at which a morbid and possibly treatable MEN1 treated tumor (prolactinoma) has been identified. An alternative, based on the fact that MEN1 morbidity is rare before the age of 20 years, is to delay carrier ascertainment until the child can make a mature decision about a test that could affect availability of insurance or job opportunities.

No identifiable MEN1 mutation is found in up to 40% of typical MEN1 kindreds, although most are believed to harbor mutations not detected by the most common DNA sequencing strategies. Carrier ascertainment in such kindreds can be established by 11q13 haplotype analysis or by streamlined tumor surveillance (see later) (Fig. 36-9). The potential for benefit from the MEN1 mutation test is generally proportional to the likelihood of finding a MEN1 mutation. Thus, obvious benefit is possible for an index case with familial or sporadic MEN1 or with a state that resembles MEN1 but does not quite meet the usual definitions. For example, a MEN1 germ line mutation was found in each of four cases with sporadic hyperparathyroidism and carcinoid tumor. The value is less clear in an apparently unaffected adult relative of a MEN1.
Germ line DNA mutation analysis can identify or rule out most MEN1 mutation carriers by a test applied only one time during the life span. This test is available through several commercial and academic laboratories (a listing can be found at [http://www.genetests.org](http://www.genetests.org)). The usual tissue surrogate for germ line DNA is blood leukocytes; MEN1-associated tumor is less satisfactory because any identified mutation could have occurred somatically. The MEN1 mutation test is based on polymerase chain reaction amplification of the nine translated exons (the open reading frame) and the intron-exon boundaries. Laboratories use modestly differing protocols.

MEN1 germ line mutations have been detected in 60% to 100% of well-defined MEN1 families. The wide variability in mutation detection is explained partly by family selection but more likely by differences in laboratory detection methods. Genetic linkage analysis previously mapped all well-characterized MEN1 kindreds to 11q13, making it likely that mutation of the MEN1 gene causes almost all familial MEN1. Only two atypical MEN1 families have not been linked to 11q13. Failure to identify a MEN1 mutation could be explained by the presence of mutations involving 3' or 3' untranslated or central intronic sequences, regions that are not normally examined, or by a large MEN1 deletion that results in no abnormal polymerase chain reaction product.

The MEN1 mutation detection rate has been lower (10% to 80%) in sporadic than in familial MEN1, probably because of differences in selection of patients. The MEN1 germ line mutation rate has been high (about 75%) in sporadic cases with hyperparathyroidism and ZES but far lower (about 10%) in sporadic cases with hyperparathyroidism and acromegaly. For the same reason, the MEN1 mutation detection rate has also been lower (0% to 30%) in sporadic cases with atypical MEN1, a truly broad category without a consensus definition. Most MEN1 mutations are familial, but about 10% arise de novo. This cannot be deduced reliably by history; rather, it requires analysis of DNA from both parents, including haplotyping (i.e., genetic "fingerprinting") to verify parental assignment. This is rarely pursued for ethical or legal objections.

When MEN1 mutation cannot be detected in the germ line DNA of a MEN1 index case, ascertainment of the carrier state in relatives is more difficult. Carrier ascertainment can still be based on streamlined periodic tumor surveillance in a relative (see later) or on haplotype analysis (similar to genetic linkage analysis) in a kindred (see Fig. 36-9). Haplotype or linkage analysis for the MEN1 trait can be done with high degrees of confidence; however, few laboratories are doing these analyses, and a substantial number (two or more depending on the informativeness of the probes used and on the family structure) of additional definitely affected relatives must participate.

Multiple Endocrine Neoplasia Type 1 Carrier Ascertainment by Streamlined Surveillance for Tumors: An Alternative to DNA Testing

In kindreds with no identifiable MEN1 mutation or possibility of 11q13 haplotype analysis, it is necessary to base assignment of carrier status on the clinical identification of one of the major tumors of MEN1. Streamlined periodic surveillance for tumors by biochemical tests should be offered every 3 to 5 years. Hyperparathyroidism is the most frequent and usually the earliest manifestation of MEN1, and therefore its recognition is central to this carrier ascertainment strategy. The preferred parathyroid tumor surveillance test is the ionized calcium test, beginning at the age of 8 or later; if it is unavailable, an albumin-adjusted calcium test is suitable and is preferable to the measurement of total serum calcium. A PTH assay should be performed at the same time.

Five years is a suggested starting age for prolactinoma surveillance, based on the occurrence of a macroadenoma in a child of that age with MEN1. Because serum prolactin rises with stress, avoidance of phlebotomy stress in a child may require an indwelling venous catheter and three blood samples at 20-minute intervals. Gastrinoma surveillance can be introduced during adulthood because of the generally later age of onset of ZES in MEN1. Only rarely is gastrinoma the first clinical tumor to occur in MEN1. Surveillance for cutaneous manifestations of MEN1, collagenomas or facial angiofibromas, may be promising but has not yet been explored in children.

False-positive test results are commonly found in MEN1 tumor surveillance through the assays of prolactin (stress, pregnancy, or psychotropic medications) or gastrin (mainly hypochlorhydria including that resulting from inhibitors of gastric acid secretion). Occasionally, a sporadic but common endocrine tumor (such as parathyroid adenoma or pituitary tumor) occurs in a family member who is not an MEN1 carrier.

Periodic Surveillance for Tumors after Proving the Multiple Endocrine Neoplasia Type 1 Carrier State

When the MEN1 carrier status has been identified by any method, it is appropriate to focus continued and increased attention on the patient with the goal of identifying and treating neoplastic manifestations at an appropriate stage. Surveillance for parathyroid tumor, prolactinoma, and insulinoma can begin at age 5 to 8 years; surveillance for gastrinoma, other islet tumors, and foregut carcinoids should be delayed until after the age of 20 years. Cost-effective surveillance combines a carefully obtained history focused on clinical symptoms associated with these tumors, limited hormonal and serum chemistry analysis, and carefully defined (i.e., selective and less frequent) use of imaging.

Some have recommended more extensive surveillance measures that include measurement of pancreatic polypeptide, insulin, proinsulin, or cortisol. Furthermore, a mealstimulated test was developed in the hope of increasing the MEN1-related diagnostic information from pancreatic polypeptide and other markers. Although a case can be made that these tests may result in earlier tumor recognition, it is unclear whether such detection results in benefit to the patient. Any abnormal tumor surveillance test should initiate as appropriate, a sequence of confirmation, exclusion of falsepositives, a decision whether to proceed with staging or surgery or both, and finally preoperative and intraoperative tumor imaging tests. The specific details differ for each tumor type.

Figure 36-10 Bilateral medullary thyroid carcinoma in multiple endocrine neoplasia type 2A. Large bilateral foci of medullary thyroid carcinoma are located in each lobe of the thyroid gland.
MULTIPLE ENDOCRINE NEOPLASIA TYPE 2

Multiple Endocrine Neoplasia Type 2A

In 1959, John Sipple was asked to see a hypertensive patient who subsequently died. At autopsy Sipple "was amazed when [he] saw large, bilateral pheochromocytomas and a 2-cm pale tan mass in each lobe of the thyroid gland and nodular enlargement of the only parathyroid gland [they] could find." He reported this case and reviewed five others from the literature. Subsequently, the familial nature of the syndrome and the recognition of the thyroid tumor as medullary thyroid carcinoma were clarified by others. Williams reasoned that because MTC was a malignancy of the C cells it might produce calcitonin, a concept that led to the use of serum calcitonin measurements for early diagnosis of MTC and of MEN2.

The clinical syndrome of MEN2A, as described by Sipple and others, consists of bilateral and multicentric MTC, unilateral or bilateral pheochromocytoma, and, less commonly, parathyroid hyperplasia or adenoma. In the decade after Sipple’s description, patients with this syndrome frequently presented with manifestations of a pheochromocytoma, a thyroid nodule, hypercalcemia, or some combination of the three. Such clinical presentations are still observed, but MEN2 syndrome identification and routine carrier ascertainment in affected families now make early thyroid C-cell hyperplasia or microscopic MTC without metastasis the most common initial presentation.}

A feature of MEN2 that differs from MEN1 is clear progression of histologic changes from normal to hyperplasia to adenoma (pheochromocytoma or parathyroid adenoma) or carcinoma (MTC). The development of hyperplasia is probably a multicentric process, with each focus of tumor derived from a single clone. This point has been proved for only one manifestation of the syndrome (MTC) but is likely to be true for other tumors as well. The heterozygous activating mutations of the RET proto-oncogene may be the stimulus for hyperplasia, whereas additional mutational events appear to be required for progression of this process (to be discussed later).

Medullary Thyroid Carcinoma in Multiple Endocrine Neoplasia Type 2

Evolution of C-Cell Abnormalities

MTC in all variants of MEN2 is a multicentric neoplasm of the parafollicular or C cell of the thyroid gland. The earliest demonstrable abnormality in the thyroid gland of individuals with this syndrome is hyperplasia of C cells, followed by progression to nodular hyperplasia, microscopic MTC, and finally frank MTC. These changes are multicentric, with the frequent occurrence of more than one type of histologic lesion in one or both lobes of the thyroid. The time required for progression through these several histologic stages is not known, but such changes have been noted as early as 3 years of age in MEN2A and during the first month of life in MEN2B. It is also not known at which earliest histologic stage metastasis occurs, but local lymph node metastasis is common when the tumor diameter is larger than 1 cm, whereas lymph node metastasis is rare in a case with only C-cell hyperplasia. Occasionally, foci of MTC occur in extrathyroidal locations such as the thymus gland. Whether these lesions are primary or metastatic has not been determined with certainty.

Tumor Markers Associated with Medullary Thyroid Carcinoma

Proteins expressed by the normal C cell and by MTC include calcitonin, calcitonin generelated peptide, somatostatin, dihydroxyphenylalanine decarboxylase, and chromogranin-A. Proteins that are not normally expressed in the C cell but that are expressed by MTC include pro-opiomelanocortin, thyrotropin-releasing hormone, gastrin-releasing peptide, VIP, neurotensin, substance P, carcinoembryonic antigen, histaminase, and others. The only reported clinical syndrome associated with ectopic hormone production is the ectopic ACTH syndrome in fewer than 5% of patients with extensive MTC. Studies to exclude hyperparathyroidism and pheochromocytoma are mandatory; pheochromocytomas should be removed before thyroid surgery.

Younger individuals diagnosed by RET mutation analysis as described subsequently should have a total thyroidectomy, C-cell hyperplasia and microscopic MTC are the most common histologic findings in these patients, although minimal or no abnormalities of C cells have been identified in some younger gene carriers. A case can be made for a central node dissection even in early disease because of the finding of metastasis in young children, although a more extensive lymph node dissection is generally not recommended. Children with MEN2B may have earlier metastasis, and consideration should be given to more extensive lymph node dissection. Surgical management for children identified by DNA testing is discussed in a subsequent section.

Monitoring after Surgery for Medullary Thyroid Carcinoma

At 3 to 6 months following thyroidectomy, patients should be reevaluated by measurement of serum calcitonin and an ultrasound examination of the neck. Earlier measurement of serum calcitonin is discouraged because serum calcitonin values may remain elevated for 3 to 6 months after thyroid surgery and become normal at a later time. The elevated calcitonin values are presumed to be related to the generalized rise in serum calcitonin that occurs during inflammation or sepsis. Equally important, the fact that calcitonin gene expression can be activated in inflamed or infected tissue unrelated to the C cell suggests that one should be careful about overinterpreting the significance of a minimal or transient rise in the serum calcitonin value. Stimuli as nonspecific as exercise can cause a serum calcitonin rise.
It may be useful to measure calcitonin after calcium or pentagastrin stimulation at 6 to 12 months if basal serum calcitonin values are undetectable. In general, it is useful to perform this procedure only if it would direct a specific clinical action (discussed in the next paragraph). Serial measurement of the serum calcitonin or carciinoembryonic antigen is a useful indicator of long-term disease progression and helps define the aggressiveness of MTC in a patient with metastatic disease. This type of information may be useful in making decisions regarding intervention with chemotherapy or reoperation. Calcitonin is a secretary peptide, and there is considerable variability of the serum concentration. Serum calcitonin values may vary by as much as 50% or more from measurement to measurement without evidence of progression of disease. However, a plot of values over months or years is useful to establish a trend of disease activity.

Management of Locally Metastatic Medullary Thyroid Carcinoma

Reoperation for persistent MTC was attempted with initially poor results. However, improvement of surgical techniques led to the normalization of serum calcitonin values in approximately one third of reoperated patients. This experience has led others to examine the usefulness of reoperation, and a larger experience suggests that approximately 15% to 20% of carefully selected patients (patients with no evidence of pulmonary, hepatic, or bone metastasis) have normal or nondetectable calcitonin values after reoperation.

A major question confronting the physician contemplating reoperation is whether the tumor is located in the neck or whether distant metastases are present. Techniques that have been used with variable success for localizing the tumor include scanning with thallium 201 and metaiodobenzylguanidine, octreotide scanning, and venous catheterization for measurement of calcitonin in blood from selectively catheterized veins. Several lines of evidence suggest that nondetectable basal and calcium- or pentagastrin-stimulated calcitonin after surgery is likely to be indicative of a cure. These include promising results after the short-term follow-up (5 to 10 years) of reoperated patients and the generally favorable long-term outcome in patients with MTC and local nodal metastasis who had nondetectable calcitonin values after primary surgery. Whether a 20% cure rate justifies the extensive repeat operative procedure is unclear, but the generally poor outcomes of reoperative strategies suggested by earlier reports should be reconsidered in light of this newer experience. One concern related to reoperation is a higher incidence of hypoparathyroidism.

Pheochromocytoma in Multiple Endocrine Neoplasia Type 2

Evolution of Pheochromocytoma

Adrenal chromaffin tissue in patients with MEN2A undergoes the same type of histologic progression as observed for the C cell, including hyperplasia, diffuse expansion of the adrenal medulla, and pheochromocytoma. The usual finding is single or multiple pheochromocytomas, with a background of hyperplastic chromaffin tissue (Fig. 36-12). The pheochromocytomas may be unilateral or bilateral. If a tumor is present in one adrenal gland, hyperplastic changes are likely in the contralateral gland. Invasion of the adrenal capsule by chromaffin cells is observed, but it is rare for these tumors to metastasize. Pheochromocytomas in MEN2B do not differ substantially from those identified in MEN2A.

Most of the pheochromocytomas in MEN2 are intra-adrenal. There are rare examples of pheochromocytomas that develop in adrenal rest tissue, including one with multiple adrenal rests. Other series have described extra-adrenal pheochromocytomas, although it is difficult to determine whether these pheochromocytomas occurred in adrenal rest tissue or along the sympathetic chain. Recurrences are observed in a small percentage of patients, predominantly in the surgical bed. It is unclear whether these recurrences represent additional pheochromocytomas that have developed in adrenal rest tissue or result from residual tissue or seeding. In most cases, they can be easily resected during a second procedure.

Malignant pheochromocytoma occurs rarely in MEN2, and most of the described cases occurred in older series and were associated with larger pheochromocytomas. Reports of malignant pheochromocytoma in MEN2A have become uncommon, suggesting that resection of smaller pheochromocytomas may eliminate not only cardiovascular risk but also the potential for malignant transformation. Malignant pheochromocytoma has been observed in the context of MEN2B. The identification of malignant pheochromocytoma in some families has led to the performance of bilateral prophylactic adrenalectomy in some kindreds, although this approach should be considered only if there is a proven pattern of adrenal medullary malignancy within a family.

The clinical syndrome caused by adrenomedullary disease has changed over the past two decades. Before prospective tumor surveillance, patients frequently presented with hypertension, headaches, cardiac arrhythmias, and large pheochromocytomas. Death caused by a cardiac arrest or stroke was as likely as death from metastatic MTC. Routine surveillance and detection of pheochromocytomas coupled with -adrenergic and -adrenergic antagonist use and improved surgical management have resulted in improved outcomes. Early adrenomedullary abnormalities may cause intermittent headaches, palpitations, and nervousness; hypertension is uncommon. Death caused by pheochromocytoma is uncommon and the reported deaths over the past decade have occurred largely in patients in whom prospective surveillance was not performed because of either an unrecognized gene carrier or noncompliance with routine surveillance. Presumably, some combination of earlier detection and routine use of - and -adrenergic antagonists has affected this change.

Clinicians should be particularly vigilant with women of childbearing age or during pregnancy because of deaths during labor and delivery. If a pheochromocytoma is identified during pregnancy, the conventional practice is to resect the tumor during pregnancy (under the coverage of adrenergic blockade). However, through an unusual set of circumstances, one of the authors (RFG) has managed a patient through a successful pregnancy and delivery. She received adrenergic antagonists throughout her pregnancy and underwent postpartum surgery for the pheochromocytoma.

Adrenomedullary abnormalities associated with MEN2A produce distinctive biochemical features. Increased urinary excretion of epinephrine and elevations of serum metanephrine are the most sensitive indicators of abnormality. Later in the course of the disease or with larger pheochromocytomas, the 24-hour excretion of epinephrine, norepinephrine, metanephrine, and normetanephrine metabolites is usually increased. Urinary vanillylmandelic acid excretion is usually normal early in the course of disease and is not useful for prospective tumor surveillance. Provocative testing (with glucagon or histamine) is rarely required and carries some risk. The diagnosis of pheochromocytoma is confirmed by CT or magnetic resonance scanning of the abdomen in the context of abnormal catecholamines. In most cases, computed tomography scanning provides greater anatomic resolution, is less expensive, and is adequate for cases in which catecholamines are abnormal. Magnetic resonance scanning provides greater specificity (bright image on T2-weighted images) and may be useful for differentiating between a small pheochromocytoma and an adrenal cortical adenoma. Scanning with metaiodobenzylguanidine, a catecholamine analogue that is selectively concentrated in adrenal chromaffin tissue, is useful for confirming the presence of functioning intra-adrenal chromaffin.
tissue and excluding the rare extra-adrenal pheochromocytoma. It is generally not useful for distinguishing between adrenal medullary hyperplasia and pheochromocytoma because it can be positive in either. Octreotide scanning, although useful for identification of extra-adrenal sporadic pheochromocytomas, has little usefulness in the management of pheochromocytomas associated with MEN2. It is rarely necessary to perform adrenal angiography; if it is performed, the patient should receive - and -adrenergic antagonists during and preferably before the procedure. (see Chapter 15).

Therapy of Pheochromocytoma

There has been an evolution of thought regarding management of pheochromocytoma in MEN2A or MEN2B, influenced by several factors. The first is the recognition that death from pheochromocytoma in this syndrome is now rare. Combined early detection and early therapy with - and -adrenergic antagonists have lessened the risk of cardiovascular death. Second, death caused by adrenal insufficiency in patients who have undergone bilateral adrenalectomy may in fact now be more common in MEN2 than death from pheochromocytoma. Third, there have been dramatic improvements in imaging and laparoscopic surgical technology over the past decade that have rendered many of earlier management discussions regarding unilateral versus bilateral adrenalectomy moot, as discussed in subsequent sections.

Radiographic evaluation is generally considered equivalent or superior to direct visualization of the adrenal gland by the surgeon. Laparoscopic adrenalectomy, particularly for small pheochromocytomas, has largely replaced other operative techniques. Current management approaches focus on the identification and treatment of adrenal medullary abnormalities before they become life-threatening. The following discussion focuses on the range of approaches for specific clinical presentations.

Multiple Endocrine Neoplasia Type 2A Family Member with Symptoms of a Pheochromocytoma but No Catecholamine Abnormality

Occasionally, a patient with known MEN2A or 2B presents with symptoms suggestive of a pheochromocytoma and no identifiable catecholamine or radiographic abnormality. In these cases clinical judgment is required to separate anxiety from intermittent, abnormal secretion of catecholamines by a hyperplastic adrenal gland. If the symptoms are persistent, not improved by antianxiety agents, and become disabling or alarming to the patient, pressure to take some action builds. In this situation, a trial of -adrenergic antagonists may, over time, make it possible to separate anxiety from symptoms caused by intermittent catecholamine production. Such an approach, if effective, may also provide a therapeutic option, thereby delaying a surgical procedure until there is a clearly defined radiographic abnormality.

Although there is no literature regarding the usefulness of metaiodobenzylguanidine or octreotide scanning in this situation, if a unilateral abnormality were defined, such information might be helpful. A decision to proceed with a blind exploration (with no imageable abnormality) should be made on a case-by-case basis and is generally discouraged. Direct visualization of the adrenal glands through an abdominal approach rarely provides greater information than imaging studies, and it is preferable to wait until a radiographic abnormality is detected.

Unilateral Pheochromocytoma

Unilateral pheochromocytomas may be identified by the development of adrenergic symptoms, abnormal urine or plasma catecholamines, or by an imaging procedure. The unilateral adrenal nodule in MEN2A or 2B is most commonly a pheochromocytoma, although metastatic MTC or benign adrenal cortical nodules have occasionally been identified. It is difficult and time consuming to differentiate between pheochromocytoma, metastatic MTC, or a benign cortical nodule in this situation. If catecholamines or metanephrines measurements are abnormal, it is appropriate to perform adrenal surgery. Computed tomography-guided fine-needle aspiration or selective adrenal vein sampling is generally discouraged.

Two different approaches have been employed to treat unilateral pheochromocytomas. The first is unilateral laparoscopic adrenalectomy. Reports of several deaths caused by adrenocortical insufficiency over the past decade raise the distinct possibility that, at present, cortisol deficiency is a greater threat to life in MEN2 than a pheochromocytoma and, therefore, the contralateral normal adrenal should remain untouched if radiographically normal. A second approach, developed to maintain adrenocortical function, is cortical sparing adrenalectomy. This is an old technique that has gained increasing favor as the risks of death from adrenocortical insufficiency associated with bilateral adrenalectomy have become apparent. This approach has been applied to small groups of patients with MEN2-related pheochromocytomas with retention of adrenal function in approximately 80% of treated patients.

Resection of an intra-adrenal pheochromocytoma with retention of cortical tissue raises the inevitable possibility of late recurrence of pheochromocytoma caused by adrenal chromaffin tissue at the corticoomedullary interface, which develops into a tumor at a later time in approximately 20% of cases. If retention of adrenocortical function is considered essential (e.g., for employment purposes), consideration should be given to performance of a cortical sparing adrenalectomy on the first identified pheochromocytoma, thereby providing an opportunity for a successful procedure on the contralateral adrenal gland if the first procedure fails. It is also difficult to know whether the first procedure is successful without detailed studies of corticoid secretory function that involve venous catheterization with selective sampling. Cortical sparing adrenalectomy is generally performed through a flank incision; although it is technically feasible to perform this procedure by a laparoscopic approach, a successful example after this approach has not been described.
Bilateral pheochromocytomas eventually develop in approximately one half of MEN2A or MEN2B gene carriers, although there has been a change in the clinical presentation. Twenty-five years ago, pheochromocytomas were generally detected at a much later stage in the clinical course of the disease and, therefore, a higher percentage of patients had bilateral pheochromocytomas at presentation. Measurement of catecholamines and metanephrines, now routinely employed, has led to routine identification of pheochromocytoma at an earlier stage, yielding a higher percentage of unilateral pheochromocytomas at initial presentation. Of those who initially present with a unilateral pheochromocytoma, 50% have a pheochromocytoma develop in the contralateral adrenal gland over a period of 8 to 10 years. Thus, the mode of clinical evolution has changed, but eventually approximately one half of MEN2A or 2B patients experience pheochromocytoma.

The management for patients with bilateral pheochromocytomas follows a paradigm similar to that described for a unilateral pheochromocytoma. Because bilateral tumors are more likely to produce catecholaminergic signs and symptoms, adrenergic antagonists and inhibition of catecholamine synthesis with -methyltyrosine become doubly important (see Chapter 15). Bilateral laparoscopic adrenalectomy or cortical sparing adrenalectomy should be performed. In patients with very large pheochromocytomas, laparoscopic adrenalectomy may not be possible, necessitating bilateral flank approaches or the less preferred anterior abdominal approach.

Hyperparathyroidism

Reports from the 1960s and early 1970s described hyperparathyroidism in 10% to 35% of individuals with MEN2A. These reports described the presence of either parathyroid hyperplasia or multiple parathyroid adenomas in association with hypercalcemia, urolithiasis, or osteitis fibrosa cystica. A careful review of the histology of these tumors has demonstrated occasional adenomatous formation with a background of parathyroid hyperplasia, a finding that is analogous to that observed for C-Cell hyperplasia in the thyroid gland and chromaffin cell hyperplasia in the adrenal medulla.

Simultaneous hyperparathyroidism is almost never seen in patients thyroidectomized for early C-cell abnormalities, although histologic findings consistent with parathyroid hyperplasia have been observed. Partly for this reason most surgeons attempt to preserve parathyroid tissue, particularly in children who have a prophylactic total thyroidectomy. Whether patients who have had a prophylactic thyroidectomy for MTC eventually experience hyperparathyroidism is unknown. Surgical management of hyperparathyroidism is similar to that described for MEN1, although recurrent hyperparathyroidism is a less common occurrence in MEN2.

One concern related to thyroid surgery is the potential for development of hypoparathyroidism after aggressive thyroidectomy. A careful review of most series documents an incidence of hypoparathyroidism that is comparable to or higher than that for other thyroid surgical indications. To address this issue, some surgeons routinely perform a total thyroidectomy and removal and reimplantation of parathyroid tissue into the nondominant arm.

A number of MEN2A variants have been described (Table 36-3). The most common is familial medullary thyroid carcinoma (FMTC) without pheochromocytoma or parathyroid disease. Kindreds with this syndrome account for approximately 10% to 15% of all those with hereditary MTC. Familial MTC is most likely to be confused with sporadic MTC because of the absence of dramatic symptomatology associated with pheochromocytomas that is found in MEN2A.

TABLE 36-3 – Multiple Endocrine Neoplasia Type 2

| Multiple endocrine neoplasia type 2A (MEN2A) | Medullary thyroid carcinoma (100%) |
| Parathormone neoplasia (10%) |
| Variants of MEN2A |
| MEN2A with cutaneous lichen amyloidosis (MEN2A/CLA) |
| MEN2A or FMTC with Hirschsprung’s disease |
| Familial medullary thyroid carcinoma (FMC) |
| Medullary thyroid carcinoma (100%) |
| Parathormone neoplasia (50%) |
| Absence of parathyroid disease |
| Marfanoid habitus (>95%) |
| Intestinal ganglioneuromatosis and mucosal neuromas (>98%) |

is most likely to be confused with sporadic MTC because of the absence of dramatic symptomatology associated with pheochromocytomas that is found in MEN2A. Because the penetrance of pheochromocytoma in MEN2A is much lower than that for MTC, it is possible to designate small kindreds with MEN2A incorrectly as having FMTC. The concern, of course, is that there will be a failure to screen for and diagnose pheochromocytomas in such kindreds.

At least 15 families have been identified with the MEN2A/cutaneous lichen amyloidosis variant. In these families, affected individuals had a pruritic skin lesion over the scapular region of the upper back consisting of multiple infiltrated papules overlying a well-demarcated plaque. The histology is that of cutaneous lichen amyloidosis (deposition of amyloid at the juncture of the epidermis and dermis) in patients with the fully formed skin lesion. Immunohistochemical staining of the amyloid for keratin but not calcitonin was observed, which indicates that the amyloid is probably of dermal origin and not the result of deposition of calcitonin gene products from the thyroid carcinoma. In most patients, intense pruriitus precedes the development of the skin lesion by 3 to 5 years, suggesting that the primary defect may be a sensory abnormality in the C6-T6 dermatomes leading to chronic irritation and friction amyloidosis. In support of this hypothesis is the fact that there is RET expression in the normal dorsal root ganglion, the site at which the sensory nerve cells for this region are located. Neurologic and electromyographic abnormalities have been observed in some of these patients.

A third variant is MEN2A associated with Hirschsprung disease, which can be differentiated by rectal biopsy from the ganglioneuromatosis identified in MEN2B. Although this variant was considered uncommon a decade ago, the increasing focus on the RET proto-oncogene in the development...
of the enteric nervous system has led to more frequent identification of this variant.
Multiple Endocrine Neoplasia Type 2B

The association of MTC and pheochromocytoma with multiple mucosal neuromas is termed MEN2B (formerly MEN3). The hallmark of this syndrome is the presence of characteristic mucosal neuromas on the distal portion of the tongue, on the lips and subconjunctival areas, and throughout the gastrointestinal tract. Thickened corneal nerves may be identified by slit lamp examination, and enlarged nerves are frequently noted during neck or abdominal surgery. Ganglioneuromatosis of the gastrointestinal tract can cause obstruction, dilation of the colon, or a colic-like childhood syndrome with associated diarrhea and may be the first clinical manifestation of MEN2B. Other features associated with this syndrome include a marfanoid habitus, pectus excavatum, slipped femoral epiphysis, and long, thin extremities.

The mucosal neuroma phenotype is associated, in all reported cases, with bilateral and multicentric C-cell hyperplasia or MTC, or both. The MTC in this syndrome is more aggressive than that in MEN2A. Metastatic C-cell disease can occur in children younger than 1 year of age and there is a shorter average survival time in patients with metastatic disease. However, the presence of multigenerational families and a more extensive compilation of outcomes suggest that long-term survival is more common than indicated by earlier reports. Unilateral or bilateral pheochromocytoma occurs in approximately half of the individuals with this disorder, and occurs at similar ages, and is histologically similar to that seen in MEN2A.

The identification of the mucosal neuroma phenotype in a child should alert the physician to the diagnosis of MTC. It is important to confirm the diagnosis of MEN2B by DNA testing (discussed later) because there are rare examples of mucosal neuromas without other features of MEN2B or a RET mutation. In most cases, a RET mutation is identified; in those who do not have a codon 918, 883, or 922 mutation, calcitonin testing would be appropriate. In children with MEN2B, total thyroidectomy should be performed during the first month of life. It is not known whether such treatment is curative because experience is limited. The expressivity of the mucosal neuroma phenotype may be less than 100%. A case report in which a mother and one child had mucosal neuromas and MTC and a second child had MTC but no evidence of the mucosal neuroma syndrome suggests this possibility.

Hyperparathyroidism is rare in MEN2B.
The net effect of these rearrangements is that the tyrosine kinase of RET is expressed in a cell type (the thyroid follicular cell) that does not normally express this gene. These rearrangements involve two different genes and differ substantially from the single nucleotide changes that are most frequently associated with MEN2, discussed later. These gene rearrangements lead to expression of a structurally normal RET kinase domain fused to one of several proteins expressed in thyroid follicular cells. Other than a few reports of rearrangement of RET and a second extracellular protein, GFR-1, together form a receptor for glial cell-derived neurotrophic factor (GDNF), a secretory peptide. A fascinating series of events led investigators to piece together the components of this signaling system. Investigators seeking to understand the roles of RET and GDNF created transgenic knockout mouse models for these two genes and independently discovered a nearly identical phenotype in both. In a parallel search for a GDNF receptor, a 468-amino-acid protein that bound GDNF was isolated from an embryonic rat midbrain complementary DNA library. This protein was designated GDNFR-1/GDNF receptor, now GFR-1. GFR-1 is a glycosyl-phosphatidylinositol-linked cell-surface protein that lacks cytoplasmic and transmembrane domains. It is expressed in GDNF-responsive cells and binds GDNF with high affinity. GFR-1

### TABLE 36-4 — Germline Mutations of the RET Proto-oncogene in Multiple Endocrine Neoplasia Type 2

<table>
<thead>
<tr>
<th>Affected Codon</th>
<th>Exon</th>
<th>Amino Acid Change Normal/Mutant</th>
<th>Nucleotide Change Normal/Mutant</th>
<th>Clinical Syndrome</th>
<th>Percentage of All MEN2 Mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>609</td>
<td>10</td>
<td>CysArg</td>
<td>TGCCGC</td>
<td>MEN2A/FMTC</td>
<td>01</td>
</tr>
<tr>
<td>611</td>
<td>10</td>
<td>CysTyr</td>
<td>TGCTAC</td>
<td>MEN2A/FMTC</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CysTrp</td>
<td>TGCTAC</td>
<td>FMTC</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CysGly</td>
<td>TGCGGC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>618</td>
<td>10</td>
<td>CysSer</td>
<td>TGCCGC</td>
<td>MEN2A/FMTC</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CysGly</td>
<td>TGCCGC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CysArg</td>
<td>TGCCGC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CysPhe</td>
<td>TGCTTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CysSer</td>
<td>TGCTCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CysTyr</td>
<td>TGCTAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CysEnd</td>
<td>TGCTGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>620</td>
<td>10</td>
<td>CysArg</td>
<td>TGCCGC</td>
<td>MEN2A/FMTC</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CysTyr</td>
<td>TGCTAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CysPhe</td>
<td>TGCTTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CysSer</td>
<td>TGCTCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CysGly</td>
<td>TGCGGC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>630</td>
<td>11</td>
<td>CysPhe</td>
<td>TGCTTC</td>
<td>FMTC</td>
<td>&lt;0.1</td>
</tr>
<tr>
<td>634</td>
<td>11</td>
<td>CysSer</td>
<td>TGCCGC</td>
<td>MEN2A</td>
<td>8090</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CysGly</td>
<td>TGCCGC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CysArg</td>
<td>TGCCGC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CysTyr</td>
<td>TGCTAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CysPhe</td>
<td>TGCTTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CysSer</td>
<td>TGCTCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>GluAsp</td>
<td>GAGGAC</td>
<td>FMTC</td>
<td>01</td>
</tr>
<tr>
<td>790</td>
<td>13</td>
<td>LeuPhe</td>
<td>TTGT TT</td>
<td>MEN2A/FMTC</td>
<td>&lt;0.1</td>
</tr>
</tbody>
</table>
is required for GDNF to bind and activate the RET receptor, forming a multisubunit receptor system in which GFR-1 is a ligand-binding component and RET is the signaling component (Fig. 36-16). Subsequent experiments showed that targeted disruption of GFR-1 produced the same phenotype as disruption of RET or GDNF. During the past 5 years, a family of GFR proteins (GFR-2, GFR-3) and GDNF-related ligands (artermin, persephin, and neurturin) has been identified. Each of these is likely to have an important developmental role, but there is currently no evidence for their involvement in MEN2.

One important and well-defined function of the RET receptor complex is to direct normal migration of several cell types during embryologic development. The targeting occurs through an interaction of the RET receptor system (RET and GFR-1) with GDNF. RET is expressed in several tissues of neural crest derivation including the neural crest (Fig. 36-15), peripheral nerve (nerve and muscle), and several extramural mesenchymal tissues (e.g., gut). The neural crest is a major source of cell types that migrate to sites of somatic tissues, where they give rise to mesenchymal cell types. In this manner, RET is involved in the normal migration of a variety of cell types. The importance of RET in normal development is illustrated by the observation that most patients with MEN2 have mutations in the RET gene, and a small number of patients with MEN2 have mutations in the GFR-1 gene. The majority of patients with MEN2 have activating mutations of RET, and a minority of patients have inactivating mutations of RET. The activating mutations of RET are located in the TK domain, whereas the inactivating mutations are located in the extracellular domain.

In contrast to MEN1, in which there is no relationship between a specific mutation and clinical phenotype, in MEN2 there is a high degree of genotype-phenotype correlation. Mutations of these two codons have been reported in Hirschsprung disease. Reported cases of MEN2/Hirschsprung disease variants have these mutations.

There is a similar relationship in the gastrointestinal tract. RET/GFR-1-expressing neurons from the neural crest migrate into the developing gastrointestinal tract, presumably enticed by GDNF-expressing cells. Predictably, disruption of any component of this system leads to a disordered migration of neurons into the gastrointestinal tract and a Hirschsprung-like phenotype. These observations are clinically relevant to both the development of mucosal neuromas (derived from the enteric nervous system) in MEN2B and disordered neural ganglia found in Hirschsprung disease. The RET mutations found in MEN2B are activating, whereas about 50% of hereditary Hirschsprung disease is attributable to inactivating mutations of RET. There is an unresolved and interesting paradox of hereditary Hirschsprung disease arising from either activating or inactivating mutation of RET.

**Figure 36-15** Molecular abnormalities of the RET proto-oncogene in multiple endocrine neoplasia type 2 (MEN2). Mutations of the RET proto-oncogene have been identified in MEN2A, familial medullary thyroid carcinoma (FMTC), MEN2A associated with Hirschsprung disease, MEN2A associated with cutaneous lichen amyloidosis (CLA), and some somatic mutations in sporadic MTC. Two regions of the RET tyrosine kinase are affected. The first is a cysteine-rich extracellular domain (Cys-Rich) important for dimerization of the ret receptor (codons 609, 611, 618, 620, 634). Mutations of individual cysteines at these codons cause RET dimerization, activation, autophosphorylation, and transformation. Mutations of the second region, the intracellular tyrosine kinase (TK) domain involving codons 768, 790, 791, 804, 883, 891, 918, and 922, causes activation, autophosphorylation, and transformation. A role for the cadherin-like region (Cadherin) has not been defined.

The diagram shows the exon-intron structure of the RET gene and the location of the mutations identified in most germline carriers of MEN2. A more complete listing of rare mutations can be found in the Human Gene Mutation Database (http://archive.uwcm.ac.uk/uwcm/mg/ns/1/120346.html#ref35) and Online Mendelian Inheritance in Man (http://www.ncbi.nlm.nih.gov/htbin-post/Omim/dispmin?164761). In addition, there are rare mutations that have been identified in individual patients or in sporadic tumors. In some cases the physical location of these rare coding changes adjacent to a sequence known to be mutated gives credibility to the role of the mutant sequence; in others the relevance is unclear. Further clarification of their relevance awaits examination of their transforming abilities or additional genotype-phenotype correlation.

Mutations of these two codons have been reported in Hirschsprung disease. Reported cases of MEN2/Hirschsprung disease variants have these mutations.

There is an unresolved and interesting paradox of hereditary Hirschsprung disease arising from either activating or inactivating mutation of RET.

**Mutations of the RET Proto-Oncogene: Correlations with Multiple Endocrine Neoplasia Type 2 Variant or with Tumor Phenotype**

RET can undergo gain of function by two types of mutations. The first, discussed two sections earlier, is a rearrangement in which genomic DNA from the RET kinase region is fused to a promoter sequence of another gene, creating the RET-PIC oncogene. The mutations of RET found in MEN2 are missense (amino acid or codon change) mutations. RET missense mutations in MEN2 fall into two broad categories: those that affect a group of highly conserved extracellular cysteine residues and a second group of intracellular mutations.

### Germ Line RET Mutations

The most common mutations in MEN2 affect codons 609, 611, 618, 620, 630, and 634, all encoding extracellular cysteine residues (see Fig. 36-15). Mutations at codon 634 account for more than 80% of mutations in MEN2, and a single coding change, cysteine to arginine at codon 634, is found in approximately 50% of cases.

In contrast to MEN1, in which there is no relationship between a specific mutation and clinical phenotype, in MEN2 there is a high degree of correlation. Kindreds with any substitution at codon 634 invariably have MEN2A or one of its variants (Table 36-3), and all reported patients with the MEN2A/cutaneous lichen amyloidosis variant have a mutation at the same codon. Mutations of codons 609, 618, and 620 have been identified in the MEN2A/Hirschsprung variant of MEN2. Other less common extracellular mutations are listed in Table 36-4.

Mutations of the intracellular domain of RET predominantly affect codons 768, 790, 791, 804, 883, 891, 918, and 922. Most cases of MEN2B have a codon 918 mutation, although there are rare examples of codon 883 and 922 mutations. Codon 883, 918, and 922 mutations account for approximately 5% of all mutations in hereditary MTC. The other intracellular domain mutations (codons 768, 790, 791, 804, and 891) are uncommon and account for 2% to 3% of all mutations.

Familial MTC has been identified in kindreds with codon
609, 611, 618, 620, 768, V804M, and 891 mutations, making the genotype-phenotype correlation for this clinical syndrome the least specific of the MEN2 variants. Indeed, there is overlap between FMTMC and MEN2A for codons 609, 611, 618, and 620 mutations. In a large French series, 60% of 148 patients with FMTMC in 47 kindreds had intracranial non-cysteine mutations (codons 768, 790, 804, 891); the remaining 40% had extracranial cysteine mutations at codons 611, 618, and 620. 1\textsuperscript{5} One experimentally based hypothesis to explain this overlap is that mutations of codons 609, 611, 618, 768, 804, and 891, relative to codons 634 or 918, have a lower in vitro transforming efficiency. 1\textsuperscript{6} Because MTC is most commonly the earliest manifestation of MEN2A, it is possible that kindreds with these RET mutations would develop other manifestations of MEN2A if they lived long enough. Alternatively, other potential explanations such as small population-based polymorphisms of the RET gene or GFR-1 or GDNF causing differential activity of this receptor system could be invoked.

The RET genotype-phenotype correlation is remarkable. Not only is it possible to predict the clinical phenotype with reasonable certainty (see Fig. 36-15), it is also possible to predict, albeit with considerably less certainty, the aggressiveness of MTC associated with a particular germ line mutation. There is an evolving literature attempting to correlate the clinical aggressiveness of MTC with in vitro studies of transforming efficiency. 1\textsuperscript{23, 24} Although it is possible to make broad correlations, they are at present imperfect in most respects. However, within a given family specific mutations appear to behave in a more predictable manner. For example, in some kindreds with intracranial domain mutations with FMTMC, there has never been a death caused by MTC, although the same mutations in other kindreds can be associated with more aggressive disease. These findings suggest that genetic differences of the RET gene or other components of the RET signaling system (GDNF, GFR-1, or downstream effectors in the kinase signaling cascade) may modify the impact of a particular mutation. Clearly, there is more to learn.

Germ line mutations of the RET proto-oncogene have also been identified in Hirschsprung disease. Many of these are inactivating mutations, 1\textsuperscript{25} 1\textsuperscript{26} 1\textsuperscript{27} 1\textsuperscript{28} 1\textsuperscript{29} 1\textsuperscript{30} 1\textsuperscript{31} although the association of Hirschsprung disease alone, or the MEN2A/Hirschsprung disease variant, with apparent activating mutations at codons 609, 618, and 620 (see Table 36-4) suggests that disordered expression can also lead to this phenotype. 1\textsuperscript{32}

Somatic mutation of the MEN1 gene contributes to the causation of certain sporadic tumors, with a spectrum similar to that of the tumors seen in MEN1. A similar phenomenon is seen for somatic RET mutations in sporadic MTC and, to a much lesser extent, pheochromocytoma. Twenty-five percent of sporadic MTCs have a somatic RET mutation (60 of 236 cases). 1\textsuperscript{33} 1\textsuperscript{34} 1\textsuperscript{35} 1\textsuperscript{36} 1\textsuperscript{37} 1\textsuperscript{38} 1\textsuperscript{39} 1\textsuperscript{40} 1\textsuperscript{41} 1\textsuperscript{42} 1\textsuperscript{43} The most common is a somatic codon 918 mutation identical to that found in the germ line in MEN2B, but other codons (609, 611, 618, 620, 631, 634, 768, 804, 883, and 891) are also mutated somatically. The presence of a codon 918 somatic mutation in an MTC is associated with a greater frequency of distant metastasis 1\textsuperscript{44} and shorter survival. 1\textsuperscript{45} In this respect, the behavior of the sporadic tumor with a somatic mutation parallels the aggressive behavior of MTC in the context of MEN2B caused by a germ line codon 918 mutation. 1\textsuperscript{46}

Somatic RET mutations are considerably less common in sporadic pheochromocytomas than in sporadic MTC. Somatic RET mutations were identified in 2 of 35, 1\textsuperscript{53} 1\textsuperscript{54} 1\textsuperscript{55} 1\textsuperscript{56} 1\textsuperscript{57} to 4 of 48, 1\textsuperscript{58} and 0 of 17 sporadic pheochromocytomas, for a total of 4 to 100 (4% to 6%). These results indicate that somatic RET mutation is an infrequent cause of sporadic pheochromocytoma. Similarly, in the only study to examine for the presence of somatic RET mutations in sporadic parathyroid tumors, none of 35 tumors examined had codon 609, 611, 618, 620, 630, 634, or 918 mutations, making it unlikely that somatic RET mutations play any significant role in sporadic hyperparathyroidism. 1\textsuperscript{59}

RET-PTC mutations in papillary thyroid cancer and in a few other neoplasms have already been described. These too are somatic RET mutations, but they have no counterpart in the form of RET germ line mutations.

Tumorigenesis: Mechanism of RET-Induced Transformation

The RET gene is normally expressed in the several cell types involved in MEN2, including the C cell, the parathyroid cell, and the adrenal medullary cell. Studies demonstrating tumorigenic mechanisms were performed in an in vitro transformation assay (the NIH3T3 cell culture system), utilizing a cell type that does not usually express RET (the only mutant RET was expressed) and is readily transformed. These studies, from several different laboratories, 1\textsuperscript{60} 1\textsuperscript{61} 1\textsuperscript{62} provide evidence for two different mechanisms of transformation.

The first mechanism is applicable to extracranial cysteine mutations (protoype was mutant codon 634). The cysteine-rich region has been shown to be important for normal RET receptor dimerization. Homozygous expression of a mutant RET receptor in NIH3T3 cells resulted in dimerization in the absence of either ligand or GFR-1 and activation of downstream signaling pathways. 1\textsuperscript{63} 1\textsuperscript{64} There is evidence that GDNF is not required for activation of the dimerized ret and, indeed, does not further activate the tyrosine kinase. 1\textsuperscript{65} A second mechanism has been demonstrated for a mutation of the tyrosine kinase domain at codon 918. This mutation causes activation of JNK (c-Jun NH2-terminal terminal.) and phosphorylation of downstream substrates (in the absence of RET dimerization or interaction with either GFR-1 or GDNF). It has been shown experimentally that GDNF can further activate RET kinase in the presence of a codon 918 mutation. 1\textsuperscript{66}

Activation of RET by GDNF or by intrinsic activating mutations leads to autophosphorylation of the tyrosine at codons 1015 and 1062 (see Fig. 36-16). 1\textsuperscript{67} There is considerable evidence that autophosphorylation of tyrosine 1062 is required for activation of downstream effector pathways and transformation. 1\textsuperscript{68} 1\textsuperscript{69} 1\textsuperscript{70} 1\textsuperscript{71} A third mechanism results from the gene termed RET-PTC; this is RET fused to regulatory elements of one of a small number of other genes, most commonly in the thyroid follicular cell (see earlier). These rearrangements drive RET expression in a cell type that does not normally express this gene and promote dimerization.

Tumorigenesis: Steps Diestal to Ret

At least three different pathways, JNK, 1\textsuperscript{72} 1\textsuperscript{73} 1\textsuperscript{74} 1\textsuperscript{75} 1\textsuperscript{76} 1\textsuperscript{77} and phospholipase C gamma, 1\textsuperscript{78} 1\textsuperscript{79} are activated through shc-grb2-src proteins linked to ret. 1\textsuperscript{80} 1\textsuperscript{81} 1\textsuperscript{82} There is also evidence that nuclear factor B is activated by mutant RET and is dependent upon activation of raf and MEKK1. 1\textsuperscript{83} It appears that additional genetic events are involved in the development or progression of the transformed phenotype. Multiple studies have demonstrated a consistent LOH of chromosome 1p, 3p, and 22q in MEN2related MTC 1\textsuperscript{84} 1\textsuperscript{85} 1\textsuperscript{86} and pheochromocytoma. 1\textsuperscript{87}

Loss of the normal RET allele or amplification of the mutant RET allele is found in a significant number of MEN2 tumors. 1\textsuperscript{88} 1\textsuperscript{89} These studies demonstrate somatic amplification of the mutant RET allele by at least two different mechanisms: trisomy of chromosome 10 (Fig. 36-17) and duplication of the mutant RET allele. 1\textsuperscript{90} 1\textsuperscript{91} 1\textsuperscript{92} 1\textsuperscript{93} Alternatively, in a small percentage of tumors there is loss of the normal RET allele. These mutational rearrangements are formally analogous to the second hit at a tumor suppressor gene (see earlier). They help explain why germ line gain-of-function tumorigenesis as in MEN2 shares major properties with germ line loss-of-function tumorigenesis as in MEN1; specifically, both processes have early onset and tumor multiplicity.

It is hypothesized that the effect of amplification of the mutant RET allele or loss of the normal copy results in predominant expression of a mutant RET receptor, a model analogous to that demonstrated for KRAS\textsuperscript{2}\textsuperscript{19} and MET. 1\textsuperscript{94} In these examples, the normal gene moderates the impact of the mutant version. Several mechanisms could be envisioned to explain this mechanism for RET that are directly related to dimerization of the receptor, interaction of the receptor variants with GDNF or linker molecules, or differential activation of downstream mediators by the two different variants. None of these has been proved. This mechanism of unbalanced RET expression was not identified earlier, presumably because LOH of chromosome 10, in fact, occurs in only a small percentage of tumors.
Testing for Carrier State and for Tumor Emergence in Multiple Endocrine Neoplasia Type 2

The identification of RET proto-oncogene mutations causing MEN2 and FMTM has simplified carrier ascertainment and MTC surveillance in families with identifiable mutations (see Table 36-4). Several different analytic techniques have been applied to the identification of mutations, although direct DNA sequencing remains the most widely used. These analytic tests are readily available throughout North America and Europe at modest cost from several commercial sources. Almost 10 years have passed since the discovery of RET mutations in MEN2, and it is now clear that DNA-based diagnosis will replace measurement of calcitonin after pentagastrin or calcium stimulation for carrier ascertainment. During this period, additional mutations have been described with some regularity and there are now few kindreds with hereditary MTC that do not have an identifiable mutation. In the rare kindred in which a mutation is not identified, continued calcitonin stimulation testing with calcium or pentagastrin (pentagastrin is not currently available in the United States but is available in other countries) is required.

Calctonin Measurement for Carrier Ascertainment and for Tumor Surveillance

The pentagastrin test is performed by measuring the serum calcitonin level before and after 2, 5, and 10 minutes after the intravenous injection of pentagastrin (0.5 μg/kg body weight). Administration of calcium immediately before the pentagastrin injection enhances the sensitivity of the test. A short calcium infusion (15 minutes) can also be used to stimulate calcitonin release. Calcium is the only potent calcitonin secretagogue currently available in the United States.

A positive test is one in which either the basal serum calcitonin concentration is elevated and is further increased by the administration of pentagastrin or calcium or one in which the basal value is normal but increases into the abnormal range.

A listing of available commercial testing sources is available through a Web site: http://www.genetests.org.

After the administration of pentagastrin or calcium. It is important that the samples be analyzed with the most sensitive assay available and with proper control samples; it is now possible to measure normal serum calcitonin levels (0.15 to 3 pmol/L [0.5 to 10 pg/mL]) routinely, thereby making it possible to separate normal subjects from those with early C-cell hyperplasia. Criteria that are useful for separation of normal from abnormal kindred members include a patient known to be affected and a consistently abnormal test result (two or more nonconsecutive test results that are abnormal). It is important to point out, however, that there is considerable overlap between the normal range of serum calcitonin after a provocative test and that observed in patients with early abnormalities of the C cell, a finding that was not fully recognized until RET mutation testing became available in 1993. In nearly every large kindred in which calcitonin abnormalities were used to determine carrier status, there are examples of false-positive calcitonin tests that in some kindreds approached 15% of the total number of carriers. In retrospect, the use of calcitonin testing was less of a “gold standard” than thought at the time.

The clinician confronted with the necessity to continue calcitonin testing in a kindred with no identifiable RET mutation should keep this experience in mind when assessing minor calcitonin abnormalities in such kindreds. Despite this caveat, it has been the authors’ experience that when an abnormal calcitonin test result has been identified, pressure builds to consider total thyroidectomy. A discussion with the family of the risks (hypoparathyroidism, need for continued thyroid supplementation, and recurrent laryngeal nerve damage) and a firm plan of continued testing at 6- to 12-month intervals often allay parental concern or provide a more balanced viewpoint. It is clear from the 25-year experience with provocative calcitonin testing that there is progression of calcitonin abnormalities over a several year period in carriers, the diagnosis. Whether a deferral of decision to operate increases the risk of metastatic disease is unknown, although long-term experience with children with codon 634 mutations indicates that 85% to 90% remain disease-free 25 years after total thyroidectomy, performed at an average age of 13 years.

An elevation of the basal serum calcitonin level with no further increase after a provocative test can be difficult to interpret. Such a test result can be seen in association with C-cell abnormalities, but this type of result is most likely to be caused by a nonspecific or false-positive increase of the serum calcitonin concentration or by production of calcitonin by a tumor other than MTC (lung carcinoma, hepatoma, pheochromocytoma, pancreatic islet cell tumor, or benign liver disease). Separation of a false-positive test result from a true elevation of the serum calcitonin concentration can be achieved by a radioimmunoassay using a different polyclonal antiserum or a two-site immunoradiometric assay. Establishment of ectopic production of calcitonin by a tumor other than MTC can be more difficult because calcitonin release is frequently enhanced by calcium or pentagastrin; however, such tumors frequently produce a high-molecular-weight form of unprocessed procalcitonin, a finding that may be helpful in selected cases.

Method of RET Sequence Testing in DNA

DNA diagnostic techniques currently in use to identify MEN2 gene carriers are based on the use of polymerase chain reaction techniques to amplify selected portions of the RET proto-oncogene known to be mutated in MEN2. All laboratories analyze for mutations in exons 10, 11, and 16 (codons 609, 611, 616, 620, 630, 634, and 918), those most commonly found in MEN2. If a mutation is not identified in one of these codons, some but not all laboratories analyze for mutations in exons 13, 14, and 15 (codons 768, 790, 791, 804, 883, and 891). If a mutation is not identified in exon 10, 11, or 16, it is important to identify a laboratory that will examine the other exons before concluding that no mutation is present. This is particularly true for kindreds with FMTM, a disproportionate number of which have an exon 13, 14, or 15 mutation. Although there is a potential for several types of errors in mutational analysis, a repeated analysis of each positive or negative test result in a separate testing facility with an independently obtained DNA sample provides nearly 100% certainty that an individual test result is accurate. About 10 years of experience with this type of testing has provided insight into its usefulness in various clinical settings.
There is now consensus that all patients with a RET mutation should be offered a total thyroidectomy. In the almost 10 years since the identification of these mutations, greater awareness of the relationship between mutation at a specific RET codon and the clinical aggressiveness of the MTC has developed. Although imperfect in many respects, a consensus has developed within the field that these difference should be considered in the decision-making process regarding early thyroidectomy. The consensus guidelines have divided hereditary MTC into three different risk categories.

Category 1. Highest risk: In the highest risk category are those with MEN2B and a codon 883, 918, or 922 RET mutation. In these children MTC with metastasis may occur during the first year of life. The presence of a pheochromocytoma is a strong indication for early thyroidectomy. The consensus guidelines recommend that these patients undergo total thyroidectomy by the age of 6 months and preferably during the first month of life. Only in exceptional circumstances, when the parents are convinced that the child may be at risk for an early thyroidectomy, is it reasonable to delay total thyroidectomy until the age of 1 year. 

Category 2. High risk: Cases with RET mutations of codons 609, 611, 618, 620, or 634 are classified as having high risk. They should have a total thyroidectomy, including removal of the posterior capsule, before the age of 5 years. This recommendation is based on the finding of microscopic MTC in two children with a codon 634 mutation at age 2 years and of nodal metastasis in one at age 5 and 6 years. There is less consensus regarding the need for central node dissection in children during the first 6 months of life and preferably during the first month. During this procedure, other level II to V lymph nodes should be sampled with performance of more extensive lymph node dissection if metastatic disease is identified.

Category 3. Intermediate risk: Cases with codons 768, 790, 791, 804, or 891 mutations are classified as having intermediate risk. The biologic behavior and clinical aggressiveness of MTC in patients with such mutations vary, but the MTC tends to be less aggressive. Lymph node metastasis and death have been identified in all of these mutations except mutation codon 791. There is no consensus regarding the age of total thyroidectomy in such patients. Two approaches are used. Some classify these patients with the high-risk category and perform a total thyroidectomy by the age of 5 years. Some recommend thyroidectomy by the age of 10 years. Others observe cases with these mutations with periodic provocative tests for calcitonin and perform a total thyroideectomy when calcitonin levels become clearly abnormal. There is consensus, however, that these patients should be followed up especially carefully if an early thyroideectomy is not performed. There are large kindreds with some of these mutations in which there has never been a death caused by MTC, making it difficult to convince family members that early thyroideectomy is indicated.

Multiple Endocrine Neoplasia Type 2 Kindred with Known RET Mutation

A normal RET analysis in a kindred with a known RET mutation excludes the carrier state with nearly 100% certainty. An individual with two independently obtained negative DNA test results in the context of a family with an identified missense mutation can be excluded from further carrier ascertainment. Pentagastrin testing in this situation adds nothing to the diagnostic accuracy and may actually confuse the clinical assessment because of a high incidence of false-positive results.

Occult Phenotypes. Should RET Germ Line Be Tested in a Case with an Apparently Sporadic Tumor in the Spectrum of Multiple Endocrine Neoplasia Type 2?

Germ Line Mutations in Sporadic Medullary Thyroid Carcinoma

A compilation of over 200 patients with apparent sporadic MTC who were examined for germ line RET mutations (codons 609, 611, 618, 620, 630, and in some cases 768) has identified mutations in approximately 6% of the total (this contrasts with data to be presented subsequently showing that approximately 25% of sporadic MTCs have a somatic RET mutation without a corresponding germ line abnormality). Subsequent investigation of the individuals in these studies with germ line RET mutations has demonstrated that the majority were members of previously unidentified kindreds or descendants of gene carriers who were separated from their kindred by many generations. Therefore, the proportion of sporadic MTCs in which a germ line mutation is identified is lower than 6%. There has been considerable disagreement among experts as to the clinical significance of finding a RET proto-oncogene mutation in an otherwise apparently sporadic tumor. The finding that 6% of patients with apparent sporadic MTC carry germ line RET mutations suggests that all patients with MTC should have a RET germ line analysis performed. This is especially important because identification of one hereditary case may have a multiplier effect, leading to the diagnosis of unsuspected cases in the family and effective early treatment. Negative family history, although useful, clearly does not exclude hereditary disease. A second reason for performing a RET germ line analysis is to reassure the patient and family members that there is no hereditary component. If the RET analysis is negative, hereditary disease can be excluded with greater than 99% certainty.

Most families of patients with sporadic MTC are assured by a less than 1% probability of hereditary disease; for those who want hereditary MTC excluded with 100% certainty, continued provocative calcitonin testing in relatives is required.

Germ Line Mutations in Sporadic Pheochromocytoma

The situation is less clear for apparently sporadic pheochromocytoma. Three genetic syndromesMEN2, VHL, and hereditary pheochromocytoma or paraganglioma syndromes present as a sporadic pheochromocytoma. This presentation is unlikely for type 1 neurofibromatosis. Estimates of frequency of heredity for pheochromocytoma in the general population have ranged from 9% to as high as 23% in some series. It is important to recognize that these estimates came from tertiary care centers and in one, located in southwest Germany, there is a high incidence of VHL (84% of those with apparent sporadic pheochromocytomas have a VHL mutation). In four of these cases in which it was tested and in most new MEN2B probands, the newly mutated allele is derived from the unaffected father, suggesting acquisition of the mutation during spermatogenesis.

The low frequency is not surprising because pheochromocytomas in MEN2 are most commonly found with codon 634 RET mutations. The 50% penetrance of pheochromocytomas in these families makes it likely that the hereditary nature of these tumors can be identified by a careful family history. Germ line mutation of GDNF has been identified in 1 of 50 sporadic pheochromocytomas. The significance of this single mutation is unclear. Similarly, germ line VHL mutations in apparent sporadic pheochromocytoma are uncommon, being found in only 9 of 173 sporadic pheochromocytomas (5.2%) in four series. Finally, there are at least four different familial pheochromocytoma-paraganglioma syndromes that have been mapped (PGL1, PGL2, PGL3, and familial pheochromocytoma-paraganglioma syndrome). Mutations of components of the succinate dehydrogenase complex gene, putative tumor suppressor genes because each of the mutations appears to be inactivating, have been identified in three of the four variants.

There is currently no information regarding the frequency of these mutations.

Estimates based on genetic testing of RET in three other studies were that 2 of 230 patients (<1%) with apparently sporadic pheochromocytoma had a germ line mutation. The low frequency is not surprising because pheochromocytomas in MEN2 are most commonly found with codon 634 RET mutations. The 50% penetrance of pheochromocytomas in these families makes it likely that the hereditary nature of these tumors can be identified by a careful family history. Germ line mutation of GDNF has been identified in 1 of 50 sporadic pheochromocytomas. The significance of this single mutation is unclear. Similarly, germ line VHL mutations in apparent sporadic pheochromocytoma are uncommon, being found in only 9 of 173 sporadic pheochromocytomas (5.2%) in four series. Finally, there are at least four different familial pheochromocytoma-paraganglioma syndromes that have been mapped (PGL1, PGL2, PGL3, and familial pheochromocytoma-paraganglioma syndrome). Mutations of components of the succinate dehydrogenase complex gene, putative tumor suppressor genes because each of the mutations appears to be inactivating, have been identified in three of the four variants.

There is currently no information regarding the frequency of these mutations.

The decision to pursue RET and VHL germ line testing in patients with apparent sporadic pheochromocytoma is somewhat more complicated than testing for MTC. There are several arguments for testing. When the incidences of RET and VHL mutations are combined (6%), they approximate the incidence observed for germ line RET mutations in apparent sporadic MTC. In addition, the argument can be advanced that identification of hereditary pheochromocytoma prevents sudden death or significant morbidity related to pheochromocytoma. In the case of MEN2A, results in earlier identification of MTC in other affected family members. The arguments against testing include the low frequency of mutations (particularly for RET), the necessity to test for both MEN2A and VHL (and possibly familial pheochromocytoma-paraganglioma syndrome) leading to a substantial expenditure, and the fact that most hereditary pheochromocytoma comes to attention because of its distinctive clinical features before serious morbidity. A pragmatic approach is to apply criteria similar to those outlined for carrier ascertainment in MEN1. The information is useful, helps families to address health problems, provides reassurance if the test is negative, and leads to appropriate screening when a mutation is identified. It is, however, not absolutely necessary because no specific clinical action other than screening for pheochromocytoma or other neoplastic components of VHL or MEN2 is based on the test result.
Mutations at codons 768, 791, V804M, and 891 have been associated exclusively with FMTC; nonetheless, it would be prudent to consider a screening urine catecholamines or metanephrine test for pheochromocytoma at the time of diagnosis and perhaps at 5-year intervals. Codon 609, 611, 618, 620, or 630 mutations have been associated with either FMTC or MEN2A. Unless there is a several generation pattern of FMTC in more than 10 affected family members, it would be prudent to screen every 1 to 3 years for pheochromocytoma. The molecular basis for the phenotypic variability (FMTC or MEN2A) with codon 609, 611, 618, 620, or 630 mutations is unknown at present, although an answer may be found in the population-based polymorphisms discussed earlier in this chapter.

Management of an Established Multiple Endocrine Neoplasia Type 2 Kindred

Counseling

Family education is an important component in the management of MEN2. Although a physician’s legal responsibility may end after immediate family members have been notified of the genetic nature of the disease, it is prudent to encourage patients to make even family members at distant risk aware of the nature of the disease. The fact that the disease appears to be benign in one generation should not deter carrier ascertainment and tumor surveillance efforts because the disease may assume a more virulent expression in a subsequent generation. This notification can be done by giving pamphlets describing the syndrome to immediate family members for distribution to more distant relatives (available at www.endocrine.mdacc.tmc.edu).

Prospects for Surgical Cure of Medullary Thyroid Carcinoma in Multiple Endocrine Neoplasia Type 2

Widespread prospective carrier ascertainment and tumor surveillance have had an impact on the natural history of the syndrome. The age at carrier ascertainment has progressively fallen from a mean of 33 years when prospective carrier ascertainment through tumor surveillance first began in 1969 to a mean below 13 years in 1988. The current mean age with widespread DNA testing is likely to be below the age of 5 years. Testing for RET mutations now makes it possible to identify carrier status at birth or in utero.

Whether prospective DNA-based ascertainment and early thyroidectomy are curative for the thyroid neoplasm is less clear. Follow-up data from several groups indicate that approximately 85% to 90% but not 100% of kindred members who were thyroidectomized for early disease on the basis of pentagastrin testing have normal or nondetectable calcitonin values at mean follow-up periods ranging from 1 to 15 years. It can be anticipated that earlier DNA-based ascertainment and treatment are likely to improve the outcome in gene carriers. None of the cases of MEN2A or FMTC identified by DNA-based carrier ascertainment and operated before the age of 5 years have had identifiable metastasis, although there was considerable variability in these small series in the numbers of nodes sampled.

The identification of MTC (without nodal metastasis) in a significant number of children with MEN2A between the ages of 2 and 5 years suggests that a child younger than 5 years with metastatic disease may be identified. There are insufficient data regarding the impact of a specific mutation on the presence or absence of metastasis in younger children, although several groups are now combining information from national databases, making it possible that such data will be available in the next few years. In summary, early carrier ascertainment and treatment by total thyroidectomy at an early age appear to identify and treat patients before there is identifiable lymph node metastasis. Thus, it is possible that death from hereditary MTC will be largely eliminated by this early intervention, making it the first example of successful use of genetic ascertainment to eliminate death from malignancy. However, it will be many years before the impact of this earlier intervention on cure of MTC will be demonstrated.
Overlap Syndromes

Overlap syndromes in single patients include gastrinoma in a MEN2 patient, adenomatous polyposis coli with MEN1 or MEN2B, posterior pituitary tumor and MEN1, pheochromocytoma in a patient with MEN2A, and pheochromocytoma in MEN1 (see earlier). There is a single case report of an ovarian strumal carcinoid tumor, a variant of an ovarian teratoma, in a patient with MEN2A. The tumor was composed of neuroendocrine cells with thyroid-like follicles that stained positive for thyroglobulin.

Most cases of overlap of MEN1 and MEN2 were published before the era of discovery of syndrome-causing genes. Currently, most can be understood as follows:

1. Unusual expressions of a syndrome. Thus, VHL can cause pheochromocytoma and islet tumor (see later), neurofibromatosis type 1 (NF1) can cause pheochromocytoma and duodenal somatostatinoma (see later), and MEN1 can cause pheochromocytoma with any other feature of MEN1 (see before).
2. Coexistence of two rare disorders.
3. Rare syndromes that have not yet been characterized sufficiently.
4. Few cases of unexplained overlap. In this regard, a mouse model (see earlier) has many features of both MEN1 and MEN2, suggesting a single pathway to both syndromes or an overlap.
Familial Occurrence of Two or More Endocrine Neoplastic Disorders

MEN1 and MEN2 are the only multiple neoplasia syndromes in which the two most prominent features are hormone-secreting tumors. In other MEN syndromes, nonhormonal tumors are more urgent. For example, the McCune-Albright syndrome (which is not hereditary) features fibrous dysplasia of bone and cafe-au-lait spots of skin (Chapter 27) and VHL syndrome features papillary renal cancer and central nervous system hemangioblastomas.

Von Hippel-Lindau Syndrome

VHL syndrome is an autosomal dominant neoplastic syndrome characterized by hemangioblastomas of the central nervous system, retinal angiomas, renal cell carcinomas, visceral cysts, pheochromocytoma, and islet cell tumors. More than 90% of gene carriers express one or more of the manifestations of this disorder by the age of 60 years. Over 70% of gene carriers have one or more central nervous system tumors. Of particular relevance to endocrinologists is the observation that 25% to 35% of these patients have unilateral or bilateral pheochromocytomas and 15% to 20% have islet cell tumors. Although the islet cell tumors may immunostain weakly for insulin, they virtually never hypersecrete it.

The VHL gene was mapped to chromosome 3p25.3 and identified by positional cloning. This gene is a tumor suppressor gene, implying that loss of function or inactivating mutations of both alleles or copies of this gene are associated with tumor formation. Studies have described an inhibitory effect of the VHL-encoded protein on transcription elongation through its binding to an elongin B/C complex. Mutation of the VHL gene, particularly in the region of codons 150 to 170, interferes with this interaction, resulting in an accelerated rate of transcription elongation.

Another mechanism may explain many of the properties of the VHL protein and disease. The action of the VHL protein to facilitate the proteasome-mediated degradation of the hypoxia-inducible factor 1 (HIF-1) protein and other proteins may prove central to many manifestations of VHL gene alteration. Mutation of codon 238 was identified in over 40% of VHL families with pheochromocytoma, suggesting that families with a mutation in this codon should be surveyed routinely for pheochromocytoma. As with other recessive oncogenes such as MEN1, p53, BRCA1, or retinoblastoma gene, a large number of inactivating mutations have been described for VHL.

Clinical management for patients with VHL syndrome is often complicated by the presence of renal or central nervous system tumors. Pheochromocytomas or islet cell tumors associated with hypertension, cardiac arrhythmias, hypoglycemia, watery diarrhea, carcinoid, or a glucagonoma-like picture should be surgically excised.

Von Hippel-Lindau Syndrome

Although the islet cell tumors may not hypersecrete insulin, they virtually never hypersecrete it. The association of pheochromocytoma and islet cell tumors can occur in familial or nonfamilial patterns. There is little information about the molecular genetics of these rare disorders, although it is possible that abnormalities of the VHL gene may be involved.

Neurofibromatosis Type 1

The main features of NF1 are neurofibromas and dermal cafe-au-lait spots. NF1 has been associated with a variety of endocrine neoplasms including pheochromocytoma, hyperparathyroidism, somatostatin-producing carcinoid tumors of the duodenal wall, MTC, and hypothalamic or optic nerve tumors that cause precocious puberty. The causative gene for NF1 encodes a ras guanosine triphosphatase (GTPase)-activating protein (GAP) of 2815 amino acids, named neurofibromin, which accelerates GTP hydrolysis on p21 ras. Loss of the GTPase-activating function of neurofibromin (through mutation or allelic loss) leads to p21 ras activation.

More specific evidence for a role of this protein in endocrine tumors is shown by allelic loss of this gene in NF1-associated sporadic pheochromocytomas. Targeted disruption of the mouse NF1 gene resulted in sympathetic ganglia hyperplasia, providing additional evidence for a potential role of this gene in the genesis of endocrine tumors derived from neural crest tissue.

Carney's Complex

Carney's complex comprises myxomas of the heart, skin, and breast; spotty skin pigmentations; testicular, adrenal cortical, and growth hormone-secreting pituitary tumors; and peripheral nerve schwannomas. Linkage analysis has identified a locus at 2p in half of the families and another locus at 17q in most others. The gene at 17q has been identified as encoding the regulatory subunit (type IA) of protein kinase A (PRKA1A), and it has tumor suppressor properties.

The activating GNAS1 mutations in McCune-Albright syndrome and the inactivating PRKA1A mutations in Carney's complex are likely to cause tumors in selected tissues with a similar tissue spectrum by raising cyclic adenosine monophosphate.
Confusion and Contrasts between Multiple Endocrine Neoplasia Type 1 and Type 2

There is occasional confusion between MEN1 and MEN2, particularly among patients, paramedical personnel, and non-specialists. The main source of confusion is the similar syndrome names. Several other similarities are recognized (see the introduction). In fact, the differences are more important than the similarities. MEN2 includes a cancer that can be prevented or cured by timely surgery. MEN1 also causes cancers, but these mostly cannot be prevented or cured. The principal tumors differ between MEN1 and MEN2, as do surveillance protocols and treatments. Gene testing in MEN2 should be started extremely early and management should be dependent on the phenotype or genotype. Gene testing in MEN1 has debatable indications for early usage and no genotype-phenotype correlation. Mutations of RET cause tumors in MEN2 by gain of function, whereas MEN1 does this by loss of function. As a result, the oncogenic RET mutations are confined to selected loci of the gene and are thus easier to identify than those in the MEN1 gene.
References


1730
References


198. Reference deleted.


211. Galett D, Losion PT, Orth DN. Two patients with Cushings disease in a kindred with multiple endocrine neoplasia type I. J Clin Endocrinol Metab 1993; 76:15801582.


312. Reference deleted.
References


1796


1957

1757


References


Benn DE, Dwight T, Richardson AL, et al. Sporadic and familial pheochromocytomas are associated with loss of at least two discrete intervals on chromosome 1p. Cancer Res 2000; 60:70487061.


1761


References


Chapter 37 - The Immunoendocrinopathy Syndromes

George S. Eisenbarth
Peter A. Gottlieb

Geneticists, immunologists, and endocrinologists have generated a wealth of new information concerning the pathogenesis of the polyendocrine autoimmune syndromes and their component disorders. In particular, the genetic loci underlying disease susceptibility and organ-specific autoantigens targeted by the immune system are being defined. Multiple distinct molecules are often the targets of autoimmunity for a single organ-specific autoimmune disorder, and in polyendocrine autoimmunity multiple molecules of multiple organs are usually targeted.

Autoantibodies highly specific for a given disorder are present before disease onset, such as anti-islet antibodies (GAD, islet cell antibody [ICA] 512 [IA-2], and insulin) in type 1 diabetes and 21-hydroxylase autoantibodies in Addison’s disease. Each specific autoantibody reacts with only a single autoantigen, although autoantigens may be present in multiple tissues. For example, 17-hydroxylase is present in both adrenal gland and gonads, and the coincidence of Graves’ hyper-thyroidism and Graves’ ophthalmopathy implies a specific shared immune target. For the most part, however, the targets of autoantibodies appear to be unrelated except for their presence as differentiation antigens in specific cells and cellular sites.

In contrast, less is known concerning the specificity of pathogenic T cells. Given the observation that cross-reactive recognition by pathogenic T-cell clones may be determined by as few as four properly spaced amino acids of a nonapeptide and the estimate that each T-cell receptor may react with a million different peptides, there is considerable potential for patterns of autoimmunity to be determined by cross-reactive T cells. One development has been the discovery within the thymus and other lymphoid tissues of what are termed peripheral antigen-expressing cells that express autoantigens such as insulin or glutamic acid decarboxylase (GAD65). Minute quantities of such molecules within the thymus can contribute to tolerance. Insulin messenger ribonucleic acid within the thymus is regulated by genetic polymorphisms of the insulin gene associated with diabetes risk.

Two distinct autoimmune polyendocrine syndromes with characteristic groupings of manifestations are readily recognized (Table 37-1). Autoimmune polyendocrine syndrome type I (APS-I) is a rare disorder with autosomal recessive inheritance that has been shown to be caused by defects in the autoimmune regulator (AIRE) gene on chromosome 21. In contrast, the most common syndrome discussed in this chapter, autoimmune polyendocrine syndrome type II (APS-II), is less well defined, including overlapping groups of disorders. A unifying characteristic within APS-II is the strong association with polymorphic genes of the human leukocyte antigen (HLA) region located on the short arm of chromosome 6 (band 6p21.3). In addition to HLA, many other genetic loci are likely to contribute to susceptibility to APS-II.

APS-II has also been known by various other names: Schmidt’s syndrome, polyglandular autoimmune disease, polyglandular failure syndrome, organ-specific autoimmune disease, and polyendocrinopathy diabetes. Such diverse names reflect the large number of studies and case reports of this syndrome and its historical importance. Studies of patients with APS-II were instrumental in the identification of the autoimmune basis of several diseases and the development of autoantibody assays (e.g., type 1 diabetes and cytoplasmic ICAs). Each of the preceding names has some shortcomings, such as failure to include the fact that both hyperfunction and hypofunction of endocrine glands can occur or failure to recognize that nonendocrine disorders such as pernicious anemia and celiac disease are a part of the syndrome.

Some authors have even broken APS-II into a third grouping called APS-III on the basis of the presence (APS-II) or absence (APS-III), for example, of autoimmune thyroid disease and type 1 diabetes without Addison’s disease. Subdividing APS-II on the basis of specific groups of component diseases appears to add little information in terms of predicting future disorders likely to develop in affected individuals or disorders likely to be present among their relatives. Therefore, in this chapter we distinguish between APS-I and APS-II. The spectrum of disorders that may be present in these syndromes is listed in Table 37-1, and the major differences between the two syndromes are highlighted in Table 37-2.
AUTOIMMUNITY PRIMER

Major determinants of autoimmune endocrine disease are T lymphocytes and autoantibodies produced by B lymphocytes. These

---

**TABLE 37-1** – Component Disorders of the Autoimmune Polyendocrine Syndromes.

<table>
<thead>
<tr>
<th>Type I</th>
<th>Type II</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Endocrine</strong></td>
<td></td>
</tr>
<tr>
<td>Addison’s disease</td>
<td>Addison’s disease</td>
</tr>
<tr>
<td>Hyoparathyroidism</td>
<td>“Geriatric” hypoparathyroidism</td>
</tr>
<tr>
<td>Primary hypogonadism</td>
<td>Primary hypogonadism</td>
</tr>
<tr>
<td>Type 1 diabetes</td>
<td>Type 1 diabetes</td>
</tr>
<tr>
<td>Hypothyroidism</td>
<td>Hypothyroidism</td>
</tr>
<tr>
<td>Graves’ disease</td>
<td>Hypophysitis</td>
</tr>
<tr>
<td><strong>Gastrointestinal</strong></td>
<td></td>
</tr>
<tr>
<td>Mucocutaneous candidiasis</td>
<td>Celiac disease</td>
</tr>
<tr>
<td>Chronic active hepatitis</td>
<td></td>
</tr>
<tr>
<td>Malabsorption</td>
<td></td>
</tr>
<tr>
<td>Oral squamous cell carcinoma</td>
<td></td>
</tr>
<tr>
<td><strong>Dermatologic</strong></td>
<td></td>
</tr>
<tr>
<td>Alopecia</td>
<td>Alopecia</td>
</tr>
<tr>
<td>Vitiligo</td>
<td>Vitiligo</td>
</tr>
<tr>
<td>Nail dystrophy</td>
<td>Dermatitis herpetiformis</td>
</tr>
<tr>
<td><strong>Hematologic</strong></td>
<td></td>
</tr>
<tr>
<td>Pernicious anemia</td>
<td>Idiopathic thrombocytopenic purpura</td>
</tr>
<tr>
<td><strong>Pure red cell hypoplasia</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Neurologic</strong></td>
<td></td>
</tr>
<tr>
<td>Myopathy</td>
<td>Myasthenia gravis</td>
</tr>
<tr>
<td>Stiff-man syndrome</td>
<td></td>
</tr>
<tr>
<td>Parkinson’s disease</td>
<td></td>
</tr>
<tr>
<td><strong>Other Manifestations</strong></td>
<td></td>
</tr>
<tr>
<td>Dental enamel hypoplasia</td>
<td>Immunoglobulin A deficiency</td>
</tr>
<tr>
<td>Keratopathy</td>
<td>Serositis</td>
</tr>
<tr>
<td>Tympnic membrane calcification</td>
<td>Goodpasture’s syndrome</td>
</tr>
<tr>
<td>Vascular calcification</td>
<td>Idiopathic heart block</td>
</tr>
<tr>
<td>Asplenism</td>
<td></td>
</tr>
</tbody>
</table>

*The most common features are shown in bold type.

---

Two arms of the immune system differ fundamentally in their recognition of target antigens. Autoantibodies react with intact molecules (including both soluble and cell-surface molecules) and usually interact with conformational determinants of the autoantigen.

In contrast, T lymphocytes recognize peptide fragments of autoantigens, often 8 to 12 amino acids in length. Furthermore, T cells can recognize peptides only if they are presented on the surface of another cell by major histocompatibility molecules, also known as HLAs. CD4⁺ (helper) T cells react with peptides that are derived from the extracellular fluid and bound by class II histocompatibility molecules (HLA-DP, HLA-DQ, or HLA-DR in humans). Effective presentation requires specialized antigen-presenting cells (APCs) such as macrophages.

---

**TABLE 37-2** – Contrasting Features of the Polyendocrine Syndromes

<table>
<thead>
<tr>
<th>Type I</th>
<th>Type II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autosomal recessive inheritance (only siblings affected)</td>
<td>Polygenic inheritance (multiple generations may be affected)</td>
</tr>
<tr>
<td>No +HLA association (linked to chromosome 21q22.3)</td>
<td>HLA-DR3 and HLA-DR4 associated</td>
</tr>
<tr>
<td>Equal sex incidence</td>
<td>Female preponderance</td>
</tr>
<tr>
<td>Onset in infancy or youth</td>
<td>Peak incidence ages 20 to 60 years</td>
</tr>
<tr>
<td>Mucocutaneous candidiasis</td>
<td>No mucocutaneous candidiasis</td>
</tr>
<tr>
<td>Destructive hypoparathyroidism</td>
<td>Hypoparathyroidism rare (antibody mediated)</td>
</tr>
<tr>
<td>Type 1 diabetes mellitus rare in children (but lifetime frequency 18%)</td>
<td>Type 1 diabetes mellitus common</td>
</tr>
</tbody>
</table>
dendritic cells, and B lymphocytes. Helper T cells secrete various lymphokines, leading to an immune response to the antigen. CD8⁺ (cytotoxic) T cells react with peptides bound by class I histocompatibility molecules (HLA-A, HLA-B, and HLA-C). Class I molecules are present on the surface of nearly all nucleated cells. The antigen peptide in this case is derived from within the presenting cell. Recognition of antigen by viral origin typically leads to the release of cytotoxic chemicals that kill the infected cell.

However, the simple expression of histocompatibility molecules and recognition of antigen by a T cell are not sufficient for T-cell activation. The interaction of major histocompatibility complex (MHC), peptide, and T-cell receptor (called signal one) is critical to the activation process, and other molecules help to define the nature of the immune response (signal two). Co-stimulatory molecules are required for T-cell activation. For example, the cell-surface molecule CD28 engages the CD80 receptor on the T cell and amplifies signal one, which leads to T-cell activation.

Cytotoxic T lymphocyte-associated antigen (CTLA)-4 is a similar receptor but one for which engagement appears to down-regulate T cells. Co-stimulatory signaling includes both cell surfacebound receptors on APCs and secreted substances called cytokines, such as interferon and interleukin-4 (IL-4), which can modify the type of immune response elicited to the same antigen. In the absence of co-stimulatory molecules, engagement of a T-cell receptor leads to inactivation rather than stimulation. Co-stimulatory molecules are expressed or secreted primarily by professional APCs. The hypothesis that expression by endocrine cells of class II molecules is the cause rather than the consequence of autoimmunity (cytokine-induced) has largely been abandoned as the preceding pathway for T-cell stimulation has been unraveled.

The crystal structure of histocompatibility molecules has been elucidated, and these molecules resemble a "hot dog bun" with the antigenic peptide (the hot dog) bound in the groove of the histocompatibility molecule (the bun). Histocompatibility molecules are extremely polymorphic, with different amino acids lining the peptide-binding groove. These variable amino acids determine which peptides are bound and presented to T lymphocytes.

Molecular HLA typing has revealed many subtypes of the older serologically defined alleles, and the unique genetic sequence encoding each polymorphic chain of the histocompatibility molecules is now given a unique identifying number. Thus, for the DQ molecule, which is the histocompatibility molecule most strongly associated with endocrine autoimmunity, a number is given for each alpha and beta chain. Examples are DQA1*0501 for the alpha chain and DQB1*0201 for the beta chain of the DQ molecule commonly encoded on DR3 (DRB1*0301) haplotypes. This DQ molecule (also termed DQ2) is strongly associated with type 1A diabetes, Addison's disease, Graves' disease, and celiac disease. Each allele with its amino acid sequence is inherited in mendelian fashion.

On activation, CD8⁺ T cells can directly lyse cells. CD4⁺ cells are also capable (without CD8⁺ T cells) of destroying target cells. CD4⁺ T cells apparently kill through indirect pathways that include the induction of macrophages to produce cytokines and free radicals. CD4⁺ T lymphocytes are also essential for activation and maturation of B lymphocytes, which produce autoantibodies.

As a simplification, there are two major pathways of CD4⁺ T-cell activation, with subsets of helper T cells, termed Th₁ and Th₂. Th₁ cells produce proinflammatory cytokines such as interferon-γ, IL-2, and tumor necrosis factor-α. Th₂ cells produce lymphokines that suppress Th₁ cells and favor antibody production (e.g., IL-4, IL-5, IL-10, transforming growth factor-β). Thus, depending on the context of T-cell stimulation, T-cell activation may actually down-regulate rather than promote autoimmunity. The Th₁/Th₂ pathway is probably important for the induction of mucosal tolerance from either the oral or nasal route and may be initiated by other forms of antigen vaccination such as the use of altered peptide ligands that can suppress autoimmunity. Such types of therapy are being studied for the prevention and reversal of several autoimmune disorders, including type 1A diabetes, multiple sclerosis, and uveitis.

The natural history of autoimmune disorders can be divided into a series of stages beginning with genetic susceptibility, followed by triggering of autoimmunity (e.g., dietary gliadin exposure in celiac disease), active autoimmunity preceding clinical manifestations (e.g., progressive glanular destruction), and finally overt disease. An etiologic classification of autoimmunity based on initiating factors can be developed, illustrating the many ways autoimmunity, even to a single organ, can be initiated (Table 37-3). For example, myasthenia gravis has a druginduced form (penicillamine), an oncogenic form (thymoma-associated), and the most common idiopathic form.
AUTOLUMINANE POLYENDOCRINE SYNDROME TYPE II

Clinical Definition

APS-II is the more common of the immune endocrinopathy syndromes. It occurs more often in females than in males, frequently has its onset in adulthood, and exhibits familial aggregation. APS-II is usually defined by the occurrence in the same individual of two or more of the following: primary adrenal insufficiency (Addison's disease) \(\text{Fig. 37-1}\), Graves' disease, autoimmune thyroiditis, type 1A diabetes mellitus, primary hypogonadism, myasthenia gravis, or celiac disease. Vitiligo, alopecia, serositis, and pemphigus anemia also occur with increased frequency in individuals with this syndrome and their family members.

The diagnosis of APS-II can be confirmed when one of the component disorders is present; an associated disorder occurs more commonly than in the general population. Furthermore, circulating organ-specific autoantibodies are often present even in the absence of overt clinical disease. In our assessment of APS-II families with Addison's disease, we have noted that up to 15% of relatives have 21-hydroxylase autoantibodies (Addison's disease), anti-islet autoantibodies, or transglutaminase (TG) autoantibodies (celiac disease). In a study of 10 families with APS-II, one in seven relatives had unsuspected illness, most commonly autoimmune thyroid disease. \(\text{11}\) The initial lesion and precipitating events that result in the syndrome are unknown, but immunogenetic and immunologic similarities are present with regard to both the time course and the pathogenesis of each of the component disorders.

The chronic development of organ-specific autoimmunity necessitates endocrinologic evaluation over time of patients with the syndrome and their families. In a family in which the syndrome has been documented, relatives should be advised of the early symptoms and signs of the principal component diseases. A list is available at www.barbaradaviscenter.org. Relatives of patients with multiple disorders should have a medical history, physical examination, and screening every 3 to 5 years with measurement of anti-islet autoantibodies, a sensitive thyrotropin assay, measurement of serum vitamin B \(_6\) levels, and, if there are any symptoms or signs or 21-hydroxylase autoantibodies \(\text{12}\) that suggest adrenal insufficiency are present, annual assay of basal corticotropin and corticotropin-stimulated cortisol levels.

Among 224 patients with Addison's disease and APS-II reported by Neufeld and colleagues, \(\text{12}\) type 1 diabetes (52%) and autoimmune thyroid disease (69%) were the most common coexisting conditions. Other components were less common, including vitiligo (5%) and gonadal failure (4%). CD8+ , CTLA4+ , intercellular adhesion molecule-positive (ICAM+) , and HLA-DR+ cells representing activated cytotoxic T cells have been found at the sites of disappearing melanocytes, suggesting their involvement in the process of destruction that leads to loss of pigmentation in vitiligo lesions. \(\text{11}\) One target antigen for the antibody response noted in this disorder is tyrosinase. \(\text{12}\) Other reports have questioned whether this is the correct or dominant antigen and suggested that a protein comigrating with tyrosinase is the antigen of interest. \(\text{12}\)

Among patients with type 1A diabetes, thyroid autoimmunity and celiac disease coexist with sufficient frequency to justify screening. Thyroid peroxidase autoantibodies are present in 10% of children with type 1 diabetes, \(\text{12}\) and this frequency increases with age. However, thyroid autoantibodies are common without progression to overt disease in the absence of subclinical elevations of thyrotropin. Thus, annual screening of patients with type 1 diabetes with determination of thyrotropin levels is recommended as a cost-effective approach.

Previous reports have suggested that 2% to 3% of patients with type 1 diabetes have celiac disease. \(\text{12}\) With the identification of tissue transglutaminase (TG) as the major endomysial autoantigen of celiac disease, radioimmunoassays were developed and demonstrated that 10% to 12% of patients with type 1 diabetes have TG autoantibodies. \(\text{12}\) The prevalence of TG autoantibodies was higher in diabetic patients with HLA-DQ2; one third of DQ2-homozygous subjects were found to express anti-TG antibody. Seventy percent of those with high-titer antibody who underwent biopsy were subsequently found to have disease. \(\text{12}\) Therefore, screening with anti-TG antibody can be carried out; if the results are positive, small bowel biopsy to document celiac disease is warranted, with institution of a gluten-free diet if the disease is present. Many individuals have asymptomatic celiac disease that may nevertheless be associated with osteopenia and impaired growth. \(\text{12}\) Untreated, symptomatic celiac disease is also associated with an increased risk of gastrointestinal malignancy, especially lymphoma.

Hypoparathyroidism frequently occurs in APS-I but is rare in APS-II. If hypocalcemia occurs in a patient with the type II syndrome, celiac disease may be a more likely diagnosis than primary hypoparathyroidism. Nevertheless, we have described several elderly patients with APS-II who had a distinct form of hypoparathyroidism that, on the basis of a small series of patients, may be termed \text{geriatric hypoparathyroidism}. \(\text{12}\) These patients form a distinct group because they have antibodies to the surface of parathyroid cells capable of suppressing parathyroid function and have a self-limited course of antibody presence and hypoparathyroidism.
Immunogenetics

Although there is familial aggregation of APS-II and its component disorders, there is no clearly discernible pattern of inheritance. Susceptibility is probably determined by multiple genetic loci (with HLA having the strongest effect) interacting with environmental factors. In the case of celiac disease, a dietary precipitating antigen (gliadin) has been identified. For type 1 diabetes, the concordance of identical twins is approximately 50%, suggesting a possible role for environmental or other nongenetic factors, such as somatic mutation or the random rearrangement of T-cell receptors that occurs during the development of the immune system. We found that diabetes developed in identical twins much more often if it arose in the proband before 25 years of age but that initially discordant twins can become diabetic even after a prolonged period of discordance. Furthermore, approximately 40% of long-term discordant twins (>7 years) have persistent autoantibodies or loss of first-phase insulin release, or both. Identical twins may also be discordant for Addison's disease as well as type 1 diabetes.

Many of the disorders of APS-II are associated with an HLA extended haplotype formed by HLA-A1, B8, DR3, DQA1*0501, DQB1*0201 and HLA-DRA. DQA1*0301, DQB1*0302. These include Graves' disease, atrophic thyroiditis, type 1A diabetes (also DR4-associated), Addison's disease (also DR4-associated), myasthenia gravis, and celiac disease. Figure 37-2 illustrates a family in which seven members have type 1 diabetes and three have Addison's disease. The HLA alleles help to explain the high frequency of autoimmunity in this family. The father is homozygous for the high-risk DR3 haplotype that was inherited by six of his eight children.

For some disorders, the complete HLA haplotype is associated with disease; for celiac disease, the most specific association is with the two chains of the DR molecule. Celiac disease occurs primarily in individuals expressing DQA1*0501, DQB1*0201, either in cis (with both of these alleles from the preceding extended DR3 haplotype on the same chromosome) or in trans (with DR5 with DQA1*0501, DQB1*0301 from a DR5 haplotype on one chromosome 6 and DQA1*0201, DQB1*0201 from DR7 on the other chromosome 6). There are more than 100 genes within the MHC on the short arm of chromosome 6, including genes that influence the processing and transport of antigenic peptides. For celiac disease, it appears that the HLA contribution to disease may be limited to the DR5 molecule (DQA1*0501, DQB1*0201). In contrast, for type 1A diabetes and several other component disorders of APS-II, HLA alleles of genes such as A2, A24, and DPB1 may contribute to susceptibility. The MHC class I chain related (MIC-A) gene encodes an unusual class I molecule in the MHC. A specific allele of the MIC-A gene, 5.1, is strongly associated with Addison's disease.

Whereas some HLA alleles increase disease risk, others are associated with protection from disease. For example, the DR alleles DQA1*0102, DQB1*0602 (usually associated with DR2) confer strong protection from type 1A diabetes in a dominant fashion but confer susceptibility to another autoimmune disorder, multiple sclerosis. Furthermore, this protection appears to be organ specific because no protection from Addison's disease is afforded by DQB1*0602.

For both type 1A diabetes and Addison's disease in APS-II, the highest risk is conferred by heterozygosity for the DR3 (DQA1*0501, DQB1*0201) and DR4 (DQA1*0301, DQB1*0302) haplotypes. The DRB1*0404 subtype is found more frequently in those with familial Addison's disease and so is consistent with APS-II. Approximately 2% to 3% of the United States population carries the high-risk DR3-DR4 haplotype, compared with about 35% of individuals who have type 1A diabetes and about 60% of patients in the United States with Addison's disease.

Given the hypothesis that HLA molecules may predominantly determine the tissue targeting of autoimmunity (although other genetic loci may predispose to autoimmunity in general), there has been an intensive effort to identify non-HLA loci. Unlike the situation in APS-I (see later), multiple genetic loci are probably involved in APS-II. For type 1A diabetes, polymorphisms of the insulin gene region contribute to disease susceptibility. There is some evidence of linkage to several loci, including insulin-dependent diabetes mellitus 17 (IDDM17), which was found on chromosome 10 and conferred a 40% risk for diabetes in combination with HLA genes in the Bedouin Arab family in which it was identified. Other loci have been described, including a locus on chromosome 6 (6q) termed IDDM15 that has been confirmed in several studies. Analysis of mutations in the IDDM15 (see later) indicates that it does not play a role in type 1A diabetes or sporadic Addison's disease, with 1 of 90 (1.1%) patients with Addison's disease (non-APS-I) and 1 of 576 (0.2%) control subjects having IDDM15 mutations. The G allele at the CTLA-4 A/G polymorphism was associated with Addison's disease, especially in patients with APS-II and in those with Graves' disease.

Several disorders of the polyendocrine autoimmune syndrome are not associated with DR3. These disorders include pemphigus anemia, goitrous thyroiditis, and vitiligo. Vitiligo is a classic component of the syndrome and was included in Addison's original description. These relatively common disorders may have more than one pathogenic mechanism, one of which is associated with polyendocrine autoimmunity. For example, a polymorphism106base pair (bp) allele: (AT)
Organ-Specific Autoantibodies

Improved assays for several organ-specific autoantibodies have been developed with the cloning of specific autoantigens and the development of assays that use recombinant antigens. These radioimmunoassays are superior to assays based on immunofluorescence with tissue sections, such as ICA testing. The most notable finding is the identification of a large number of different autoantigens that are targeted even in single autoimmune disorders. Most of the endocrine autoantigens are hormones (such as insulin) or enzymes associated with differentiated endocrine function: thyroid peroxidase in thyroiditis, glutamic acid decarboxylase, carboxypeptidase H, and ICA512/IA-2 in type 1 diabetes; 17-hydroxylase and 21-hydroxylase in Addison’s disease (Fig. 37-3) and the parietal cell enzyme H+/K+ -adenosine triphosphatase in pernicious anemia.

In type 1 diabetes, the four most informative assays currently available determine autoantibodies that react with insulin, GAD65 (glutamic acid decarboxylase), ICA512/IA-2, and ICA512/IA-2. In a similar manner, a radioassay format for the detection of autoantibodies that react with the enzyme 21-hydroxylase in Addison’s disease has been developed and provides excellent disease specificity and sensitivity. Adrenal autoantibodies reacting with recombinant 21-hydroxylase usually precede the development of Addison’s disease. However, as with thyroid autoantibodies, there may be individuals who present with antibodies but have normal production of cortisol in response to adrenocorticotropic hormone (ACTH). Continued rescreening every year initially and then biyearly is indicated in this situation.

In contrast to the autoimmune polyendocrine disorders with T cell-mediated glandular destruction, autoantibodies may also be pathogenic. A hallmark of pathogenic autoantibodies is the existence of a neonatal form of the disorder, secondary to transplacental passage of the autoantibody. Examples include neonatal Graves’ disease (antithyrotropin receptor autoantibodies) and neonatal myasthenia gravis (antiacetyl choline receptor autoantibodies).
Thyroxine therapy can precipitate a life-threatening Addisonian crisis in a patient with untreated adrenal insufficiency and hypothyroidism. Thus, it is necessary to evaluate adrenal function in all hypothyroid patients in whom the syndrome is suspected before the institution of such therapy.

Because of the autoimmune nature of these disorders, several studies have evaluated immunosuppressive drugs. Such studies have contributed to our understanding of autoimmunity, and drugs such as cyclosporine have preserved some residual insulin secretion. However, the nephrotoxicity and potential for oncogenicity of cyclosporine have precluded its more generalized use. Newer immunosuppressive agents (mycophenolate mofetil or sirolimus) and biologics such as anti-IL-2 receptor (daclizumab) or "nonmitogenic" CD3 antibodies, for example, hOKT3(ala-ala), may be more effective and have a better safety profile, and these are currently in clinical trials in patients with new-onset diabetes. With further research, immunomodulatory therapies have been developed that are remarkably effective in preventing type 1A diabetes of the BioBreeding (BB) rat and nonobese diabetic (NOD) mouse models of type 1 diabetes.

There are also relatively benign therapies, such as administration of high doses of the vitamin nicotinamide or administration of immune adjuvants that appear to interfere with selected effector pathogenic mechanisms. Small trials of bacille Calmette-Guérin (BCG) vaccine and nicotinamide treatment suggest that these agents have little if any effect in humans, but large randomized trials of nicotinamide are under way in Europe that should define its potential within the next 2 years.

Mucosal administration of antigens is frequently associated with bystander immunosuppression, in which T cells specific to the antigen are apparently induced to produce suppressive T_{H}2-like or T_{H}3-like cytokines (e.g., IL-4, IL-10, and transforming growth factor). In addition, subcutaneous administration of insulin prevents diabetes and insulits in animal models, and subcutaneous administration of insulin peptides in adjuvants can prevent diabetes but not insulits. When such animal trials were extended to humans, oral insulin at diabetes onset had no effect but a small pilot trial suggested that a combination of daily subcutaneous insulin and intermittent intravenous insulin may delay the development of type 1A diabetes in humans. A large National Institutes of Health trial, the Diabetes Prevention Trial-Type I (DPT-I), is under way in the United States to test this hypothesis directly. DPT-I has two arms: (1) intravenous-subcutaneous for those at high risk and (2) oral insulin for those at moderate risk. Subcutaneous insulin injection did not slow progression to diabetes, and results of the oral trial should be known by 2003.

Hashizume and co-workers reported a lower relapse rate in women with Graves' disease who were given thyroxine at 100 µg daily than in those who did not receive thyroxine. Takasu and colleagues reported that autoimmune thyroiditis may be reversible with the disappearance of autoantibodies and maintenance of a euthyroid state in a minority of patients after the cessation of thyroxine therapy. However, subsequent trials have not confirmed such protection, suggesting that other factors may influence the occurrence of disease (iodine intake, for example). In the BB rat model, the administration of thyroxine resulted in a reduced frequency of appearance of lymphocytic thyroiditis. In prediabetic Addison's disease, a short course of glucocorticoids appeared to suppress the expression of adrenal autoantibodies and prevent progressive adrenal destruction.

De Bellis and co-workers screened patients with organ-specific autoimmune disorders and found that 0.9% were positive for adrenal autoantibodies. Three patients with high-titer adrenal autoantibodies and an impaired cortisol response to ACTH experienced remission and had negative results on assay for adrenal autoantibodies when treated with corticosteroids for Grave's ophthalmopathy. In contrast, in 11 other patients with high-titer adrenal autoantibodies, the antibodies persisted and various abnormalities developed, including elevated plasma renin activity, impaired cortisol response to ACTH, elevated ACTH, and overt Addison's disease. Feedback inhibition of endocrine gland function may decrease the exposure of autoantigens to the immune system or decrease the susceptibility of the targeted tissue to immune attack. This preliminary observation requires direct testing in a larger population and in randomized fashion.

Thyroxine therapy can precipitate a life-threatening Addisonian crisis in a patient with untreated adrenal insufficiency and hypothyroidism. Thus, it is necessary to evaluate adrenal function in all hypothyroid patients in whom the syndrome is suspected before the institution of such therapy.

A decreasing insulin requirement in a patient with insulin-dependent diabetes mellitus can be one of the earliest indications of adrenal insufficiency, occurring before the development of hyperpigmentation or electrolyte abnormalities.

In patients with both adrenal insufficiency and primary hypothyroidism, thyroid function may improve after glucocorticoid replacement. Addisonian crisis that responds to mineralocorticoid therapy can occur in patients receiving high-dose glucocorticoids for inflammatory disease. Immuneologic therapies, especially in patients with an autoimmune disease, may induce autoimmunity. A remarkable example is the treatment of patients with multiple sclerosis with an anti-CD52 monoclonal antibody. One third of 27 patients given the monoclonal developed antithyrotropin receptor autoantibodies and hyperthyroidism. Interferon therapy for hepatitis has been associated with thyroid autoimmunity and potentially type 1 diabetes.

![Figure 37-4](image4) Adrenal antibody (AA) titers and levels of adrenocorticotropic hormone (ACTH), cortisol, plasma renin activity (PRA), and aldosterone in three antiladrenal autoantibody-positive patients treated for 6 months with glucocorticoids for concomitant Graves' ophthalmopathy. (From De Bellis A, Bizzaro A, Rossi R, et al. Remission of subclinical adrenocortical failure in subjects with adrenal autoantibodies. J Clin Endocrinol Metab 1993; 76:1021-1037.)

![Figure 37-5](image5) Age at onset of mucocutaneous candidiasis, hypoparathyroidism, and adrenal insufficiency in patients with autoimmune polyendocrine syndrome type I. (From Neufeld M, Maclaren NK, Blizzard RM. Two types of autoimmune Addison's disease associated with different polyglandular autoimmune [PGA] syndromes. Medicine [Baltimore] 1981; 60:355-362.)
AUTOIMMUNE POLYENDOCRINE SYNDROME TYPE 1

Clinical Features

APS-I, also known as autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), is characterized by the classic triad of mucocutaneous candidiasis, autoimmune hypoparathyroidism, and Addison's disease, although the presence of all three is not needed to make the diagnosis and various other manifestations may be present (see Table 37-1). The association of mucocutaneous candidiasis with glandular failure was recognized by Thorpe and Handley in 1929. More than 140 patients have since been reported, including two large series from Finland and the United States.

APS-I is characteristically recognized in early childhood, whereas APS-II has its peak incidence in middle age. Chronic mucocutaneous candidiasis is often the first manifestation, followed by hypoparathyroidism and Addison's disease, but new components can develop at any age. Decades may elapse between the diagnosis of one disorder and the onset of another in the same individual. Consequently, lifelong follow-up is important to allow the early detection of additional components.

In a series of 68 Finnish patients described by Ahonen and co-workers, all had chronic candidiasis at some time, 79% had hypoparathyroidism, 72% had Addison's disease, and 51% had all three of these classical components. Gonadal failure (60% in women, 14% in men) and hypoplasia of the dental enamel (77%) were also frequent findings. Other manifestations that occurred less frequently included alopecia (29%), vitiligo (13%), intestinal malabsorption (18%), type 1 diabetes (18%), pernicious anemia (13%), chronic active hepatitis (12%), and hypothyroidism (4%). The onset of chronic active hepatitis suggested by hepatomegaly, jaundice, or elevated liver enzymes is a serious complication requiring therapy. The presence of chronic candidiasis suggests that a defect in T-cell function may underlie the development of multiple autoimmune disorders in three of these classical components. Gonadal failure (60% in women, 14% in men) and hypoplasia of the dental enamel (77%) were also frequent findings. Other manifestations that occurred less frequently included alopecia (29%), vitiligo (13%), intestinal malabsorption (18%), type 1 diabetes (18%), pernicious anemia (13%), chronic active hepatitis (12%), and hypothyroidism (4%).

Other infections do not occur with such increased frequency. Ectodermal dystrophy is another non-autoimmune part of the syndrome (manifested by pitted nails, keratopathy, and enamel hypoplasia) and cannot be attributed to hypoparathyroidism. Enamel hypoplasia may precede the onset of hypoparathyroidism and, despite adequate replacement therapy, may also affect teeth forming after the onset of hypoparathyroidism. Friedman and colleagues reported the frequent occurrence of asplenia and cholelithiasis as additional features of APS-II. Malabsorption with steatorrhea is of uncertain origin, is usually intermittent, and may be exacerbated by hypocalcemia. Bereket and associates reported a case with patchy intestinal lymphangiectasia discovered by endoscopically directed biopsy. Pancreatic insufficiency has been treated with cyclosporine.

Antiparathyroid and antiladrenal antibodies have been reported. Cross-reactive autoantibodies may play a role in the multigland involvement of these disorders, as noted in several case reports. A target antigen for the autoantibodies found in autoimmune hypoparathyroidism was described by Li and colleagues, who showed that 56% of 25 patients, 17 with APS-I, reacted to the extracellular domain of a membrane-associated antigen of 120 to 140 kd, which was identified as the calcium-sensing receptor. Whereas 21-hydroxylase appears to be the major autoantigen in isolated Addison's disease and in Addison's disease associated with APS-II, autoantibodies against 17-hydroxylase and P450 side-chain cleavage enzyme (CYP11A1) have also been reported in Addison's disease associated with APS-I. Other autoantibodies that may be involved in other components of this disorder have been reported. These include antibodies to tryptophan hydroxylase in intestinal disease, tyrosine hydroxylase in alpoeica areata, and L-amino acid decarboxylase in hepatitis and vitiligo, and phenylalanine hydroxylase and antibodies reacting with hair follicles.

As with APS-II, Tuomi and co-workers originally observed that many more patients (41%) express anti-GAD65 autoantibodies than become diabetic. Among patients with APS-I in this study in whom diabetes developed, GAD autoantibodies could be detected up to 8 years before the onset of overt diabetes. They subsequently noted that many of the antibody-negative patients can demonstrate T-cell responses to GAD65. Nearly 76% of all tested subjects showed either autoantibody or T-cell responses to GAD65 but only 18% had diabetes (8 of 44 subjects), again suggesting that reactivity to this single autoantigen has low predictive value.

### TABLE 37-4 – Mutations of the AIRE (Autoimmune Regulator) Gene

<table>
<thead>
<tr>
<th>Population</th>
<th>Exon 8 13-bp deletion (10941106)</th>
<th>R257X</th>
<th>R139X</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Britain (n = 12)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APS-I</td>
<td>71% (17/24)</td>
<td>4% (1/24)</td>
<td>21% (5/24)</td>
<td></td>
</tr>
<tr>
<td>Normal population</td>
<td>0.2% (1/576)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>United States (n = 16)</td>
<td>50% (16/32)</td>
<td>16% (5/327)</td>
<td></td>
<td>(711/32)</td>
</tr>
<tr>
<td>Sardinia (n = 10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APS-I</td>
<td>5% (1/20)</td>
<td>90% (18/20)</td>
<td>5%</td>
<td></td>
</tr>
<tr>
<td>Normal population</td>
<td>1.7%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northern Italy (n = 9)</td>
<td>27% (5/18)</td>
<td>56% (10/18)</td>
<td>17%</td>
<td></td>
</tr>
</tbody>
</table>

APS-I autoimmune polyendocrine syndrome type I.
Genetics

APS-I is unique among autoimmune endocrine disorders, in that it is not associated with high-risk class II HLA alleles, although the protective allele DQB1*0602 may protect against type 1 diabetes but not Addison's disease.\(^1\)\(^3\) Addison's disease in APS-II is strongly associated with HLA-DR3 and HLA-DR4. APS-I shows an autosomal recessive pattern of inheritance, with a 25% recurrence risk for siblings of affected individuals.\(^8\) The disorder has a high prevalence in Finland and in consanguineous Iranian Jewish families.\(^8\)

The etiologic gene was localized to the short arm of chromosome 21 (near markers D21s49 and D21s171 on 21p22.3) by Aaltonen and co-workers\(^6\) and identified as AIRE. The gene encodes a putative deoxyribonucleic acid (DNA)-binding protein of unknown function expressed in the thymus and in lymphoid and other tissues.\(^7\)\(^8\)\(^9\) Multiple mutations are causative, and the frequency of specific mutations varies in different populations (Table 37-4). For example, in Sardinia, a deletion of amino acid 257 is present in 90% of mutated alleles. A 136-bp deletion in exon 8 is present in 71% of British alleles and 56% of alleles in the United States. Analysis of haplotypes indicates that this deletion has arisen on multiple occasions.
Therapy

Note the following principles:

1. The treatment of adrenal insufficiency and hypoparathyroidism is the same as that discussed in Chapter 14 and Chapter 26, respectively, with the caveat that malabsorption may complicate treatment.

2. The therapy for mucocutaneous candidiasis has been improved with orally active antifungal drugs such as fluconazole \[90\] and ketoconazole \[91\]. Infection often recurs when the drug is discontinued or when the dosage is decreased. Patients must be monitored carefully because ketoconazole may inhibit adrenal and gonadal steroid synthesis and may precipitate adrenal failure. It is also associated with transient elevation of liver enzyme levels and, occasionally, hepatitis. Fluconazole is associated with a lower frequency of hepatiis and does not inhibit steroidogenesis when given in the recommended doses. \[90\]

3. Screening to allow the early detection of new disorders before overt symptoms and signs develop is recommended, including autoantibody studies, electrolytes, calcium and phosphorus levels, thyroid and liver function tests, blood smear, and plasma vitamin B \(_12\) levels. Patients at risk for adrenal failure can be screened by measurement of basal ACTH and supine plasma renin activity (PRA) levels, \[92\] followed by dynamic testing as appropriate. Evaluation for asplenism \[73\] with abdominal ultrasonography and blood smear examination for Howell-Jolly bodies is warranted, with pneumococcal vaccination and appropriate antibiotic coverage for affected patients.

4. Hypocalcemia has been associated with the intermittent steatorrhea characteristic of APS-I, and therapies that restore calcium levels have been beneficial. Treatment includes magnesium replacement for hypomagnesemia. Nevertheless, in individual patients, specific etiologic factors have been implicated, including pancreatic insufficiency, *Giardia lamblia* infection, and lymphangiectasia, and individualized therapy is required for these potentially diverse causes.

5. There are case reports of severely affected patients who have benefited from immunosuppressive therapy. For example, Ward and colleagues \[93\] treated a 13-year-old patient who had keratoconjunctivitis, hepatitis, and severe pancreatic insufficiency. Treatment with cyclosporine was associated with normalization of stool fat (from 31.5 g/day to 2.5 g/day).
OTHER POLYENDOCRINE DEFICIENCY AUTOIMMUNE SYNDROMES

Anti-Insulin Receptor Antibodies

In this rarely reported disorder (25 patients), also known as type B insulin resistance and acanthosis nigricans, insulin resistance is due to the presence of anti-insulin receptor antibodies. Approximately one third of patients with these antibodies have an associated autoimmune illness such as systemic lupus erythematosus (SLE) or Sjögren's syndrome. Arthralgia, vitiligo, alopecia, and secondary amenorrhea have also been reported. One patient had a daughter with hyperthyroidism and a granddaughter with SLE. Autoimmune thyroid disease has been described in two such patients, one with hypothyroidism and the other with antithyroid antibodies. Anti-nuclear antibodies and an elevated erythrocyte sedimentation rate, hyperglobulinemia, leukopenia, and hypocomplementemia are common.

The major clinical manifestations are related to the anti-insulin receptor antibodies. Severe insulin resistance is profound, and up to 175,000 U of insulin given intravenously per day may be ineffective in lowering the elevated glucose level. Despite hyperglycemia and marked insulin resistance, ketoacidosis is uncommon. The course of the diabetes is variable, and several patients have had spontaneous remissions. Other patients have had severe hypoglycemia (perhaps related to the insulin-like effects of anti-insulin receptor antibodies demonstrable in vitro). The acanthosis nigricans, which is due to hypertrophy and folding of otherwise histologically normal skin, appears to be related to the insulin-resistant state. Other forms of marked insulin resistance in the absence of antireceptor antibodies are also associated with acanthosis nigricans.
POEMS Syndrome

The components of the multisystem disorder POEMS (plasma cell dyscrasia with polyneuropathy, organomegaly, endocrinopathy, M protein, and skin changes or Crow-Fukase syndrome) consist of diabetes mellitus (50% of patients), primary gonadal failure (70% of patients), plasma cell dyscrasia, sclerotic bone lesions, and neuropathy. Patients usually present with severe progressive sensorimotor polyneuropathy, hepatosplenomegaly, lymphadenopathy, and hyperpigmentation. On evaluation, they are found to have plasma cell dyscrasia and sclerotic bone lesions.

POEMS is assumed to be secondary to circulating immunoglobulins, but binding of antibody directly to involved tissues has not been demonstrated. There is evidence implicating cytokines such as IL-1, IL-6, and tumor necrosis factor in addition to the M protein in the pathogenesis of this disorder. Several studies have also demonstrated that elevated levels of vascular endothelial growth factor (VEGF) correlate with the disease state and that treatment with immunosuppressive agents reduced both symptoms of the disease and levels of VEGF, suggesting that this growth factor may play a role in the disease. A therapeutic trial of an anti-VEGF antibody would provide more definitive evidence for this hypothesis. The diabetes mellitus responds to small subcutaneous doses of insulin. The hypogonadism is associated with elevated plasma levels of follicle-stimulating hormone and luteinizing hormone. Temporary resolution of disease, including a return of the blood glucose level to normal, may occur after radiotherapy for localized plasma cell lesions of bone.
Kearns-Sayre Syndrome

The rare Kearns-Sayre syndrome, also known as oculocraniosomatic disease or oculocraniosomatic neuromuscular disease with ragged red fibers, is characterized by myopathic abnormalities leading to ophthalmoplegia and progressive weakness in association with several endocrine abnormalities, including hypoparathyroidism, primary gonadal failure, diabetes mellitus, and hypopituitarism. Crystalline mitochondrial inclusions are found in muscle biopsy specimens, and such inclusions have also been observed in the cerebellum. The relationship between the mitochondrial disorders and endocrinologic abnormalities is not known. Other abnormalities include retinitis pigmentosa and heart block.

Antiparathyroid antibodies have not been described; however, antibodies to the anterior pituitary gland and striated muscle have been found, and the disease may have autoimmune components.
Thymic Tumors

The thymus has a central role in the ontogeny of cell-mediated immunity. DiGeorge described congenital aplasia of the thymus and parathyroid glands, both of which are derived from the third and fourth pharyngeal pouches. Affected infants present with tetany secondary to hypocalcemia, severe infections with markedly suppressed T-cell immunity, and normal humoral immunity.

The thymus is a complex tissue with a specialized endocrine epithelium that synthesizes a variety of biologically active peptides involved in the control of T-cell maturation. This epithelium is derived from the neural crest and contains complex gangliosides that react with monoclonal antibody (A2B5) and tetanus toxin in a manner similar to that of pancreatic islets. The role of these biologically active peptides of the thymus has not been defined, but they may be trophic factors in T-cell activation and increase in situations of primary failure of T-cell activation, just as the levels of trophic hormones increase in primary endocrine failure.

The illnesses associated with thymomas are similar to those in APS-II,\[105\] although the frequency of specific disorders is different. In one review of patients with thymoma, myasthenia gravis occurred in 44% of the patients, red blood cell aplasia in approximately 20%, hypoglobulinemia in 6%, autoimmune thyroid disease in 2%, and adrenal insufficiency in 1 of 423 patients. The frequency of autoimmune thyroid disease reported in patients with thymoma is probably an underestimate, given the frequency of unsuspected thyroid disease in patients with myasthenia gravis. Mucocutaneous candidiasis in adults is also associated with thymomas. In most patients, the thymomas are malignant, although temporary remissions of the autoimmune disease can occur with resection of the tumor.
Trisomy 21

Down syndrome, or trisomy 21, is associated with the development of insulin-dependent diabetes mellitus and thyroiditis. We have observed one patient with a partial distal translocation "leading" to trisomy 21 and "associated" with adrenal insufficiency, celiac disease, hypothyroidism, and insulin-dependent diabetes. Patients with trisomy 21 also have T-cell abnormalities, including increased Ia-positive T cells and a premature increase in the 3G5 age-related T-cell subset. It is not known whether the observed chromosomal abnormality influences the development of autoimmunity or whether part of the susceptibility to autoimmunity is associated with chromosomal disorders. Organ-specific autoimmunity also occurs with gonadal dysgenesis.
Congenital Rubella

Patients with congenital rubella have an almost 20% risk of acquiring diabetes mellitus and a higher than normal risk of acquiring thyroiditis and hypothyroidism. Those at highest risk for diabetes express diabetes-associated HLA-DR3 and HLA-DR4 alleles. Rubella appears to be associated with diabetes primarily after fetal infection, and it is not known whether the virus increases the probability of subsequent autoimmunity because it has permanent effects on the developing immune system.

Organ-specific autoimmunity is readily induced in animal models by perturbations of neonatal immune function (neonatal thymectomy, neonatal cyclosporine administration). Congenital diabetes, although rare, may occur as illustrated by a case report describing lymphocytic infiltration of the pancreas with CD45RO+ memory phenotype cells and the presence of autoantibodies to GAD and insulin in a child born with congenital diabetes and without serologic evidence of viral infection.
Wolfram Syndrome

Wolfram syndrome is a rare autosomal recessive disease that is also called DIDMOAD, which stands for diabetes insipidus, diabetes mellitus, progressive bilateral optic atrophy, and sensorineural deafness. In addition, neurologic and psychiatric disturbances are prominent in most patients and may cause severe disability.

Linkage analysis of several families with this disorder has identified a locus of chromosome 4 that was highly associated with the disease. Segregation analysis of the mutations found in familial and sporadic cases of Wolfram syndrome led to the identification of wolframin, or WFS1, a 100-kd transmembrane protein encoded by a gene located at 4p16.1. The disease has been mapped to a locus on the short arm of chromosome 4.

Atrophic changes in the brain have been found with magnetic resonance imaging.

Wolfram syndrome appears to be a slowly progressive neurodegenerative process, and there is also (non-autoimmune) selective destruction of the pancreatic beta cells. This association may be related to the many molecules and metabolic pathways shared by islets and neurons. Diabetes mellitus with an onset in childhood is usually the first manifestation. Diabetes mellitus and optic atrophy are present in all reported cases, but expression of the other features is variable. Linkage to other loci in addition to WFS1 may explain the variability in phenotype seen in this disorder. In one case report, two related children with Wolfram syndrome had megaloblastic and sideroblastic anemia that responded to treatment with thiamine. Furthermore, thiamine treatment was associated with a marked decrease in insulin requirements.
X-Linked Syndrome of Polyendocrinopathy, Immune Dysfunction, and Diarrhea (XPID)

XPID was first described in 1982 and was noted to be an X-linked immunodeficiency syndrome characterized by autoimmune enteropathy, polyendocrinopathy, atopic dermatitis, and fatal infections. Activated T cells can be found in the circulation and in the infiltrates along affected gut tissue. Linkage analysis demonstrated that a 17-cM stretch of the X chromosome is associated with XPID; this site is distinct from the adjacent site, which is associated with the similar Wiskott-Aldrich syndrome. The syndrome consists of loss of pancreatic islets (diabetes mellitus), infection, severe enteropathy, thrombocytopenia and anemia, endocrinopathy and eczema, and growth retardation, and is often fatal in infancy. It differs from Wiskott-Aldrich syndrome in that activated CD4+ T cells are found. Characteristics of the scuffy mouse (gene sf) bear a number of similarities to this syndrome, and abnormalities in the sf gene lead to abnormalities in the amount and function of scurfin, a DNA-binding protein. Wildin and colleagues studied four individuals with XPID for abnormalities in the sf gene and found that each had either nonconservative substitutions or deletion or insertion mutations that could affect function of this protein. This study indicates that this disorder is tied to abnormalities in the sf gene.
Summary

The most basic pathogenic lesion of the polyendocrine autoimmune syndromes is an inherited tendency to target self-molecules immunologically. The disease associations and the inheritance pattern make it possible to detect additional components of these syndromes in patients before the appearance of serious manifestations and to make the diagnosis in some first-degree relatives with unrecognized disease. Detection and diagnosis can now be facilitated by autoantibody assays with excellent sensitivity and specificity.
References


---


Endocrine tumors originating from islet, or enteroendocrine, cells may present with unique clinical symptoms that reflect the biologic actions of secreted peptide hormones. In this chapter, we discuss how endocrine cell lineages develop during organogenesis in the endocrine pancreas and intestine and review the biologic actions of peptide hormones produced in pancreatic and intestinal endocrine cells and enteric nerves. Although numerous physiologic actions of these peptides are still poorly understood and under active investigation, excessive production of one or more of these peptides frequently accounts for the clinical symptoms attributable to endocrine tumors arising from the gastrointestinal tract and pancreas.
ENDOCRINE CELL DEVELOPMENT IN THE PANCREAS

The endocrine and exocrine pancreas develop from the primitive foregut endoderm. Pancreatic morphogenesis is a complex process that begins with the evagination of the embryonic foregut into ventral and dorsal buds at 28 days' gestation in humans and at embryonic day (ED) 8 in mice. Upon rotation of the stomach and duodenum during development results in simultaneous rotation of the ventral bud that undergoes fusion with the dorsal bud to give rise to the primitive pancreas. The ventral bud develops into the posterior portion of the pancreatic head, including the uncinate process, while the remaining pancreas derives from the dorsal bud. In mice, a complex tree-like, epithelial-lined ductal system develops within the pancreatic diverticula with glucagon immunoreactive cells detected as early as ED 9.5, followed by detection of cells containing insulin at ED 10.5. Stem cells that give rise to both terminally differentiated endocrine and exocrine acinar cells are thought to reside within the ductal epithelium.

In humans, islet formation begins at gestational week (GW) 12 with the aggregation of polyclonal endocrine cells. Between GWs 13 and 16, small aggregates of endocrine cells arise from the pancreatic duct and develop their own blood supply. At GWs 17 to 20, fewer islets are observed in contact with the ducts, and a mantle of non-beta endocrine cells forms around the beta cells. Between GWs 21 and 26, a continual increase in the proportion of islet tissue and in the average size of the islets is observed with occasional non-beta cells in the center of the islet, a morphologic appearance that is characteristic of the postnatal islet.

At birth, the endocrine pancreas accounts for 1% to 2% of the entire pancreatic cell mass. The neuroendocrine marker nestin appears to be expressed on islet and ductal cells that exhibit properties consistent with human islet stem cells in vitro. Although genetic studies in mice have yielded valuable insights into the ontology of islet development, the relative order of appearance of unique populations of hormone-producing islet endocrine cells differs in humans and mice. Somatostatin- and pancreatic polypeptide (PP)-positive cells are detected at GW 7 in the human pancreas scattered among ductal cells. One week later, glucagon cells appear, and by

<table>
<thead>
<tr>
<th>TABLE 38-1</th>
<th>Effects of Disrupting Genes on Pancreatic Endocrine Cell Development</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene</td>
<td>Phenotype in Homozygous (I) Mutant Mice</td>
</tr>
<tr>
<td>Hes-1</td>
<td>Increased glucagon-positive alpha cells, pancreatic hypoplasia</td>
</tr>
<tr>
<td>Nkx2-2</td>
<td>Dorsal lobe agenesis, small islets, reduced beta cells</td>
</tr>
<tr>
<td>Isl-1</td>
<td>Loss of differentiated islet cells</td>
</tr>
<tr>
<td>Nkx6.2</td>
<td>Absent mature beta cells, reduced alpha and pancreatic polypeptide cells</td>
</tr>
<tr>
<td>NeuroD</td>
<td>Reduced beta cells, arrested islet morphogenesis</td>
</tr>
<tr>
<td>ngn3</td>
<td>Absent islet cells and defective enterodocrine cell formation</td>
</tr>
<tr>
<td>Pax4</td>
<td>Absent islet beta and delta cells</td>
</tr>
<tr>
<td>Pax6</td>
<td>Absent alpha cells</td>
</tr>
<tr>
<td>Pdx-1</td>
<td>Pancreatic agenesis</td>
</tr>
</tbody>
</table>

GW 9 to 10, insulin-producing cells are detectable. In mice, both insulin- and glucagon-expressing cells are first detected between ED 9.5 and 10.5, whereas somatostatin and PP are expressed later, by ED 15.5. Although cells coexpressing insulin and glucagon are detected during early islet development, cell lineage studies employing specific transgenes that mark or ablate islet cell precursors suggest that the alpha and beta cell lineages arise independently during ontogeny in the mouse. Peptide YY (PYY) co-localizes with each of the four main islet hormones in the developing pancreas. However, genetic evidence for an essential role of a PYY-producing precursor cell in pancreatic endocrine development has not yet been forthcoming.

Delineation of the genetic determinants that regulate the developmental formation and organization of pancreatic endocrine cell populations has been facilitated by studies of mice with disruption of candidate regulatory genes, principally islet transcription factors (Table 38-1). The homeobox transcription factor Pdx-1 is required for insulin gene transcription in the adult beta cell and for developmental formation of the entire pancreas. Mice homozygous for a null mutation in Pdx-1 fail to develop a pancreas, whereas restricted inactivation of Pdx-1 in the murine beta cell produces insulin deficiency and diabetes mellitus. Similarly, pancreatic agenesis has also been reported in human subjects homozygous for a loss of function Pdx-1 mutation. Subjects heterozygous for Pdx-1 develop a form of maturity-onset diabetes mellitus of the young (MODY4).

Targeted disruption of the LIM domain Isl-1 gene in mice results in abnormal development of the dorsal pancreatic mesenchyme and abnormal differentiation of islet cells, whereas a heterozygous human ISL-1 mutation has been reported in a single patient with type 2 diabetes mellitus. Although mutations in the Pax4 and Pax6 genes produce profound abnormalities in developmental formation of murine pancreatic endocrine cells, islet function has not been extensively studied in human subjects with PAX mutations. Nevertheless, binding sites for the MODY genes Pdx-1, HNF1, and HNF4 have been identified in the Pax4 promoter, suggesting that MODY genes may be upstream regulators of genes critical for islet cell formation and islet function in the pancreas.

Genes encoding members of the notch receptor family, their ligands, and downstream targets are essential for developmental formation of the endocrine pancreas. Mice lacking neurogenin 3 (ngn3), a basic helix-loop-helix (bHLH) transcription factor, fail to develop pancreatic endocrine cells and die of diabetes mellitus postnatally, whereas overexpression of ngn3 produces accelerated differentiation of pancreatic endocrine cells. These findings, taken together with the loss of Isl-1, Pdx-1, Pdx-6, NeuroD expression in ngn3 / mice, implicate ngn3 as a key upstream regulator of pancreatic endocrine cell development.

Research into the identification of upstream control mechanisms and downstream targets that promote islet cell formation, growth, and differentiation is likely to proceed rapidly in the next few years, providing scientists and clinicians with an enhanced understanding of the genetic determinants regulating the growth of normal and neoplastic endocrine cells. A summary of genetic mutations associated with abnormal formation of pancreatic endocrine cells is provided.
ENDOCRINE CELL DEVELOPMENT IN THE INTESTINE

Stem cells associated with the intestinal epithelium differentiate into four different cell lineages: (1) enterocytes, (2) Paneth cells, (3) goblet cells, and (4) enteroendocrine cells.

The enteroendocrine cell population comprises less than 1% of all intestinal epithelial cells but represents the largest mass of endocrine cells in the body. Compared with studies of pancreatic enteroendocrine cell development, much less is known about the molecular control of enteroendocrine cell formation and differentiation.

Numerous enteroendocrine cell types have been identified that can be classified according to morphologic criteria and expression of one or more secretory products. In the stomach, gastrin cells first appear in the duodenum, followed by their localization to the antrum and pylorus in adult gastric mucosa. In the small bowel, a secretin precursor cell appears important for enteroendocrine cell lineage formation. In the murine colon, PYY is the first detectable hormone marking appearance of enteroendocrine cells and is coexpressed in most enteroendocrine cells in the large intestine as they first differentiate.

The notch signaling pathway is essential for developmental formation of enteroendocrine cells. Activation of notch results in increased expression of the bHLH transcriptional repressor Hes1 that functionally antagonizes bHLH genes that regulate cellular differentiation. Mice deficient in Hes1 demonstrate premature cellular differentiation and severe pancreatic hypoplasia due to depletion of pancreatic epithelial precursors. These mice also demonstrate excessive differentiation of multiple endocrine cell types in the developing stomach and gut, suggesting that Hes1 is a negative regulator of endodermal endocrine differentiation. Both Notch1 and ngn3 act upstream of BETA2/NeuroD, a bHLH protein important for differentiation of endocrine cells in both the pancreas and intestine (Table 38-2).

Mice homozygous for a null mutation in the Pdx-1 gene demonstrate poorly differentiated duodenal intestinal epithelium with absence of Brunner’s glands and a deficiency of gastrin cells in the stomach. Just distal to the abnormal epithelium, a reduction in the number of enteroendocrine cells is observed. In contrast, expression of Pdx-1 in gut epithelial cells redirects cell lineage toward an enteroendocrine phenotype. Inactivation of BETA2/NeuroD in mice results in absence of secretin-producing and cholecystokinin (CCK)-producing enteroendocrine cells. The complexity of lineage relationships between gut endocrine cell populations is further illustrated by studies in mice with targeted ablation of secretin-producing cells. These mice exhibit nearly complete elimination of enteroendocrine cell populations producing CCK and PYY/glucagon and a reduction in cells producing gastric inhibitory polypeptide (GIP), somatostatin, and serotonin.

Members of the Pax gene family are also essential for the formation of enteroendocrine cells (see Table 38-2). Targeted disruption of Pax4 markedly reduces the number of murine duodenal cells immunopositive for serotonin, secretin, GIP, PYY, and CCK and decreases the number of somatostatin- and serotonin-positive cells in the stomach. Complete disruption of the Pax6 locus more selectively reduces the number of duodenal cells expressing GIP and CCK and decreases the number of gastrin-immunopositive and somatostatin-immunopositive cells in the stomach, whereas SEV Pax6 mice that express a dominant negative mutant Pax6 allele demonstrate markedly reduced levels of proglucagon messenger RNA (mRNA) transcripts in both the small and large intestine, with almost complete depletion of enteroendocrine cells exhibiting glucagon-like peptide 1 (GLP-1) and GLP-2 immunoreactivity (Fig. 38-1; see also Color Plate).

At present, the classification of enteroendocrine cells is based principally on the phenotype ascribed to the production of one or more peptide hormones. Nevertheless, it seems likely that additional enteroendocrine cell subpopulations will be described in different regions of the gut that exhibit considerable biologic complexity beyond that which is currently appreciated.
AMYLIN

Amylin, also known as islet amyloid-associated peptide, is a 37-amino acid hormone produced in islet beta cells and in scattered endocrine cells in the stomach and in the proximal small intestine. Exogenous administration of amylin inhibits gastric emptying and glucagon secretion in rodents and humans. Excess amylin secretion and deposition in the endocrine pancreas have been implicated as a potential pathogenic feature in some subjects with type 2 diabetes mellitus. Amylin exerts its physiologic actions through interaction with the calcitonin receptor in the presence of a receptor activity-modifying protein (RAMP). Mice deficient in amylin display modest perturbations in islet function and enhanced glucose clearance following glucose challenge.

The role of gut-derived amylin in human physiology has not been clearly established. Although amylin expression has been detected in both pancreatic and gut endocrine tumors, a specific syndrome attributable to amylin overexpression has not been delineated.
Calcitonin Gene-Related Peptide

Calcitonin gene-related peptide (CGRP) is a member of a larger family of peptides that includes calcitonin, amylin, and adrenomedullin. In humans, distinct genes CALC-A and CALC-B encode for both calcitonin and CGRP and give rise to two 37 amino acid C-terminal amidated neuropeptides, designated \(-\text{CGRP}\) and \(-\text{CGRP}\). These neuropeptides share considerable amino acid sequence homology differing by only three amino acids in humans. \(-\text{CGRP}\) is expressed predominantly in primary afferent sensory neurons arising from the spinal cord, whereas \(-\text{CGRP}\) is expressed in enteric neurons. Two calcitonin/CGRP seven-transmembrane domain G-protein coupled receptors both interact with a family of RAMPs; coexpression of calcitonin receptor-like receptor with RAMP1 results in ligand specificity for CGRP, whereas expression of the same receptor with RAMP2 results in specificity for adrenomedullin.

CGRP immunoreactivity has been localized to enteroendocrine cells of the human rectum and to endocrine cells and neurons in the small intestine. Intestinal CGRP is released in response to glucose and by gastric acid secretion. CGRP produces marked vasodilation in the stomach, splanchnic, and peripheral circulation through stimulating nitric oxide (NO) release. CGRP also inhibits gastric acid and pancreatic exocrine secretion likely through stimulating somatostatin release. Although focal CGRP positivity has been detected in some human carcinoid and pancreatic endocrine tumors (PETs), its usefulness as a tumor marker has not been firmly established.
Cholecystokinin

Cholecystokinin (CCK) was first characterized as a factor that stimulates gallbladder contraction. The CCK gene is expressed in "open-type" enteroendocrine I cells in the proximal

<table>
<thead>
<tr>
<th>Hormone</th>
<th>Cell/Tissue of Origin</th>
<th>Related Peptides</th>
<th>Actions</th>
<th>Secretory Stimuli</th>
</tr>
</thead>
</table>
| Amylin                       | Pancreatic B cell, endocrine cells of stomach and small intestine                     | Calcitonin, calcitonin generelated peptide, adrenomedulin                         | 1. Inhibits gastric emptying  
2. Inhibits arginine-stimulated and postprandial glucagon secretion  
3. Inhibits insulin secretion  
4. Satiety factor                                                                 | 1. Co-secreted with insulin in response to oral nutrient ingestion                  |
| Calcitonin generelated peptide (CGRP) | -CGRP is expressed predominantly in afferent sensory nerves from the spinal cord; -CGRP is expressed in enteric neurons and enteroendocrine cells of the rectum | Calcitonin, amylin, adrenomedulin                                                 | 1. Produces marked vasodilatation in the splanchnic and peripheral circulation by stimulating nitric oxide release  
2. Inhibits gastric acid and pancreatic exocrine secretion  
3. Induces intestinal smooth muscle relaxation                                                                 | 1. Glucose and gastric acid secretion                                                |
| Cholecystokinin (CCK)        | Enteroendocrine I cells and enteric nerves, central nervous system, pituitary corticotrophs, C cells of the thyroid, adrenal medulla, and the acrosome of developing and mature spermatozoa |                                                                               | 1. Inhibits proximal gastric motility while increasing antral and pyloric contractions  
2. Regulates meal-stimulated pancreatic enzyme secretion and gallbladder contraction  
3. Trophic effects on pancreatic acini in rats  
4. Postprandial satiety  
5. In the brain, CCK affects memory, sleep, sexual behavior, and anxiety               | 1. Oral nutrient ingestion  
2. Several intestine-derived hormones, including GRP and bombesin  
3. Activation of -adrenergic receptors                                                                 | |
| Galanin                      | Central and peripheral nervous systems, pituitary, neural structures of the gut, pancreas, thyroid, and adrenal gland |                                                                               | 1. In the brain, regulation of food intake, memory and cognition, and antinociception  
2. Inhibits pancreatic exocrine secretion and intestinal ion transport  
3. Induces both contraction and relaxation of intestinal smooth muscle, depending on the species examined  
4. Delays gastric emptying and prolongs colonic transit times  
5. Inhibits the secretion of insulin, PYY, gastrin, somatostatin, enteroglucagon, neurtensin, and pancreatic polypeptide | 1. Intestinal distention  
2. Chemical stimulation of the intestinal mucosa  
3. Electrical stimulation of periarterial nerves  
4. Extrinsic sympathetic neurons                                                                 | |
| Gastric inhibitory polypeptide/glucose-dependent insulinitropic polypeptide (GIP) | Neuroendocrine K cells in the duodenum and proximal jejunum                      |                                                                               | 1. Inhibits gastric acid secretion and gastrointestinal motility  
2. Increases insulin release and regulates glucose and lipid metabolism  
3. Exerts anabolic actions in bone                                                                 | 1. Oral nutrient ingestion, especially long-chain fatty acids  
2. Humoral and neural influences, including the vagus nerve, -adrenergic and -aminobutyric acid neurons, and gastrin-releasing peptide                                                                 | |
| Gastrin                      | Predominantly enteroendocrine G cells of the stomach and duodenal bulb; central and peripheral nervous systems, pituitary, adrenal gland, genitl tract, respiratory tract, fetal pancreas |                                                                               | 1. Induces gastric acid secretion  
2. Amidated gastrins are trophic to the oxyntic mucosa of the stomach and gastrin induce colonic epithelial proliferation  
3. Pregastrin and glycine-extended gastrin induce colonic epithelial proliferation | 1. Luminal contents, especially partially digested aromatic amino acids, small peptides, calcium, coffee, and ethanol  
2. Humoral and neural influences, including the vagus nerve, -adrenergic and -aminobutyric acid neurons, and gastrin-releasing peptide  
3. Somatostatin inhibits secretion                                                                 |
<table>
<thead>
<tr>
<th>Peptide</th>
<th>Source and Targets</th>
<th>Functions</th>
</tr>
</thead>
</table>
| Gastrin-releasing peptide (GRP) and related peptides | Central nervous system, enteric nervous system; reproductive tract, and lung, where it acts as a neurotransmitter. GRP neurons also distributed throughout the human pancreas | 1. Stimulates smooth muscle contraction in the stomach, intestine, and gallbladder  
2. Stimulates the release of CCK, gastrin, GIP, glucagon, GLP-1, GLP-2, motilin, PP, PYY, and somatostatin  
3. Stimulates gastric acid secretion via direct effect on G cells  
4. In the brain, regulates appetite, memory, thermogenesis, and cardiac function  
5. Stimulates pancreatic growth  
6. In the lung, growth factor for both normal and neoplastic tissue |
| Bombesin, neuromedin B, neuromedin C | | 1. Cholinergic stimulation |
| Ghrelin | Central nervous system, stomach, small intestine, and colon | Motilin 1. Stimulates growth hormone release  
2. Stimulates gastric kinetic activity  
3. Orexigenic activity  
4. Stimulates energy production and signals hypothalamic regulatory nuclei that control energy homeostasis  
5. Stimulates pancreatic growth  
6. In the lung, growth factor for both normal and neoplastic tissue |
| Glucagon | Pancreatic A cell, central nervous system | 1. Primary counterregulatory mechanism to restore plasma glucose levels in the setting of hypoglycemia by increasing gluconeogenesis, glycogenolysis, and protein-lipid flux in both the liver and periphery  
2. Gastrointestinal smooth muscle relaxation  
3. Neural and humoral factors released in response to hypoglycemia  
4. Gastrointestinal smooth muscle relaxation  
5. Inhibition of food intake |
| Glucagon-like peptide 1 (GLP-1) | Enteroneocline L cells located in the ileum and colon, central nervous system | 1. Enhances glucose disposal following nutrient ingestion by inhibiting gastric emptying, stimulating insulin secretion, and inhibiting glucagon secretion  
2. Inhibits food intake  
3. Stimulates pancreatic islet neogenesis and proliferation  
4. Inhibits sham feeding-induced gastric acid secretion |
| Glucagon-like peptide-2 (GLP-2) | As described above for GLP-1 | 1. Induces small intestinal and colonic mucosal growth by stimulating crypt cell proliferation and inhibiting apoptosis  
2. Inhibits centrally induced antral motility and meal-stimulated gastric acid secretion  
3. Enhances intestinal epithelial barrier function  
4. Stimulates intestinal hexose transport  
5. Inhibits short-term control of food intake |
| Motilin | Brain, bronchoepithelial cells, and enteroendocrine M cells located in the duodenum and proximal jejunum | Ghrelin 1. Induces phase III contractions in the stomach  
2. Stimulates gastric and pancreatic enzyme secretion  
3. Induces contraction of the gallbladder, sphincter of Oddi, and lower esophageal sphincter  |
| Neuropeptide Y (NPY) | Central and peripheral nervous systems, pancreatic islet cells | PYY and PP 1. Potent stimulator of oral nutrient intake  
2. Activation of the sympathetic nervous system  
3. Reduces gastrointestinal fluid and electrolyte secretion  
4. Inhibits gastric and small intestinal motility  
5. Induces marked vasoconstriction of the splanchnic circulation |
<table>
<thead>
<tr>
<th>Neurotensin (NT)</th>
<th>N cells located in the small intestinal mucosa, especially the ileum; central and peripheral nervous systems, including the enteric nervous system; heart, adrenal gland, pancreas, and respiratory tract</th>
<th>Neureomodulin N, xenin, and xenopsin</th>
<th>1. Stimulates growth of the colonic epithelium</th>
<th>1. Luminal nutrients, especially lipids, but not amino acids or carbohydrates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>2. Inhibits postprandial gastric acid secretion and pancreatic exocrine secretion</td>
<td>2.GRP and bombesin</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. Stimulates colonic motility but inhibits gastric and small intestinal motility</td>
<td>3. Somatostatin inhibits secretion</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4. Facilitates fatty acid uptake in the proximal small intestine and induces histamine release from mast cells</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5. Trophic in some pancreatic and colon cancer cell lines in vitro</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6. In the brain, neuromodulator of dopamine transmission and anterior pituitary hormone secretion, hypothermia, analgesic effects; reduces food intake</td>
<td></td>
</tr>
<tr>
<td>Pancreatic polypeptide (PP)</td>
<td>Major site of expression is pancreatic endocrine cells located in periphery of islets in pancreatic head and uncinate process</td>
<td>NPY and PYY</td>
<td>1. Reduces CCK-induced gastric acid secretion</td>
<td>1. Stimulated by nutrients, hormones, neurotransmitters, gastric distention, insulin-induced hypoglycemia, and direct vagal nerve stimulation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2. Increases intestinal transit times by reducing gastric emptying and upper intestinal motility</td>
<td>2. Hyperglycemia, bombesin, and somatostatin inhibit secretion</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. Inhibits postprandial exocrine pancreas secretion via a vagal-dependent pathway</td>
<td></td>
</tr>
<tr>
<td>Peptide YY (PYY)</td>
<td>Enteroendocrine cells, developing endocrine pancreas, subpopulation of pancreatic A cells in mature islets</td>
<td>NPY and PP</td>
<td>1. Entero gastrone inhibits both gastric acid secretion and gastric motility</td>
<td>1. Following oral nutrient ingestion, early secretion is mediated by the vagus nerve and hormonal influences; subsequently, secretion occurs as a result of direct L-cell stimulation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2. Increases intestinal transit time by reducing intestinal motility</td>
<td>2. Bile acids and fatty acids</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. Inhibits pancreatic exocrine secretion</td>
<td>3. Amino acids administered intracolically</td>
</tr>
<tr>
<td>Pituitary adenylate cyclase activating peptide (PACAP)</td>
<td>Brain, respiratory tract, and enteric nervous system</td>
<td>Vasoactive intestinal peptide (VIP), peptide histidine isoleucine (PHI), and peptide histidine methionine (PHM)</td>
<td>1. Stimulates histamine release from the stomach</td>
<td>1. Activation of the central nervous system</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2. Increases the secretion of pancreatic fluid, protein, and bicarbonate</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. Stimulates insulin and cate-cholamine release</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4. Neural regulation of gastric acid secretion</td>
<td></td>
</tr>
<tr>
<td>Secretin</td>
<td>Central nervous system, fetal endocrine pancreas, and enteroendocrine S cells located in the duodenum and proximal jejunum</td>
<td></td>
<td>1. Principal hormonal stimulant of pancreatic and biliary bicarbonate and water secretion</td>
<td>1. Gastric acid, bile salts, and luminal nutrients, especially fatty acids, peptides, and ethanol</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2. Regulates pancreatic enzyme secretion</td>
<td>2. Somatostatin inhibits secretion</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. Stimulates gastric secretion of pepsinogen</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4. Inhibits lower esophageal sphincter tone, postprandial gastric emptying, gastrin release, and gastric acid secretion</td>
<td></td>
</tr>
</tbody>
</table>
Somatostatin
- Central nervous system, pancreatic D cells, enteroendocrine D cells
- 1. Inhibits secretion of islet hormones, including insulin, glucagon, and PP
- 2. Inhibits the secretion of gut peptides, including gastrin, secretin, VIP, CCK, GLP-1, and GLP-2
- 3. Inhibits pancreatic exocrine secretion
- 4. Acts in a paracrine manner on G cells, enterochromaffin-like cells, and parietal cells to inhibit gastric acid secretion
- 5. Reduces splanchnic blood flow, intestinal motility, and carbohydrate absorption while increasing water and electrolyte absorption

Tachykinins
- Throughout the central and peripheral nervous systems, including the respiratory tract, skin, sensory organs, and the urogenital tract; in the gastrointestinal tract, neurons localized in the submucous and myenteric plexuses, extrinsic sensory fibers, and enterochromaffin cells in the gut epithelium
- Substance P, neurokinin A, and neurokinin B
- 1. Regulate vasomotor and gastrointestinal smooth muscle contractility
- 2. Chemotaxis and activation of immune cells, mucus secretion, water absorption and secretion
- 3. Role in visceral inflammation, hyperreflexia, and hyperalgesia

Thyrotropin-releasing hormone (TRH)
- Central and enteric nervous system, colon, G cells of the stomach, pancreatic islet beta cells
- 1. Suppresses pentagastrin-stimulated gastric acid secretion
- 2. Chronic administration induces pancreatic hyperplasia and inhibits amylase release
- 3. Attenuates CCK-induced gallbladder smooth muscle contraction
- 4. Inhibits cholesterol synthesis within the intestinal mucosa

Vasoactive intestinal peptide (VIP)
- Widely expressed in the central and peripheral nervous system (including the enteric nervous system)
- PACAP, PHI and PHM
- 1. Induces relaxation of vascular and nonvascular smooth muscle
- 2. Mediates relaxation of the lower esophageal sphincter, sphincter of Oddi, and anal sphincter
- 3. Regulates relaxation-associated gut contraction and may be involved with reflex vasodilation in the small intestine
- 4. Inhibits gastric acid secretion
- 5. Stimulates biliary water, bicarbonate, pancreatic enzyme, and intestinal chloride secretion
- 6. Some evidence suggesting a role in regulating pancreatic release of insulin and glucagon

Small intestine (Table 38-4) and in nerve fibers branching to the gastric and colonic myenteric plexus and submucosal plexus where CCK acts as a neurotransmitter. CCK-immunoreactive peptides are found in the cerebral cortex and limbic system and in pituitary corticotrophs, C cells of the thyroid, adrenal medulla, and the acrosome of the developing and mature spermatozoa. The CCK gene encodes a 94-amino acid prohormone that is post-translationally processed in a tissue-specific fashion into multiple molecular forms of CCK-83, -58, -39, -33, -22, -8, and -5, all sharing a common C-terminus. The major active form, CCK-8, is an octapeptide containing a sulfated tyrosine residue and an amidated C-terminal phenylalanine residue, whereas CCK-33 appears to be the predominant circular form in human plasma.\(^{[29]}\)

CCK binds with high affinity to the CCK-A receptor, a seven-transmembrane domain G-proteincoupled receptor expressed.

### Table 38-4 -- Location of Enteroendocrine Cells and Their Associated Peptide Hormones in the Gastrointestinal Tract

<table>
<thead>
<tr>
<th>Hormones</th>
<th>Enteroendocrine Cell</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Somatostatin</td>
<td>D</td>
<td>Stomach, duodenum, small intestine, colon</td>
</tr>
<tr>
<td>Gastrin, TRH</td>
<td>G</td>
<td>Stomach and duodenum</td>
</tr>
<tr>
<td>CCK</td>
<td>I</td>
<td>Duodenum and jejunum</td>
</tr>
<tr>
<td>GIP</td>
<td>K</td>
<td>Duodenum and proximal jejunum</td>
</tr>
<tr>
<td>GLP-1, GLP-2, PYY</td>
<td>L</td>
<td>Ileum, colon, and rectum</td>
</tr>
<tr>
<td>Motilin</td>
<td>M</td>
<td>Duodenum and proximal jejunum</td>
</tr>
<tr>
<td>Neurotensin</td>
<td>N</td>
<td>Small intestine, especially ileum</td>
</tr>
<tr>
<td>Secretin</td>
<td>S</td>
<td>Duodenum and proximal jejunum</td>
</tr>
</tbody>
</table>

CCK, cholecystokinin; GIP, gastric inhibitory polypeptide; GLP-1, -2, glucagon-like peptide 1, 2; PYY, peptide YY; TRH, thyrotropin-releasing hormone.
in pancreatic acinar cells, gallbladder, smooth muscle, chief and D cells of the gastric mucosa, and the central and peripheral nervous systems. In the stomach, CCK inhibits proximal gastric motility while increasing the force of antral and pyloric contractions. CCK also regulates meal-stimulated pancreatic enzyme secretion and gallbladder contraction.

CCK exhibits trophic effects on pancreatic acini in rats. Experimental manipulations that increase levels of circulating CCK, such as treatment with soybean trypsin inhibitor, or long-term pancreatobiliary diversion result in pancreatic growth and premalignant changes. Elevated circulating levels of CCK also enhanced the development of preneoplastic acinar lesions induced by azaserine, a pancreatic carcinogen in rats. In contrast, the Otsuka Long-Evans Tokushima Fatty (OLETF) rat fails to express a functional CCK-A receptor and demonstrates reduced pancreatic size.

Exogenous administration of CCK decreases the size of spontaneously ingested meals, whereas CCK-A receptor antagonists increase appetite. A human subject with polyglandular syndrome type I exhibited severe diarrhea and malabsorption in association with reduced numbers of enteroendocrine cells and CCK deficiency. Thus, CCK secretion in response to oral nutrient ingestion likely regulates nutrient absorption and postprandial satiety. Nevertheless, CCK receptors do not appear essential for weight regulation in vivo because mice with targeted disruption of the CCK-A and CCK-B receptors exhibit normal food intake and weight gain well into adult life.
Galanin was initially isolated from porcine intestine as a 29-amino acid C-terminally amidated neuropeptide. In humans, two molecular forms of galanin exist that are 19 and 30 amino acids in length. Galanin is expressed in the central and peripheral nervous systems and pituitary gland and in neural structures of the gut, pancreas, thyroid, and adrenal gland. In the intestine, galanin immunoreactivity is detected predominantly within enteric neurons located in the myenteric and submucosal plexuses that innervate the mucosa and the circular and longitudinal smooth muscle layer. Galanin is released by enteric neurons in response to intestinal distention, chemical stimulation of the mucosa, electrical stimulation of periarterial nerves, and extrinsic sympathetic neurons.

At least three different galanin receptor subtypes have been identified: GalR1, GalR2, and GalR3. These subtypes are widely expressed in gastric and intestinal smooth muscle cells, pancreas, and the central nervous system. The actions of galanin include regulation of food intake, memory and cognition, antinociception, and modulation of multiple neuroendocrine systems in the pituitary, pancreas, and gut. The importance of galanin for pituitary lactotroph biology is exemplified by studies of galanin knockout mice that exhibit normal growth rates but reduced levels of prolactin and complete failure of lactation.

Although galanin can inhibit GIP-induced and GLP-1-induced proinsulin gene transcription and insulin secretion, infusion of galanin in humans has no effect on levels of plasma insulin. Galanin also inhibits both pancreatic exocrine secretion and intestinal ion transport and induces the contraction and relaxation of intestinal smooth muscle. In humans, intravenous administration of galanin delays gastric emptying and prolongs colonic transit times. Although galanin expression has been detected in hypothalamic, pituitary, and adrenal tumors, galanin immunopositivity in pancreatic or gut endocrine tumor cells is rare.
Gastric Inhibitory Polypeptide

Also known as glucose-dependent insulinotropic polypeptide, GIP is a 42-amino acid peptide secreted by enteroendocrine K cells located in the duodenum and proximal jejunum. GIP levels rise immediately following nutrient ingestion, leading to modest inhibitory effects on gastric acid secretion and gastrointestinal motility.

The precise role of GIP as an enterogastrone remains controversial because supraphysiologic concentrations of GIP are required to inhibit both gastric acid secretion and gastric emptying in humans. The actions of GIP on the pancreatic beta cell are primarily those of an incretin, a gut-derived peptide that stimulates insulin secretion in the setting of raised plasma glucose levels following oral nutrient ingestion. GIP receptor knockout mice are viable but exhibit impaired oral glucose tolerance and enhanced susceptibility to diabetes mellitus following high-fat feeding. Although GIP-secreting endocrine tumors are rare, gut-derived GIP may contribute to the development of food-induced Cushing’s syndrome in a subset of patients with adrenal adenomas that express the GIP receptor.
Gastrin

A single mRNA transcript encodes a pre-progastrin precursor of 101 amino acids that undergoes post-translational processing into multiple biologically active molecular forms of circulating gastrin, including G34, G17, and G14. Gastrin is produced predominantly in G cells located in the gastric antrum and duodenal bulb; however, gastrin immunoreactivity has also been detected in the central and peripheral nervous systems, pituitary, adrenal gland, genital tract, and respiratory tract and in tumors. The fetal endocrine pancreas produces large amounts of amidated gastrin, suggesting a possible role of gastrin in pancreatic development. However, gastrin-deficient mice do not demonstrate overt abnormalities in pancreatic islet morphology. A possible role for gastrin in human islet biology derives from studies of the CCK-2 receptor on pancreatic A cells, which secrete glucagon in response to gastrin in vitro.

G cells are open-type endocrine cells subject to regulation by luminal contents in addition to humoral and neural influences. The effects of gastrin on acid secretion are mediated by the fully processed amidated forms of gastrin (G17 and G34) at the CCK-2 receptor (formerly known as the CCK-B/gastrin receptor) located on the enterochromaffin-like (ECL) cells of the oxyntic mucosa. Gastrin stimulates histamine synthesis and release from ECL cells, which then induce acid secretion by binding to the histamine 2 (H₂) receptor located on the basolateral aspect of the parietal cell. Gastrin also stimulates acid secretion from parietal cells via the CCK-2 receptor.

The physiologic roles of progastrin and glycine-extended gastrin (G-Gly) are less completely defined but may involve regulation of the growth and differentiation of the gastrointestinal tract. Amidated gastrin is trophic to the oxyntic mucosa of the stomach, where it stimulates proliferation of gastric stem cells and ECL cells, resulting in increased parietal and ECL mass. G-Gly exerts trophic effects on the colonic mucosa and mucosal thickness and are more prone to formation of aberrant crypt foci following treatment with azoxymethane, whereas inactivation of the gastrin gene results in reduced basal rates of colonic proliferation.

Gastrin has been reported to induce proliferation of colon cancer cell lines expressing the CCK-2 receptor; however, most colon cancers and normal colonic epithelium do not normally express the CCK-2 receptor. A truncated gastrin-binding receptor has been described in some colon cancer cell lines, and a constitutively active CCK-2 receptor mutant that confers ligand-independent growth to transfected cells has been identified in human colorectal cancers. The trophic effects of gastrin have led to studies of gastrin-neutralizing antisera for the potential treatment of intestinal neoplasia.
Gastrin-Releasing Peptide and Related Peptides

The bombesin family of peptides was originally isolated from frog skin and includes bombesin, gastrin-releasing peptide (GRP, the mammalian homologue of bombesin), neuromedin B (NMB), and neuromedin C (NMC). Gastrin-releasing peptide is a 27-amino acid peptide, whereas both NMB and NMC are decapetides. These peptides share an identical C-terminal amidated heptapeptide sequence that is essential for biologic activity. GRP is expressed in the central, peripheral, and enteric nervous systems, reproductive tract, and lung, where it acts as a neurotransmitter. NMB is expressed predominantly in the brain and gastrointestinal tract. Within the intestine, GRP and NMB are localized to neurons in the submucosal and myenteric plexuses of the stomach, small intestine, and colon. GRP-containing neurons are also distributed throughout the human pancreas. Bombesin and GRP stimulate smooth muscle cell contraction in the stomach, intestine, and gallbladder. GRP stimulates the release of CCK, gastrin, GIP, glucagon, GLP-1 and GLP-2, motilin, PP, PYY, and somatostatin in some, but not all, species.

Three GRP receptor subtypes have been cloned that are seven-transmembrane domain G-protein-coupled receptors that bind bombesin-like peptides, including a GRP-preferring subtype (expressed throughout the intestine), NMB-preferring subtype (expressed in the esophageal and intestinal muscularis), and a third subtype, designated bombesin receptor subtype 3, which preferentially binds GRP over NMB and is expressed in testes and small cell lung cancer. GRP regulates appetite, memory, and thermoregulation and suppresses appetite following intracerebroventricular or systemic administration. GRP stimulates pancreatic growth in part via a CCK-dependent mechanism. The expression of GRP in human tumors with neuroendocrine properties such as small cell carcinoma and medullary thyroid carcinoma, taken together with its autocrine and endocrine effects on cell growth, suggests that GRP may contribute to regulation of tumor cell growth.
Ghrelin

Ghrelin, a motilin-related peptide, is a 28-amino acid growth hormone-releasing factor originally purified from rat stomach that stimulates growth hormone release via the growth hormone secretagogue receptor (GHS-R). Fasting increases gastric ghrelin gene expression, and ghrelin exhibits gastric prokinetic activity and orexigenic activity following both intracerebroventricular and peripheral administration via the ghrelin receptor expressed in hypothalamic nuclei. Ghrelin expression is also induced by stressors, and ghrelin may play a role in the anxiogenic stress response in a corticotropin-releasing hormone (CRH)-dependent manner in mice.

The majority of rat and human gut endocrine cells that express ghrelin are localized to the stomach, with a small number of ghrelin-positive cells identified in the small and large intestine. GHS-R is also expressed in the gut; however, the function of the intestinal ghrelin/GHS-R axis remains poorly understood. Circulating levels of ghrelin in human subjects increase and fall before and after food ingestion, consistent with a role for ghrelin in appetite regulation.
Glucagon, Glucagon-like Peptide 1, and Glucagon-like Peptide 2

The proglucagon gene is expressed in the pancreatic A cell, intestinal L cell, and specialized regions of the brain, primarily neurons in the brain stem and, to a lesser extent, hypothalamus. In mammals, a single proglucagon precursor is differentially processed to yield multiple proglucagon-derived peptides (PGDPs), including glucagon in the islet A cell, and glicentin, oxyntomodulin, glucagon-like peptide 1 (GLP-1), GLP-2, and several spacer or intervening peptides in the gut enteroendocrine L cell.

Pancreatic glucagon is a 29-amino acid peptide that regulates plasma glucose levels via effects on gluconeogenesis and glycogenolysis. Increased glucagon secretion functions as the primary counterregulatory mechanism to restore normal levels of plasma glucose in the setting of hypoglycemia. In contrast,

**Figure 38-2** Glucagon-like peptide II receptor (GLP-2R) expression in subsets of endocrine cells in the human stomach (ST) and large bowel (LB). Most cells exhibiting positivity with antisera against the human GLP-2R also exhibited immunopositivity for an endocrine marker such as chromogranin (CHROM). In contrast, most endocrine cells in the stomach and both small and large intestine did not express the GLP-2R. Arrows denote cells positive for both the GLP-2R and chromogranin, and arrowheads denote cells positive for the GLP-2R or chromogranin.

GLP-1 secreted from the gut endocrine cell enhances glucose disposal following nutrient ingestion by inhibition of gastric emptying, stimulation of insulin secretion, and inhibition of glucagon secretion. GLP-1 also inhibits food intake and stimulates pancreatic islet neogenesis and proliferation, biologic actions that facilitate long-term control of nutrient homeostasis.

GLP-2 is a 33-amino acid peptide that is co-secreted with GLP-1, oxyntomodulin, and glicentin from enteroendocrine cells in a nutrient-dependent manner. GLP-2 inhibits both centrally induced antral motility and meal-stimulated gastric acid secretion. GLP-2 exhibits trophic actions in the small intestine and colon via stimulation of crypt cell proliferation and reduction of apoptosis within the crypt and villus compartments. GLP-2 also exerts actions independent of intestinal growth, including enhancement of intestinal epithelial barrier function and stimulation of intestinal hexose transport. The beneficial therapeutic actions of GLP-2 in experimental models of intestinal injury and in human subjects with short bowel syndrome suggest that GLP-2 may be useful for preventing injury and enhancing repair and regeneration in the gastrointestinal epithelium.

Although the actions of GLP-1 and GLP-2 are transduced via distinct GLP-1 and GLP-2 receptors, both GLP-1 and GLP-2 are rapidly inactivated by the same enzyme, dipeptidyl peptidase IV (DPP-IV). The GLP-1 receptor is widely expressed in GLP-1 target tissues, and disruption of the GLP-1 receptor in mice produces mild glucose intolerance and betacell dysfunction. The GLP-2 receptor is expressed in subsets of enteroendocrine cells; hence, many of the gastrointestinal actions of GLP-2 are likely indirect and mediated via enteroendocrine-derived factors. The coexpression of the GLP-2 receptor in minor subsets of distinct enteroendocrine subpopulations in the stomach and small and large intestine further illustrates the phenotypic complexity in defining subsets of functionally unique enteroendocrine cells (Fig. 38-2) (see also Color Plate). The GLP-2 receptor is also expressed in the central nervous system, where both GLP-1 and GLP-2 regulate the short-term control of food intake.

In contrast with the biologic actions of GLP-1 and GLP-2, those of glicentin and oxyntomodulin are less well established. Glicentin appears trophic for the gut mucosal epithelium, whereas oxyntomodulin inhibits food intake and pentagastrin-stimulated gastric acid secretion both in vitro and in vivo. Although distinct G protein-coupled receptors (GPCRs) for glucagon, GLP-1, and GLP-2 have been characterized, separate receptors that mediate the actions of glicentin and oxyntomodulin have not yet been identified.
Motilin

Motilin is a 22-amino acid peptide originally isolated from porcine intestine. Motilin immunoreactivity has been detected in open-type enteroeendocrine epithelial M cells located predominantly in the duodenum and proximal jejunum. Secretion of motilin occurs in a cyclical manner during the interdigestive state between meals. The presence of nutrients in the duodenum suppresses the endogenous release of motilin in both dogs and humans. Duodenal alkalinization, sham feeding, gastric distention, and administration of opioid agonists promote motilin secretion. A putative motilin receptor has been cloned that exhibits 52% amino acid identity with the human receptor for growth hormone secretagogues. The motilin receptor is expressed in multiple regions of the gastrointestinal tract, predominantly in smooth muscle and enteric neurons, and also recognizes the macrolide antibiotic erythromycin.

Motilin induces phase 3 contractions in the stomach, an effect that can be abolished by food ingestion, duodenal acidification, somatostatin, pentagastrin, and CCK. Atropine and 5-hydroxytryptamine antagonists also abolish phase 3 contractions, emphasizing the importance of the cholinergic and serotoninergic neuronal pathways. Motilin stimulates gastric and pancreatic enzyme secretion and induces contraction of the gallbladder, sphincter of Oddi, and lower esophageal sphincter.
Neuropeptide Y

Neuropeptide Y (NPY) is primarily synthesized and secreted by neurons in the central and peripheral nervous systems. In the brain, NPY is expressed not only in the hypothalamus, where it exhibits extremely potent effects on nutrient intake, but also in the cortex, hippocampus, basal forebrain striation, limbic structures, amygdala, and brain stem. In the peripheral nervous system, expression occurs predominantly in sympathetic neurons and in the myenteric and submucous plexuses of the enteric nervous system.

NPY and vasoactive intestinal peptide (VIP) are often coexpressed in enteric neurons. NPY is synthesized in and released from pancreatic islet cells and inhibits glucose-stimulated insulin secretion via the Y1 receptor. Elevated circulating NPY levels are observed following sympathetic nervous system activation and in patients with pancreatic endocrine tumors, carcinoid tumors, and neurogenic tumors, including neuroblastomas and pheochromocytomas.

Elevated circulating NPY levels are observed following sympathetic nervous system activation and in patients with pancreatic endocrine tumors, carcinoid tumors, and neurogenic tumors, including neuroblastomas and pheochromocytomas. Intravascular administration of NPY is associated with marked vasoconstriction of the splanchnic circulation, an effect that is not altered by -adrenergic or -adrenergic blockade.

NPY exerts its actions via several receptor subtypes, including the Y1 and Y2 receptors that bind NPY and PYY with similar affinities and the Y3 receptor that exhibits a preference for NPY over PYY. NPY and PYY are targets for N-terminal degradation by the enzyme DP-IV, leading to the generation of NPY (336) and PYY (336), peptides that exhibit preferential binding to the Y2 receptor. In the gastrointestinal tract, NPY reduces fluid and electrolyte secretion and inhibits both gastric and small intestinal motility.
Neurotensin

Neurotensin (NT) is a 13-amino acid peptide originally detected in bovine hypothalamus. NT-related peptides include neuromedin N, a 6-amino acid NT-like peptide co-encoded in proneurotensin, as well as xenin, and xenopsin. In the gastrointestinal tract, NT processing favors the generation of NT in N cells of the ileum and in enteric neurons. NT is also produced in the central and peripheral nervous systems, heart, adrenal gland, pancreas, and respiratory tract. NT secretion is stimulated by luminal nutrients, especially lipids, but not amino acids or carbohydrates. GRP also stimulates NT release, whereas somatostatin exerts an inhibitory effect.

At least three different NT receptor/binding proteins (NTS-1 to NTS-3) have been identified. NTS-1 and NTS-2 belong to the GPCR family, whereas NTS-3 represents a structurally unrelated protein with NT-binding properties. NTS-1 is expressed in both the brain and intestine, whereas NTS-2 and NTS-3 are expressed exclusively in the brain.

NT administration to rats augments the adaptive response to small bowel resection in the intestinal remnant, and NT stimulates growth of the colonic epithelium in vivo. NT also inhibits postprandial gastric acid secretion and pancreatic exocrine secretion, stimulates colonic motility, and inhibits gastric and small intestinal motility. NT facilitates fatty acid uptake in the proximal small intestine and induces histamine release from mast cells. NT receptor expression has been detected in a subset of human colon and pancreatic ductal cancers, and NT is trophic for some pancreatic and colon cancer cells in vitro.
Pancreatic Polypeptide

Pancreatic polypeptide (PP) was isolated from chicken pancreatic extracts as a by-product of insulin purification. The majority of PP is expressed in pancreatic endocrine cells located predominantly in the periphery of islets in the pancreatic head and uncinate process. Elevated plasma levels of PP have been detected in patients with gastrointestinal endocrine tumors; hence, PP may be used as a tumor marker in appropriate clinical scenarios. Nutrients, hormones, neurotransmitters, gastric distention, insulin-induced hypoglycemia, and direct vagal nerve stimulation regulate PP secretion, whereas hyperglycemia, bombesin, and somatostatin inhibit PP secretion.

The actions of PP are mediated by the Y4 receptor, a GPCR coupled to the inhibition of cyclic adenosine monophosphate accumulation. The human Y4 receptor is expressed in the stomach, small intestine, colon, pancreas, prostate, and enteric nervous system and in select central nervous system neurons. Exogenous administration of PP reduces CCK-induced gastric acid secretion and increases intestinal transit times by reducing gastric emptying and upper intestinal motility. PP also inhibits postprandial exocrine pancreas secretion via a vagal-dependent pathway. Transgenic mice that overexpress PP exhibit reduced weight gain and rate of gastric emptying and decreased fat mass. The biologic actions of PP in the gastrointestinal tract and pancreas are in part centrally mediated, because intracisternal injections of PP cause an increase in gastric acid secretion and gastric motility and a reduction in pancreatic secretion.
Peptide YY

Peptide YY (PYY), together with NPY, and PP are members of the PP family. These peptides consist of 36 amino acids, contain several tyrosine residues, and share considerable amino acid identity with amidated C-terminal ends. Although these peptides likely share a common ancestry, they exhibit unique actions and patterns of tissue-specific expression, with PYY and PP acting as hormones, whereas NPY acts primarily as a neurotransmitter.

PYY is expressed in the fetal and adult gastrointestinal tract in enteroendocrine cells. Distinct enterodendocrine subpopulations have been identified that express PYY alone, or both PYY and GLP-1, in the ileum, colon, and rectum. Immunoreactive PYY has also been detected in the developing endocrine pancreas and in a subpopulation of glucagon-producing A cells in mature islets. PYY is secreted as a 36-amino acid peptide and circulates as two molecular forms: PYY (136) and an N-terminally truncated form, PYY (336). Luminal nutrients, CCK, GRP, and vagal tone regulate PYY secretion.

PYY exerts its actions in part through the NPY Y1 and Y2 receptors and a separate PYY receptor. Whereas PYY (136) binds both Y1 and Y2 receptors, PYY (336) is selective for the Y2 receptor. PYY demonstrates inhibitory effects on gastrointestinal secretion, motility, and blood flow. In the stomach, PYY functions as an enterogastrone, inhibiting both gastric acid secretion and gastric emptying. PYY also increases intestinal transit times by inhibiting small and large intestinal motility. The role of PYY as an intestinal epithelial growth factor remains unclear because some, but not all, studies demonstrate an intestinotrophic effect of PYY in rodents. In the pancreas, both PYY (136) and PYY (336) inhibit pancreatic exocrine secretion.
Pituitary Adenylate Cyclase-Activating Peptide

Pituitary adenylate cyclase-activating peptide (PACAP), VIP, and GRF are structurally related members of the glucagon-secretin superfamily. PACAP-immunoreactive nerve fibers are distributed along the gastrointestinal tract from the esophagus to the colon. Both PACAP-38 and PACAP-27 are detected in many tissues, with PACAP-38 generally the predominant peptide. PACAP stimulates histamine release from the stomach; increases the secretion of pancreatic fluid, protein, and bicarbonate; and stimulates insulin secretion and catecholamine release. PACAP signaling in gastric ECL cells may also constitute an important component of the neural regulation of gastric acid secretion.

Three PACAP receptors (PAC1, VPAC1, and VPAC2) have been cloned and bind PACAP and VIP with varying affinities. Consistent with the putative importance of PACAP for islet function, PAC1 receptor knockout mice exhibit defective glucose-stimulated insulin secretion.
Secretin

Secretin is a 27-amino acid peptide synthesized predominantly in the brain and gastrointestinal tract. In the gut, secretin is produced by the enteroendocrine S cell in the duodenum and proximal jejunum. Gastric acid, bile salts, and luminal nutrients stimulate, and somatostatin inhibits, the release of secretin. Secretin stimulates pancreatic and biliary bicarbonate and water secretion and may regulate pancreatic enzyme secretion. Secretin also stimulates the gastric secretion of pepsinogen and inhibits lower esophageal sphincter tone, postprandial gastric emptying, gastrin release, and gastric acid secretion.

Although secretin is expressed in the fetal endocrine pancreas, its function in islet biology remains uncertain. To date, only a single secretin receptor has been isolated and characterized. Secretin has been proposed as a treatment for autism; however, clinical trial results examining this issue have not been consistently positive.
Somatostatin

Somatostatin, originally isolated as a hypothalamic growth hormone release inhibiting factor, is also expressed in the intestine and pancreas. Post-translational processing of prosomatostatin results in the generation of SS-14 and SS-28, biologically active peptides corresponding to the C-terminal 14 and 28 amino acids of prosomatostatin. SS-28 is the predominant molecular form liberated by enteroeendocrine D cells, whereas SS-14 is the predominant species liberated by D cells in the stomach and pancreas.

Five somatostatin receptor subtypes (SST-1 to SST-5) have been identified that are expressed in a tissue-specific manner. Somatostatin’s actions are generally inhibitory; that is, somatostatin inhibits the secretion of growth hormone and thyrotropin in the pituitary, and insulin, glucagon, and PP in the endocrine pancreas. In the gastrointestinal tract, somatostatin inhibits the secretion of a broad range of gut peptides. Somatostatin inhibits pancreatic exocrine secretion and also acts in a paracrine manner on G cells, ECL cells, and parietal cells to inhibit gastric acid secretion.

The inhibitory properties of somatostatin make it suitable for the treatment of conditions characterized by excess hormone secretion. Although the circulating half-life of native somatostatin is short, longer-acting synthetic somatostatin analogues such as octreotide and lanreotide are useful in the treatment of neuroendocrine tumors, acromegaly, and portal hypertension. Both octreotide and lanreotide are octapeptides that bind the SST-2 and SST-5 somatostatin receptor subtypes, receptors commonly expressed in neuroendocrine tumors. Somatostatin analogues are also employed for the treatment of portal hypertension and gastrointestinal bleeding. Tumor-associated somatostatin receptor expression forms the basis for the radiolabeled octreotide scan, a test that appears useful for the detection of a broad spectrum of human neoplasms. Somatostatin-deficient mice exhibit normal growth but defects in sexually dimorphic hepatic gene expression.
Tachykinins

The family of tachykinins includes substance P (SP), neurokinin-A (NKA), and neurokinin-B (NKB), all of which share a common C-terminal pentapeptide sequence essential for biologic action. Two genes encode the tachykinins: a pre-protachykinin-A gene that encodes SP and NKA and a preprotachykinin-B gene that encodes NKB. Tachykinins are synthesized within neurons localized to the submucous and myenteric plexuses, extrinsic sensory fibers, and in enterochromaffin cells in the gut epithelium. Tachykinins are also widely distributed throughout the central and peripheral nervous systems, the respiratory tract, skin, sensory organs, and the urogenital tract.

Four different tachykinin receptors (NK1 to NK4) have been cloned and bind tachykinin peptides with different affinities. NK1 receptors preferentially bind SP, NK2 preferentially binds NKA, and both NK3 and NK4 preferentially bind NKB.

The tachykinins regulate vasomotor and gastrointestinal motor activity. The ability of tachykinins to induce vasodilatation or vasoconstriction appears to be specific to the species and to the vascular bed. Tachykinins exhibit both direct and indirect effects on intestinal smooth muscle contractile activity.

Activation of NK1 receptors on the interstitial cells of Cajal and NK2 receptors on intestinal smooth muscle cells directly promotes peristalsis, whereas activation of NK3 receptors on enteric neurons exerts a prokinetic effect that is indirectly mediated through cholinergic stimulation of enteric smooth muscle cells. The NK1 and NK3 receptors can exhibit inhibitory effects on intestinal motility by inducing the release of inhibitory molecules such as NO and VIP from inhibitory neurons. NK2 receptors can also inhibit intestinal motility by either stimulation of sympathetic ganglia or activation of nonadrenergic inhibitory mechanisms. NK2 receptor antagonists reduce or prevent trinitrobenzene sulfonic acid induced weight loss and intestinal injury, whereas an NK1 receptor antagonist exhibits protective effects in acetic acid induced colitis.

Tachykinins are commonly produced by gut carcinoids and may be responsible for mediating some of the clinical manifestations associated with these tumors.
Thyrotropin-Releasing Hormone

Originally isolated as a hypothalamic regulatory peptide, thyrotropin-releasing hormone (TRH) is expressed throughout the gastrointestinal tract, including the stomach, colon, and pancreas. In the pancreas, TRH is most abundantly expressed during perinatal development. Pre-pro-TRH is synthesized by islet beta cells, G cells in the stomach, and neurons comprising the myenteric plexus of the esophagus, stomach, and intestine. In the stomach, histamine and serotonin stimulate, and endogenous opioids inhibit, TRH release.

TRH acts via two related GPCRs: TRH receptor 1 (TRHR1) and TRH receptor 2 (TRHR2). TRH suppresses pentagastrin-stimulated gastric acid secretion, and chronic administration of TRH induces pancreatic hyperplasia and inhibits amylase release. TRH also attenuates CCK-induced gallbladder smooth muscle contraction and inhibits cholesterol synthesis within the intestinal mucosa.
Vasoactive Intestinal Peptide

Vasoactive intestinal peptide (VIP) is a 28-amino acid member of a peptide superfamily that includes PACAP, peptide histidine isoleucine, and peptide histidine methionine, all neurotransmitters and neuromodulators of the enteric nervous system. The VIP gene is widely expressed in the central and peripheral nervous systems. Receptors for VIP and PACAP belong to the same family of GPCRs. The PAC1 receptor binds both PACAP (1-27) and PACAP (1-38) with the same affinity but is unable to bind VIP, whereas the VPAC1 and VPAC2 receptors recognize both VIP and PACAP.

In the digestive tract, VIP functions as an inhibitory neurotransmitter that induces relaxation of vascular and nonvascular smooth muscle. VIP mediates the relaxation of the lower esophageal sphincter, the sphincter of Oddi, and the anal sphincter. VIP also regulates relaxation associated with gut contraction and may be involved in reflex vasodilatation in the small intestine, in part through a NO-dependent mechanism. In humans, VIP and PACAP may be co-localized to some neuronal subpopulations and are co-released as neurotransmitters leading to NO regeneration. VIP inhibits gastric acid secretion but stimulates biliary water and bicarbonate, pancreatic enzyme, and intestinal chloride secretion. VIP may also regulate pancreatic release of both insulin and glucagon and exerts either trophic or growth inhibitory effects on both normal and neoplastic cells.
Miscellaneous Gut Endocrine Peptides

In addition to the several peptide hormones outlined earlier and summarized in Table 38-3, several other gut endocrine peptides exist. Chromogranins and secretogranins are a family of secretory proteins that are found in secretory vesicles of both endocrine cells and neurons. Chromogranin-A (CgA) is a protein belonging to this family of peptides and is secreted into the circulation by several neuroendocrine tumors, especially small gastrinomas and pheochromocytomas. A direct correlation exists between circulating levels of CgA and tumor burden, making this a well-suited marker for assessing treatment response. In addition, opioid peptides regulate intestinal motility and gastric acid secretion and inhibit secretion.

Neuromedin U is a neurotransmitter that is expressed in the enteric nervous system, where it regulates intestinal motility and ion secretion.

A number of hormones are secreted by the gastrointestinal tract directly into the lumen, where they modulate secretion and the release of other hormones. Guanylin and uroguanylin stimulate water, bicarbonate, and chloride secretion by the intestine while inhibiting sodium reabsorption.

Other luminally secreted peptides include sorbin (a 153amino acid peptide involved with monitoring fluid and sodium fluxes in the duodenum) and monitor peptide (a 61amino acid peptide that stimulates CCK release).
PANCREATIC AND GUT ENDOCRINE TUMORS

Understanding the ontogeny of pancreatic and gut endocrine cell development provides some insight into the molecular pathophysiology of pancreatic endocrine tumors. Although gastrin is not normally produced in human adult islets of Langerhans, the finding of gastrinomas arising from the adult endocrine pancreas may reflect the dedifferentiation of neoplastic endocrine tumor cells that recapitulates, in part, patterns of islet gene expression observed during embryonic development. Similarly, the observation that pancreatic and gut endocrine tumors are frequently plurihormonal is consistent with studies demonstrating co-localization of peptide hormones in both fetal and adult endocrine cells in the pancreas and gut.

Pancreatic endocrine tumors may present in isolation or as part of a genetic syndrome such as multiple endocrine neoplasia type 1 (MEN-1), or the phakomatoses such as von Hippel-Lindau disease, von Recklinghausen's disease (neurofibromatosis type 1), and tuberous sclerosis. Defects in distinct tumor suppressor genes account for the phenotypic manifestations and development of tumors in these syndromes (Table 38-5). Loss of heterozygosity at 10q has been detected in several sporadic pancreatic endocrine tumors, with cellular rather than nuclear localization of PTEN (phosphatase and tensin homologue deleted on chromosome 10) detected in a substantial proportion of malignant pancreatic endocrine tumors. Similarly, loss of heterozygosity at the 11q13 MEN1 locus has also been detected in a few sporadic pancreatic endocrine tumors.

Genetic mutations in the menin gene give rise to the MEN-1 syndrome associated with an increased incidence of endocrine tumors in many organs, including the pancreas and gut carcinoids in the stomach. The MEN1 gene encodes a 610 amino acid nuclear protein that interacts with the RUN domain transcription factor, presumably resulting in derepressed cell growth. About 10% of all MEN1 germline mutations arise de novo. The current usefulness of genetic testing for all patients with suspected MEN-1 syndrome remains unclear.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Disease</th>
<th>Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>menin</td>
<td>MEN-1</td>
<td>Parathyroid, pituitary, and pancreatic endocrine tumors</td>
</tr>
<tr>
<td>VHL</td>
<td>von Hippel-Lindau disease</td>
<td>Pancreatic endocrine tumors, hemangiomomas, and multiple neoplasms</td>
</tr>
<tr>
<td>NF-1</td>
<td>Neurofibromatosis</td>
<td>Neurofibromas, pheochromocytomas</td>
</tr>
<tr>
<td>TSC1/2</td>
<td>Tuberous sclerosis</td>
<td>Pancreatic endocrine tumors, hamartomas</td>
</tr>
<tr>
<td>MEN-1</td>
<td>multiple endocrine neoplasia type 1</td>
<td></td>
</tr>
</tbody>
</table>

Figure 38-3 Clinically "nonfunctioning" tumors are often found to express one or more peptide hormones after immunocytochemical analyses. The photomicrographs represent histologic sections from the identical nonfunctioning human pancreatic endocrine tumor that exhibit immunopositivity for glucagon (A) and pancreatic polypeptide (B). (Courtesy of Dr. G. Rindt, Brescia, Italy.)

Owing to the large number of heterogeneous mutations identified in the menin gene, potentially affected family members may find utility in ruling out the diagnosis with genetic testing, thereby precluding years of biochemical testing and imaging studies.

A search for clinical manifestations of diseases associated with these genetic syndromes is an important component in the initial diagnosis and ongoing management of patients with pancreatic endocrine tumors. More careful clinical phenotypic analyses have ascertained that facial angiolipomas, collagenomas, lipomas, leiomomas, and adenocortical tumors all may be seen with increased frequency in patients with MEN-1. Moreover, somatic mutations of the menin gene have been described in isolated cases of gastrinomas, insulinomas, and gut endocrine tumors. The secretion of one or more peptide hormones resulting in the production of symptoms attributable to hormone excess, such as hypoglycemia, gastric ulceration, or profuse watery diarrhea in patients with insulinomas, or VIPomas, respectively, clearly facilitates the diagnosis of a hormone-secreting endocrine tumor. In some instances, pancreatic or gut endocrine tumors may not be associated with clinically or biochemically detectable hormone excess, and the development of a recognizable syndrome and analysis of the tumor may fail to reveal evidence for peptide hormone biosynthesis.

Nonfunctioning pancreatic endocrine tumors are more common, often larger, and more frequently malignant at the time of diagnosis. The term nonfunctioning may be a misnomer, because these tumors may produce peptide hormones (Fig. 38-3) (see also Color Plate) whose biologic actions are less clinically apparent. In some instances, tumor-associated defects in post-translational processing may preclude the efficient synthesis and secretion of peptide hormones. Factors affecting prognosis include the presence of liver metastases, incomplete resection of the primary tumor, and poorly differentiated tumor cells.

The use of somatostatin receptor scintigraphy (SRS) and measurement of gene products commonly expressed in endocrine cells such as chromogranin, PP, neuron-specific enolase, or glycoprotein hormone subunits may be useful as an adjunct for monitoring the tumor response to therapy. The widespread expression of receptors for somatostatin and multiple peptide hormone GPCRs has stimulated efforts directed at developing novel radiolabeled peptide ligands for the localization and treatment of both endocrine and nonendocrine neoplasms.

Despite the large number and complexity of endocrine cell populations in the human small bowel, gut endocrine tumors, including ileal carcinoids, are comparatively rare. Similarly, although human colon cancer remains a major cause of cancer-associated morbidity and mortality, peptide hormonesecreting carcinoid tumors arising from the colon are comparatively much less common compared with colonic adenocarcinomas. The molecular basis for the infrequent malignant transformation of human gut endocrine cells remains incompletely understood. Mutations in the regI gene have been identified in a subset of patients with ECL tumors and associated hypergastrinemia; however, the contribution of this genetic mutation to transformation of ECL cells remains unclear. The clinical presentation, diagnosis, and treatment of several more common pancreatic and gut endocrine tumors are reviewed later, and both medical and surgical perspectives to treatment have been reviewed.

Gastrinoma (Zollinger-Ellison Syndrome)

In 1955, Zollinger and Ellison described two patients with intractable peptic ulcer disease and pancreatic islet cell tumors. Subsequent studies demonstrated elevated levels of circulating gastrin associated with gastric acid hypersecretion in patients with Zollinger-Ellison syndrome (ZES).

Although the gastrin gene is not normally expressed in the adult pancreas, gastrinomas commonly arise from within the pancreas and present as endocrine carcinomas, solitary adenomas, microadenomas, or endocrine cell hyperplasia. Gastrinomas and insulinomas represent the two most common pancreatic endocrine tumors. A smaller proportion of gastrin-secreting tumors (20% to 40%) arise from the duodenum. Most (75%) gastrinomas occur sporadically, whereas approximately 25% are associated with MEN-1 syndrome. MEN-1-related patients tend to exhibit a younger age of onset at the time of diagnosis. MEN-1-associated tumors are usually multiple but may be more localized at the time of diagnosis. It is estimated that 50% to 60% of gastrinomas are malignant, based on the presence of metastases at the time of diagnosis, perhaps due in part to the long delay between the initial clinical presentation and the diagnosis of ZES. Nevertheless, gastrin-secreting tumors are often slow growing and associated with prolonged survival, despite complications.
Clinical manifestations of gastrinomas are usually related to excessive gastric acid secretion resulting in severe refractory peptic ulceration complicated by hemorrhage, perforation, and stricture. Many patients report symptoms for 5 to 6 years before the diagnosis of ZES is established. Abdominal pain, diarrhea, and heartburn are common presenting symptoms, with diarrhea and pain observed in more than 70% of patients with ZES. The diarrhea results in part from fat malabsorption due to pancreatic lipase degradation by excess gastric acid. Small-bowel inflammation and impaired nutrient absorption may also arise from excess gastric acid. Antisecretory therapy usually abolishes the diarrhea and diminishes many clinical features of ZES.

The diagnosis of gastrinoma is based on the detection of elevated fasting circulating gastrin levels (>200 pg/mL) and gastric acid hypersecretion (basal acid output >15 mEq/hour with an intact stomach or >5 mEq/hour after ulcer surgery) in patients off all acid antisecretory medication (14 days for H⁺,K⁺-ATPase inhibitors and 3 days for H₂-receptor antagonists). Although many patients with ZES have serum gastrin values that exceed 500 pg/mL, a secretin stimulation test may be performed when the serum gastrin levels are in the range of 200 to 500 pg/mL to confirm the diagnosis. Provocative testing requires overnight fasting and the intravenous administration of secretin (2-unit/kg bolus), followed by serial measurement of circulating gastrin levels at 2, 5, 10, 15, and 20 minutes. A rise in the serum gastrin level of more than 200 pg/mL within 15 minutes or a doubling of the fasting gastrin levels strongly suggests the presence of a gastrinoma.

Localization of small primary tumors or endocrine hyperplasia can be difficult. Conventional endoscopy or an upper gastrointestinal series can occasionally be used to directly visualize duodenal lesions; however, tumors are often confined to the submucosa, making detection and biopsy challenging. Radiolabeled octreotide scanning can be useful for detecting the primary tumor and metastases. Magnetic resonance imaging (MRI) or computed tomography (CT) scans can also be informative; however, the primary tumor may not be detected with these modalities alone. Endoscopic ultrasonography has been used for tumor localization with increasing success, and, less commonly, angiography with selective venous sampling may be helpful in localizing occult tumors. Primary tumors may also be localized to lymph nodes, and ectopic gastrinomas in sites such as the ovary have also been reported.

Initial treatment of patients with gastrinoma is directed at pharmacologic reduction of gastric acid secretion. Although H⁺ blockers have been used with some success, H⁺,K⁺-ATPase inhibitors such as omeprazole have become the drug of choice owing to their longer duration of action. Doses should be titrated to keep the H⁺ ion output to less than 10 mEq/hour (5 mEq/hour in patients with previous acid-reducing surgery) for the hour prior to receiving the next dose of the drug.

As outlined in Figure 38-4, in the absence of resectable disease all patients with sporadic gastrinoma should undergo surgical exploration with the intent of curative surgical resection. Exploration should include a combination of duodenal palpation, endoscopic transillumination, intraoperative ultrasonography, and duodenotomy. In as many as 20% of patients undergoing surgical exploration, the primary tumor remains undetected at laparotomy despite meticulous exploration of the abdominal cavity. Total gastrectomy should be performed only under rare circumstances in patients with severe ulcer disease refractory to medical therapy in which the primary tumor cannot be resected. Surgery is generally not indicated in patients with gastrinoma and MEN-1 syndrome because these individuals often have multiple, small pancreatic tumors that are not all amenable to surgical resection.
Glucagonoma

Most glucagomas are pancreatic in origin. Approximately 80% of tumors occur sporadically, with the remainder associated with MEN-1 syndrome. Most glucagomas (75%) are malignant and have metastasized by the time of diagnosis.

The clinical presentation reflects the various actions of the PGDPs and may vary depending on the profile of PGDPs liberated due to tumor-specific differences in the post-translational processing of proglucagon. A hallmark of this syndrome is necrolytic migratory erythema, a skin rash that usually begins in the groin and perineum as a raised, erythematous patch with occasional bullae that may also involve the lower extremities and perioral area. The exact etiology of the skin rash remains unknown, and elevated plasma glucagon levels, as well as deficiencies of zinc, amino acids, and fatty acids may represent contributing factors.

Patients with glucagomas may exhibit weight loss, abdominal pain, diabetes mellitus, stomatitis, glossitis, cheilitis, nail dystrophy, thromboembolic events, anemia, hypoaminoacidemia, and neuropsychiatric symptoms. The triad of hyperglucagonemia, necrolytic migratory erythema, and a pancreatic tumor is seen in a minority of cases. Intestinal obstructive symptoms and increased intestinal transit times have also been reported and may reflect tumor-specific liberation of GLP-1 and GLP-2, peptides with antimotility and intestinotrophic properties, respectively.

The diagnosis may be confirmed by the presence of significantly elevated levels of plasma glucagon in association with a pancreatic mass. Extremely high levels of glucagon are more often seen with the classic glucagonoma syndrome, whereas more modest elevations of glucagon are detected in the setting of plurihormonal tumors. In contrast with insulinomas, glucagomas are often large and more easily localized with imaging modalities. SRS is effective in detecting metastatic disease that most commonly involves the liver, lymph nodes, adrenal glands, or vertebrae.

Therapy with a somatostatin analogue may be useful in the setting of metastatic disease by reducing levels of circulating glucagon via the SST2 receptor, improving the skin rash, and promoting weight gain. The skin rash may also respond to selective nutrient supplementation. Although somatostatin analogues may reduce glucagon secretion and tumor-associated symptoms, effects on tumor growth are often modest. Patients with nonresectable or recurrent disease can be treated with chemotherapeutic agents such as streptozotocin and dacarbazine or interferon or with the selective use of arterial embolization.
Somatostatinoma

Somatostatinomas are extremely rare tumors that arise in the pancreas and the duodenum. Most clinical symptoms observed in the originally described somatostatinoma syndrome reflect the inhibitory properties of somatostatin on most digestive organs. A classic triad involving mild diabetes mellitus, steatorrhea, and cholelithiasis is observed in a few patients due to reduced insulin secretion, reduced biliary and pancreatic secretions, and inhibition of gallbladder motility. More prominent symptoms seen with duodenal tumors may include weight loss, postprandial fullness and abdominal pain, cholelithiasis, anemia, and hypochlorhydria. Many patients do not present with the classic triad and exhibit only nonspecific symptoms. As a result, somatostatinomas are frequently malignant with extensive metastasis to the liver by the time of diagnosis.

Duodenal tumors are more frequently seen in association with neurofibromatosis type 1 or, less commonly, von Hippel-Lindau disease, and therefore may be associated with pheochromocytomas. Most duodenal tumors are not associated with symptoms of classic somatostatinoma syndrome and may present with local obstruction and abdominal pain. Pancreatic tumors usually occur sporadically or as part of MEN-1 syndrome and are most commonly located in the head of the pancreas.

The diagnosis is confirmed by the presence of markedly elevated levels of plasma somatostatin. CT scan and both conventional and endoscopic ultrasonography may localize the duodenal tumors (Fig. 38-5) (see also Color Plate). In the small proportion of patients with localized disease, surgical resection can be curative. Patients with incurable or recurrent disease can be treated with the chemotherapeutic agents streptozotocin and dacarbazine.
Vasoactive Intestinal PeptideSecreting Tumors

The VIPoma syndrome is also known as pancreatic cholera, Verner-Morrison syndrome, or WDHA (watery diarrhea, hypokalemia, and achlorhydria) syndrome. Approximately 90% of individuals present with a pancreatic endocrine tumor that secretes VIP and often prostaglandins. The remaining tumors are extrapancreatic, usually involving the sympathetic chain or adrenal medulla. VIPomas may present as sporadic tumors or as part of MEN-1 syndrome.

Clinical manifestations include intermittent, severe watery diarrhea that contains large quantities of potassium, bicarbonate, and chloride. As a result, patients may exhibit signs and symptoms of hypokalemia, metabolic acidosis, and dehydration. Hypotension can occur due to dehydration and the vasodilator effects of VIP. Diarrhea is secretory and does not respond to antidiarrheal medications. Gastric analysis usually reveals hypochlorhydria or achlorhydria, although an appropriate increase in acid secretion is observed in response to a pentagastrin challenge. Glucose intolerance may be present due to hypokalemia and altered insulin sensitivity. Cutaneous flushing of the head and trunk may be observed in 15% of patients, usually during a bout of diarrhea, and may be associated with a patchy erythematous rash.

The diagnosis of VIPoma may be challenging due to the intermittent nature of the symptoms. A history of severe recurrent, severe diarrhea, together with elevated fasting levels of plasma VIP (>200 pg/mL), should prompt a search for a pancreatic tumor. Increased circulating levels of peptide histidine methionine, PP, NT, and prostaglandins have also been detected in patients with VIP-producing tumors. VIPomas can be localized by ultrasonography, CT scan, and SRS imaging, and exploratory laparotomy with intraoperative ultrasonography may also be used to identify the tumor.

Initial treatment of patients with VIPoma syndrome involves aggressive fluid and electrolyte replacement. Somatostatin analogues may be used preoperatively to control the diarrhea by both lowering circulating VIP and directly inhibiting intestinal secretion. Definitive treatment requires surgical resection of the tumor, commonly located in the body or tail of the pancreas. Although tumors are usually solitary, 60% are malignant at the time of diagnosis, with 75% metastasizing to the liver and regional lymph nodes, and, less commonly, to the lungs, mediastinum, stomach, and kidney.

If a pancreatic tumor cannot be identified, exploration of the retroperitoneum that includes the adrenal glands and sympathetic chains is indicated. If no pancreatic tumor is identified, some patients may elect to be closely monitored, whereas others may opt for a 80% distal pancreatectomy. This latter strategy may be beneficial for the 10% to 20% of symptomatic patients with diffuse islet cell hyperplasia. In patients with inoperable or metastatic tumor, a combination of 5-fluorouracil and streptozotocin may be effective.
Miscellaneous Gut Hormone-Producing Tumors

Although rare, pancreatic endocrine tumors may secrete PTH, growth hormonereleasing hormone, and adrenocorticotropic hormone, leading to the development of hypercalcemia, acromegaly, and Cushing's syndrome, respectively. A large number of peptide hormones may be produced by pancreatic endocrine tumor cells, including PYY, calcitonin, NT, melanocyte-stimulating hormone, corticotropin-releasing hormone, NPY, NMB, CGRP, GRP, and motilin. In some cases, the hormone precursors may be produced, but the correctly processed intact hormone may not be secreted by the tumor. Accordingly, excessive production of many of these hormones may not always be associated with characteristic signs and symptoms. Similarly, carcinoid tumors of the gastrointestinal tract often exhibit immunopositivity for multiple peptide hormones in the absence of a recognizable clinical syndrome.
Summary

A large number of peptides are synthesized in and secreted by endocrine cells of the pancreas and gastrointestinal tract. Many of these peptides circulate as hormones, but they also function as paracrine modulators or neurotransmitters not only in the gut but in the central and peripheral nervous systems. Although some biologic actions for many of these peptides have been delineated, it seems likely that new peptides, receptors, and novel biologic functions will continue to be discovered, which may provide new opportunities for understanding the pathophysiology, diagnosis, and treatment of endocrine disease.
References


1794


BREAST CANCER

Etiology

A variety of data suggest that estrogens directly cause or contribute to the development of breast cancer. Administration of exogenous estrogens to various animal species results in breast cancer. Spontaneous development of breast cancer in aging rats can be prevented by oophorectomy or administration of aromatase inhibitors to block estrogen production. Oophorectomy before age 35 years in women lowered the risk of breast cancer by 75% over a 25-year period. Administration of antiestrogens caused a 50% reduction in breast cancers diagnosed in women at high risk for development of breast cancer. These observations and other data have led to the classification of estrogen as a carcinogen.
Sources of Estrogen

The breast derives estradiol from three separate sources: the ovary, extraglandular tissues, and the breast itself. First, direct glandular secretion by the ovary results in delivery of estradiol to the breast through an endocrine mechanism in premenopausal women. Second, after the menopause, extraglandular production of estrogen from ovarian and adrenal androgens provides the second source of estradiol. Third, the breast itself can synthesize estradiol through aromatization of androgens to estrogens or cleavage of estrone sulfate to estrone by the enzyme sulfatase. Estradiol acts through paracrine, autocrine, and intracrine mechanisms on cells in the breast. Several factors regulate in situ estradiol synthesis but the most important is the degree of obesity, which increases the amount of aromatase in breast and, consequently, estradiol production.
Estrogen-Induced Carcinogenesis

Mutations of key genes involved in cell proliferation, deoxyribonucleic acid (DNA) repair, or apoptosis must accumulate to produce cancer. It is likely that mitogenic as well as mutagenic effects of estradiol act in concert to initiate and promote the development of breast cancer. As a rule, the frequency of mutations increases in parallel with the number of mitotic divisions in a proliferating tissue. Accordingly, estradiols may initiate mutations leading to neoplastic transformation by increasing the rate of cell proliferation. As cells divide more rapidly, less time is available for DNA repair. Estradiols also enhance tumor promotion by increasing the rate of cell division with propagation of the mutations already present.

Metabolites of estradiol may be directly mutagenic through a pathway involving the 1B1 cytochrome P450 enzyme. This catalyzes conversion of estradiol to the catechol-estrogen 4-hydroxyestradiol, which is then further metabolized to 3,4-estradiol-quinone. This highly reactive species binds covalently to guanine or adenine molecules in the DNA helix, which activates a glycosidase and results in depurination. Error-prone or replicative repair of the depurinated site leads to point mutations. Recent studies directly demonstrated estradiol-guanine conjugates in benign and malignant human breast tissue, establishing the activity of the genotoxic pathway in women.

Estradiol is not the sole factor mediating the development of breast cancer. A number of specific genetic mutations are associated with a high incidence of breast cancer. Studies in twins suggest that approximately 27% of breast cancers arise because of genetic factors. The most common mutations involve the BRCA1 and BRCA2 genes and cause approximately 5% of breast cancer cases. Rarer genetic syndromes include mutations of the p53 gene in the Li-Fraumeni syndrome, impaired cell cycle checkpoint surveillance in the ataxia-telangiectasia syndrome, mutations in the PTEN gene in the Cowden syndrome, the MLH1 and MSH2 genes in the Muir-Torre syndrome, and an STK11 mutation in the Peutz-Jeghers syndrome. Dietary and environmental factors play a key role in breast cancer etiology and contribute to the fourfold difference in incidence between Japan, with a rate of 23 women per 100,000 per year, and the United States, at 90 per 100,000 per year. Epidemiologic observations suggest a role for a high-fat diet and resultant obesity in the genesis of breast cancer. In Japan, the rate of breast cancer peaks at the age of menopause; in the United States, however, the incidence continues to increase until age 90 years.
Hormonal Risk Factors for Breast Cancer

Most risk factors for breast cancer are related to the duration or intensity of a woman's exposure to endogenous or exogenous estrogens. Early menarche or late menopause, or both increase breast cancer risk. Elevations in circulating estradiol levels predict the risk for development of breast cancer over the ensuing years in postmenopausal women. An estradiol level in the top quintile is associated with an increase in the relative risk of breast cancer by five-fold. Putative markers of long-term estrogen exposure such as bone density are also predictive. Women in the top quartile of bone density have a threefold increase in risk of breast cancer.

Late first births increase the risk 2.8-fold and is believed to be related to the lack of the differentiating effect of pregnancy on the type of breast lobule present. Gain of at least 20 kg as an adult increases breast cancer risk twofold, probably as a result of increasing aromatase activity and estrogen production by adipose tissue. Increased waist-hip ratio is associated with a similar increase in risk. Several but not all studies suggest that alcohol intake can increase the risk of breast cancer, perhaps by decreasing the clearance of estradiol. Increased exposure to estradiol in utero, as shown by twin studies, may increase the risk of breast cancer by as much as twofold. Early pregnancy and prolonged duration of breast-feeding diminish the risk. More dramatic is the 75% reduction in risk caused by bilateral oophorectomy before age 35 years.

Mammographic density represents the most powerful risk factor for breast cancer. Increased breast density probably reflects either an increase in exposure to estrogen or sensitivity to it. Exogenous estrogens increase and antiestrogens reduce breast density. Because of these effects, hormone replacement therapy (HRT) alters the sensitivity and specificity associated with interpreting mammograms. The increase in breast cancer risk from the lowest to the highest breast density category is on the order of sixfold, depending on the age of the patient, with a greater relative risk seen in older women.

Exogenous Estrogens and Breast Cancer Risk

In premenopausal women, use of oral contraceptives for 10 or more years increases the relative risk of breast cancer by approximately 10%, However, this increase in relative risk affects few women because the age-related incidence of breast cancer is quite low in women taking oral contraceptives. Controversy surrounds the concept that HRT in postmenopausal women increases the risk of breast cancer. More than 50 observational studies have examined this question but have reported conflicting results. However, a key study clarified the factors responsible for the different conclusions in the various reports. A meta-analysis from the Collaborative Group on Hormonal Factors in Breast Cancer examined data for 52,705 women with breast cancer and 108,411 without. Five objective factors that confounded the interpretation of prior studies were identified.

1. The relative risk of breast cancer associated with estrogen replacement therapy is small, and large studies are required to minimize type I and type II statistical errors.
2. The risk of breast cancer appears to increase linearly with duration of HRT use. Accordingly, comparisons of "ever users" with "never users" are invalid because the duration of estrogen use is not considered.
3. The increased risk of breast cancer associated with estrogens appears to dissipate within 4 years of cessation of therapy. Therefore, only women using estrogen within 4 years of the study might be found to be at increased risk.
4. Breast cancer risk diminishes over a 4-year period after the menopause, presumably reflecting decreased estrogen levels. As a result, analyses of observational studies need to match users and nonusers according to time after menopause.
5. The increased risk of breast cancer appears to be limited to nonobese women (i.e., body mass index [BMI] < 25 kg/m²). Inclusion of a large proportion of obese women in a single study might obscure an association between estrogen use and breast cancer risk.

Taking into account these five factors, the meta-analysis concluded that the relative risk of breast cancer increases linearly by 2.3% per year of HRT for up to 25 years. Both the slope of the risk-time analysis (P value) and the overall risk of breast cancer among estrogen users (P value) were highly statistically significant. This study detected no increased risk associated with HRT use in obese women (i.e., BMI > 25 kg/m²). A hypothetical explanation for the differences between obese and thin women involves the degree of in situ estrogen production. Obese women may have an increase in breast tissue estrogen as a result of increased aromatase activity, whereas lean women have lower levels. Exogenous estrogen might then produce a greater percentage increase in estrogen levels in thin than in obese women.

Data suggest that addition of a progestin to an estrogen replacement therapy regimen enhances the risk of breast cancer to a greater extent than estrogen alone. Although data from cell cultures or animal studies are not conclusive, the weight of evidence from patients suggests that progestins are mitogenic in breast tissue. The physiologic basis for this conclusion rests on data indicating that progestins are mitogenic to breast tissue in contrast to their antimitogenic effects on the uterus. Although data from cell cultures or animal studies are not conclusive, the weight of evidence from patients suggests that progestins are mitogenic in breast tissue. Mammographic studies demonstrate that breast density in women taking an estrogen-progestin combination is greater than breast density in those taking estrogen alone. Histologic examination demonstrated increased cell proliferation and percentage of the breast containing glandular tissue as a function of duration of progestin usage. Increased proliferation would be expected to increase both initiation and promotion of breast cancer in a manner similar to that for estrogens.

A number of observational studies have suggested an increase in breast cancer risk in women receiving an estrogen-progestin combination, but these were not considered conclusive. Recent results from a large, prospective, randomized, placebo-controlled trial in postmenopausal women now provide compelling evidence for this effect. Nearly 16,000 postmenopausal women with an average age of 63 enrolled in the
over time. Shaded area represents the 95% confidence limits of that risk. 8. Increase in relative risk of breast cancer in women taking estrogen plus a progestin as HRT. Solid line represents the mean increase in relative risk over time. Shaded area represents the 95% confidence limits of that risk. (From Santen RJ, Pinkerton J, McCartney C, et al. Risk of breast cancer with progestins in combination with estrogen as hormone replacement therapy. J Clin Endocrinol Metab 2001; 86:1623.)

<table>
<thead>
<tr>
<th>Hormone</th>
<th>Age Initiated (yr)</th>
<th>Duration of Use (yr)</th>
<th>Per 100 Women</th>
<th>As Numeric Chance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estrogen alone</td>
<td>50</td>
<td>2</td>
<td>0.0101</td>
<td>1:9925</td>
</tr>
<tr>
<td>Estrogen-prog</td>
<td>50</td>
<td>2</td>
<td>0.0806</td>
<td>1:1241</td>
</tr>
<tr>
<td>Estrogen alone</td>
<td>50</td>
<td>10</td>
<td>0.252</td>
<td>1:397</td>
</tr>
<tr>
<td>Estrogen-prog</td>
<td>50</td>
<td>10</td>
<td>2.016</td>
<td>1:50</td>
</tr>
<tr>
<td>Estrogen alone</td>
<td>60</td>
<td>10</td>
<td>0.350</td>
<td>1:286</td>
</tr>
<tr>
<td>Estrogen-prog</td>
<td>60</td>
<td>10</td>
<td>2.780</td>
<td>1:36</td>
</tr>
</tbody>
</table>


Women’s Health Initiative study and received either placebo or conjugated estrogens plus medroxyprogesterone-acetate for 5 years on average. The study was terminated early because of an increased incidence of breast cancer in the HRT group with a relative risk of 1.26 and 95% confidence interval of 1.00 to 1.59. The absolute excess of cases was small, with only 8 more invasive breast cancers per 10,000 person-years in the HRT group. Nonetheless, these data confirm the prior observational studies and indicate a relative risk increase of 5.5% per year in those receiving HRT. This is similar to the 8% per year increase reported in the observational studies (Fig. 39-4). At the present time it is not known whether the type or dose of progestin can alter breast cancer risk. Various observational studies have examined differences between combined continuous and sequential regimens and various types of progestins but data are insufficient to make definitive conclusions. It has been suggested that certain types of progestin may exert no effects on breast cancer risk, but this requires further study. With respect to estrogen alone and breast cancer, another Women’s Health Initiative study is currently comparing placebo with estrogen alone in women who had previously undergone a total abdominal hysterectomy. This study continues and has not at this time shown a sufficient increase in breast cancer with estrogen alone to warrant terminating the study.

Review of the totality of basic and clinical data strongly supports an adverse effect of adding progestins to estrogens in women. Regarding the effects of estrogen alone on breast cancer risk, one must interpret the existing data cautiously because of the retrospective/observational nature of available studies. Definitive information will await prospective information from the Women’s Health Initiative study that is comparing estrogen alone to placebo. In the meantime, it is considered prudent to advise patients about the worst case conclusion, that estrogen alone may increase their risk of breast cancer. Additional risks may also be imparted for ovarian and uterine cancer.
Relative, Absolute, and Attributable Risks

Epidemiologists use relative risk analysis as a tool that provides substantial power to determine statistical significance. However, this term is misleading to patients because actual risk is quite small in magnitude. The lay press, patients, and many physicians confuse the terms relative, absolute, and attributable risk.

Relative risk is defined as the ratio of risk under one condition to risk under another condition and does not take into account the frequency of occurrence of that condition.

Absolute risk is determined by multiplying the underlying incidence rate in the group being considered by relative risk. For example, the absolute risk of breast cancer development for an average 50-year-old woman in the United States is 2.52 per 100 women over a 10-year period. A 10% increase in relative risk resulting from estrogens alone would increase a woman's absolute risk of breast cancer to 2.77 per 100 women.

Attributable risk is defined as the number of women who would develop breast cancer that would not have occurred unless they had used estrogen replacement. Using the preceding example, the difference between breast cancer risk of 2.52 per 100 and 2.77 per 100 represents the increased risk attributable to estrogen or 0.252 per 100 women.

Attributable risks can be translated into the numeric odds of one's being adversely affected by estrogen (Table 39-1). For example, the chance of an estrogen-induced breast cancer in a 50-year-old woman receiving estrogen replacement therapy over 2 years is 1 in 9925; over 10 years, it would be 1 in 397. For a woman taking estrogen plus a progestin, the comparable odds are 1 in 1241 for 2 years of use and 1 in 50 for 10 years.

Figure 39-5 Relative risk of invasive breast cancer related to several benign breast lesions in women with long-term follow-up.

<table>
<thead>
<tr>
<th>Lesion</th>
<th>Relative Risk</th>
<th>Absolute Risk (2 yrs)</th>
<th>Absolute Risk (10 yrs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Placebo</td>
<td>1.00</td>
<td>2.52</td>
<td>2.52</td>
</tr>
<tr>
<td>Estrogen</td>
<td>1.10</td>
<td>2.77</td>
<td>2.77</td>
</tr>
</tbody>
</table>

Attributable risk is the difference between the two, or 0.252 per 100 women.
Benign Breast Disease and Risk of Breast Cancer

Benign breast lesions with an enhanced rate of proliferation or increased degree of atypical features predict an increased incidence of breast cancer over time. Some of these lesions contain genetic mutations, and many lesions are analogous to adenomas because of clonal populations of mutated cells. Hyperplastic ductal lesions are often multicentric, suggesting that some type of underlying abnormality or field defect is present that predisposes to such lesions. The multifocal nature of the associated benign hyperplastic lesions is most apparent in breast tissue from women with cancer. Examination of tissue adjacent to an invasive breast cancer or in the contralateral breast reveals one or more additional hyperplastic lesions in approximately 40% of patients.

The nature of the field defect has not been specifically identified, but hypothetically it may represent a single mutation of a gene controlling local estrogen production, cellular proliferation, DNA repair, metabolism of procarcinogens to carcinogens, or other cellular events. Preliminary data suggest progression from adenomas to frank neoplasms. Eighty to 90% of hyperplastic lesions contain DNA mutations similar to those in the contiguous tumors. Extensive molecular genetic studies have now shown progression of abnormalities in the spectrum of breast lesions.

A major consideration for women who present with breast problems is whether they have a higher than normal risk of breast cancer. Certain breast lesions, such as fibrocystic changes, are associated with no increased risk of subsequent breast cancer (see Fig. 39-5), whereas other lesions, such as atypical ductal hyperplasia, involve an increased risk. The relative risk of development of invasive cancer is increased 10-fold to 12-fold when ductal carcinoma in situ (DCIS) and lobular carcinoma in situ (LCIS) are present.
Estimating Breast Cancer Risk

To aid in assessment of breast cancer risk, a questionnaire developed by Gail and co-workers utilizes answers to seven questions to calculate the 5-year and lifetime risks of breast cancer. This model has recognized deficiencies, in that it does not consider breast density, plasma estradiol levels, bone density, BMI, weight gain in adulthood, second-degree relatives with breast cancer, proliferative lesions of breast other than atypical ductal hyperplasia, alcohol intake, or birth control pill and HRT use.

Nonetheless, two major prospective studies validated the Gail model in a high-risk (National Surgical Adjuvant Breast and Bowel Project [NSABP] prevention study) and in an average-risk population of women (the Nurses Health Study). The ratio of observed to expected cancers with this tool is 1.03 (95% confidence limits of 0.88 to 1.21) in the high-risk patients and 0.94 (95% confidence interval [CI] 0.89 to 0.99) in average-risk women, and both are highly statistically significant. This risk tool, called the RISK DISK, is available from the National Cancer Institute. When second-degree relatives with breast cancer represent the major risk factor, the Claus model is a more valid risk assessment tool.
Breast Cancer Risk and Clinical Decisions

Knowledge about underlying breast cancer risk influences advice given by health care providers and choices made by patients. Women known to be at high risk for breast cancer frequently choose a surrogate for estrogen (see later) to obtain its benefits while avoiding its risks. Those at low risk usually choose estrogen or an estrogen-progestin combination to relieve symptoms of estrogen deficiency or to prevent osteoporosis. Those at intermediate risk have several options, including use of a selective estrogen receptor modulator (SERM), other alternatives to estrogen, watchful waiting, or HRT.

As a working guide, we arbitrarily define risk categories on the basis of the Gail model:

- **High risk**, more than a 3% chance of breast cancer in 5 years
- **Intermediate risk**, 1.5% to 3% chance
- **Low risk**, less than 1.5%

Women classified as **high risk** include patients with a strong family history of breast cancer (particularly if associated with ovarian cancer), a prior history of atypical ductal hyperplasia or LCIS, and age older than 60 years when combined with early menarche, late menopause, or late first live birth. **Intermediate risk** patients have some risk factors but not others. **Low-risk** patients are younger than 60 years, have a late onset of menarche, early menopause, early age of first live birth, and no family history of breast cancer or prior predisposing breast lesions.

Because no formal risk tool incorporates breast density, bone density, or plasma estradiol levels in its assessment, a physician must take these factors into account in advising patients.
Prevention of Breast Cancer

Depending on their risk category, women may wish to take tamoxifen or raloxifene to prevent breast cancer. Only the antiestrogen tamoxifen is approved for this use in the United States. Although evidence supporting antiestrogens for prevention is substantial, results from individual studies are conflicting. Five clinical trials provided data regarding the efficacy of antiestrogens for prevention of breast cancer.\[72\]

The most definitive trial, the NSABP P-1 study, involved 13,388 women randomly assigned to receive either placebo or 20 mg of tamoxifen daily.\[72\] Eligibility for the study required an intermediate or high risk of breast cancer, defined in the study as a 1.67% or greater chance of development of a new breast cancer over a 5-year period. The rate of breast cancers was 9.4 per 1000 women-years in the placebo group versus 4.7 in the tamoxifen group, a relative risk reduction of 0.50.\[72\]

Two other published tamoxifen cancer prevention trials did not find a statistically significant reduction of breast cancer risk. Both studies were much smaller than the NSABP trial (2471 women in the United Kingdom study and 5408 women in the Italian study)\[72\] and are considered flawed because of concomitant estrogen use (the United Kingdom study) or frequency of oophorectomy (the Italian study).\[72\]

A fourth trial (Multiple Outcomes of Raloxifene [MORE]) involved more than 7705 women taking the antiestrogen raloxifene or a placebo.\[72\] Although breast cancer incidence was only a secondary study end point, a highly significant 70% reduction in new breast cancers (relative risk 0.30) was observed after 48 months of raloxifene use.\[72\] A fifth study examined the effect of tamoxifen on contralateral breast cancer in women with diagnosed DCIS and found a significant reduction.\[72\] Review of these prevention trials led a panel of experts commissioned by the American Society of Clinical Oncology (ASCO) to conclude that tamoxifen does reduce the incidence of newly diagnosed breast cancer in high-risk women.\[72\]

Tamoxifen in the breast cancer treatment setting is considered well tolerated and safe. However, for use in otherwise normal women, infrequent side effects and toxicity become more important. Up to 40% of women who start tamoxifen do not continue it because of perceived side effects, including depression and mood changes.\[72\] Tamoxifen belongs to the class of agents called SERMs. Agents in this class exert antiestrogenic effects on tissues such as breast but estrogenic effects on other tissues such as uterus and liver. These various actions of the SERMs must be factored in when estimating risks and benefits in the setting of breast cancer prevention.

To compare risks and benefits in a meaningful way, the NSABP data are expressed as the number of women per 100 with a specific benefit or adverse event after 5 years of study. With respect to benefits, new invasive breast cancers were prevented in 1.7 of 100 women, noninvasive cancers in 0.67, and bone fractures in 0.50, for a total of 2.87 in 100 benefitted after 5 years. Risks of tamoxifen are primarily related to its estrogenic effects on the uterus, a prothrombotic effect, and adverse effects on the lens of the eye. To estimate actual risks validly for a patient, one must correct for the underlying risks in the population under study. For that reason, analysis of adverse events includes determination of attributable risk and involves subtracting the underlying rate from the total observed with tamoxifen. This analysis indicates an excess of 1.5 of 100 women who experienced xerostomia, 0.69 with endometrial cancer (nearly exclusively in postmenopausal women), 0.27 with cerebrovascular accident (CVA), 0.25 with deep venous thrombosis (DVT), and 0.23 with pulmonary embolus, or a total of 2.94 per 100.\[73\] The new endometrial cancers were exclusively stage I and pulmonary emboli nonfatal. Neither risk nor benefit was observed regarding cardiovascular effects,\[73\] but the number of patients was not sufficient for statistical significance.

The increased incidence of uterine cancer in patients receiving tamoxifen presents a clinical management problem. Substantial study has been addressing how best to detect this cancer early.\[72\]\[73\]\[74\]\[75\]\[76\]\[77\]\[78\]\[79\]\[80\]

Transvaginal ultrasonography was initially proposed as a means of assessing both endometrial hyperplasia, as a precursor lesion, and early cancer. A prospective study demonstrated a mean increase in endometrial thickness from 3.5 ± 1.1 to 9.2 ± 5.1 mm in response to 5 years of tamoxifen. In 20% of women, endometrial thickness exceeded 10 mm or appeared suspicious for neoplasm. In this subset of women, biopsies revealed atrophy in 73%, polyps in 17%, and hyperplasia in 7.7%.\[81\] Endometrial thickness does not usually indicate hyperplasia but, rather, the presence of edema and dilated myometrial glands.\[81\] It should be noted that nonprospective studies reported a higher prevalence of benign polyps (5% to 55%) and hyperplasia (8% to 16%),\[81\]\[82\]\[83\]\[84\] but these are likely to overestimate resulting from selection bias.

On the basis of these studies, recommendations for screening include a yearly gynecologic examination with reservation of endometrial sampling and ultrasonography for patients with signs of abnormal vaginal bleeding.\[82\]\[84\]

Guidelines for Breast Cancer Prevention

Premenopausal women with a 5-year risk of breast cancer greater than 1.67% are candidates for tamoxifen unless they are at increased risk for DVT or pulmonary emboli. Postmenopausal women are candidates for this approach if they no longer have a uterus and lack a predisposing risk for DVT or pulmonary emboli. The decision to take tamoxifen should be made by the patient in partnership with her health care provider and based on a full discussion of individual risks and benefits expressed in absolute and not relative terms.\[84\]\[85\]

Critical Assessment of Breast Cancer Prevention Strategies

Estimates indicate that between 20 and 100 women (depending on underlying risk) must be treated to prevent one breast cancer. Clinical decisions depend on analysis of risk/benefit ratios and should be made after full discussion between health care provider and patient.\[86\] Regarding efficacy, available data demonstrate a reduction in newly diagnosed breast cancers. It is too early to assess overall survival. It is not known whether tamoxifen prevents breast cancer, cures some preexisting subclinical cancers, or delays the onset of small tumors. Mathematical modeling techniques suggest that the "preventive effects" are equally divided between blockade of growth of occult tumors and prevention of new ones.\[86\] Tumors whose growth was blocked by tamoxifen during 5 years of administration might
be expected to regrow later. For this reason, data on overall survival (to be available from ongoing European studies) are critically important.

Because reports indicate that both tamoxifen and raloxifene prevent breast cancer, it is expected that many women will ask their physicians about these two agents. The use of raloxifene exclusively for breast cancer prevention is not currently recommended. Although the MORE study demonstrated a reduction of breast cancers diagnosed in women taking raloxifene, breast cancer prevention was not a primary end point. The ongoing STAR (Study of Tamoxifen and Raloxifene) trial is currently testing the efficacy of tamoxifen versus raloxifene for breast cancer prevention.

Future studies are needed that utilize additional factors to select women at higher risk for development of breast cancer. A study of raloxifene examined women with factors suggesting high long-term exposure to estrogen (i.e., increased bone mineral density, high BMI, and high plasma estradiol levels). Results indicated that raloxifene was more effective in preventing breast cancer in women with high, long-term estrogen exposure than in those with low exposure (94% versus 56% for BMD; 82% versus 64% for BMI; and 77% versus 55% for estradiol levels, \(P < .005\)).
Treatment of Established Breast Cancers

In the year 2000, there were 184,000 new cases of breast cancer diagnosed in the United States and 41,000 deaths. The mechanisms whereby estradiol stimulates breast cancer growth are complex and involve direct regulation of genes involved in control of proliferation and apoptosis, induction of growth factors through secondary actions, and cross-talk between growth factor and estrogen pathways at both upstream and downstream levels. Nongenomic effects of estradiol on mitogenic pathways may also be involved. Treatment strategies utilize agents that abrogate the effects of estrogen on these pathways and thus inhibit growth and induce apoptosis.

Common approaches involve blockade of estrogen action with antiestrogens or estrogen synthesis with aromatase inhibitors or gonadotropin-releasing hormone (GnRH) superagonist analogues. Patients who initially respond to hormonal therapy eventually have relapses. The biologic mechanisms for this phenomenon are not fully understood. A standard situation is that responders to the first-line treatment benefit from second-ary and tertiary hormonal therapies upon relapse. The sequential responses to hormonal therapies suggest an adaptive process whereby tumors do not become totally resistant to hormonal therapy but develop a transitional state in which alternative means of blocking hormonal pathways cause tumor regression.

Patients who have relapses after oophorectomy or tamoxifen treatment commonly respond secondarily to inhibitors of estrogen production (aromatase inhibitors) or, rarely, to the withdrawal of tamoxifen. A hypothesis to explain this phenomenon is that these tumors become hypersensitive to lower amounts of circulating estrogen or to the estrogen agonist properties of tamoxifen. Experimental support for this concept comes from observations in xenograft models and in vitro studies. Breast tumor xenografts adapt to long-term exposure to tamoxifen by responding to it as an estrogen rather than as an antiestrogen. Under these circumstances, the pure antiestrogen ICI 182,780 (fluvestrant [Faslodex]) caused tumor regression. In an in vitro model system, long-term deprivation of estradiol rendered breast cancer cells hypersensitive to the proliferative effects of this sex steroid. This adaptive process is associated with increments in mitogen-activated protein (MAP) kinase, an enzyme that stimulates cell proliferation. The hypersensitivity concept may explain secondary responses to aromatase inhibitors in patients who have relapses after oophorectomy or tamoxifen and secondary responses to the pure antiestrogen Faslodex.
Development of Hormonal Resistance

Women respond to each hormonal therapy on average for 12 to 18 months and then experience relapse. At some time in their course, tumors become totally resistant to further hormonal therapy. Several explanations for development of resistance have been suggested, including:

2. Constitutive increase in growth factor production as a result of additional oncogene mutations.
3. Enhanced growth factor receptor functionality.
5. Outgrowth or selection of hormone-resistant clones of tumor cells.
6. Other adaptive mechanisms.
Prognostic Factors

Clinical decision making requires knowledge of the degree of aggressiveness and natural history of the diagnosed breast cancer. Numerous histologic types occur, and pathologic features provide information related to the prognosis. Much attention has been directed toward multivariate and neural network analysis to calculate the precise prognosis in individual women. In general, these methods have not been particularly useful in practical decision making and individual factors are used in treatment algorithms. \[158\] \[159\] \[160\] \[161\]

**Figure 39-7** illustrates the prognostic power of various biologic factors by comparing the effects of individual parameters on 5-year survival. \[141\] \[142\] \[144\] \[145\] The most powerful prognostic parameters are related to tumor size, invasiveness, nodal status, histologic grade and type, and hormone and human epidermal growth factor receptor 2 (HER2/neu receptor status) (see Fig. 39-7). Other less useful prognostic characteristics include DNA labeling index (LI), percent S phase, percent PCNA or Ki67 positivity (markers of proliferation), degree of aneuploidy, and overexpression of certain oncogenes or coactivators (e.g., D1 cyclin, A1B1, MAP kinase, Ras, HER3 and HER4, heregulin, c-Src, and ODC [ornithine decarboxylase] levels). \[149\] \[150\] \[152\] \[154\] \[156\] \[158\] \[160\] 

Finally, tumors diagnosed in postmenopausal women receiving HRT have a lower histologic grade and a 10% better prognosis than those in women not receiving HRT. \[160\] \[165\] \[166\]
Predictive Factors

Certain other biologic parameters allow assessment of the potential effectiveness of certain therapies. Older age, long disease-free survival, high degree of tumor differentiation, and prior response to endocrine therapy predict a higher likelihood of responses to hormonal therapy. The ability to measure receptors markedly improves the process of selection of patients for hormonal therapy. Absence of estrogen receptor (ER) in the tumor predicts that fewer than 5% to 10% of women will respond to hormonal therapy. If both ER and progesterone receptor (PR) are negative, an even lower percentage respond. Patients with both ER-positive and PR-positive tumors respond to hormonal therapy 50% to 75% of the time, whereas 30% to 50% of patients with ER-positive but PR-negative tumors are responsive. Preliminary but conflicting data suggest that patients with ER- and PR-positive tumors respond to hormonal therapy less frequently if HER2/neu is positive.

Receptor measurements are commonly performed by immunocytochemical analysis, which correlates well with ligand-binding assays. Immunocytochemical techniques measure only ER alpha and not ER beta, but 76% of ER alpha-positive tumors also contain ER beta (S. Fuqua, personal communication, 2001). Currently, it is not clear whether measurement of ER beta would provide important clinical information. Data suggest that certain tumors make ER beta variant proteins that can heterodimerize with full-length ER alpha or beta and exert dominant negative effects. This concept suggests that further refinement of receptor assays may improve their predictive value.

Predictive parameters for chemotherapy are more limited. Measurement of HER2/neu content predicts responses to trastuzumab (Herceptin), an antibody directed against HER2/neu. Preliminary data suggest that this also predicts which ER-positive patients will not respond to hormonal therapy. A concept proposed many years ago was that ER-negative tumors responded better to chemotherapy than ER-positive tumors because of the more rapid proliferation rate of ER-negative tumors. Although originally controversial, meta-analyses now demonstrate the positive predictive value of a negative ER for chemotherapy.
Hormonal Therapies for Breast Cancer: Mechanism of Action

Surgical Ablative Therapies

Historically, the initial approaches to hormonal therapy involved surgical removal of endocrine glands responsible for synthesis of estrogen or its precursors. In 1896, Beatson first demonstrated that oophorectomy caused regression of breast cancer in premenopausal women; comprehensive studies later documented responses in one third of patients. The availability of glucocorticoid replacement therapy in the 1940s enabled the use of adrenalectomy or hypophysectomy. These maneuvers produced similar rates of response as a result of a reduction in estradiol levels and perhaps also of pituitary factors in the case of hypophysectomy. The development of medical means to block hormone synthesis or action has largely replaced adrenalectomy and hypophysectomy but surgical oophorectomy persists.

Hormone Additive Therapies

Clinicians learned from empirical observations that high doses of estrogen, androgen, or progestins caused tumor regression. The mechanism whereby estrogens paradoxically inhibit tumor growth is unknown. This therapy is most effective in women who experienced menopause several years previously. A variety of observations suggest that an increased ratio of androgens to estrogens exerts an antagonistic effect on breast tissue. Progestins have glucoceptic actions that suppress circulating estrogen levels and can also act through progesterogen mechanisms.

Medical Ablative Therapies

Medical means of abating hormone secretion or action avoid the need for major surgery and can effectively replicate the hormonal and clinical effects of these procedures. High doses of GnRH agonist analogs act as agonists, with concomitant reduction in plasma estradiol. Physiologically, the pituitary requires pulsatile exposure to GnRH to maintain gonadotropin secretion. GnRH agonists suppress luteinizing hormone (LH) and follicle-stimulating hormone by exposing the pituitary to a continuous GnRH stimulus that causes a paradoxical gonadotropin inhibition. Preparations lasting 3 to 4 months can be given by intramuscular injection. For the first several days after initiation of therapy, an increase in LH, follicle-stimulating hormone, and estradiol occurs, but thereafter these hormones fall to suppressed levels.

Antiestrogens with Mixed Agonist-Antagonistic Actions

Blockade of estrogen action rather than synthesis provides an additional strategy. These agents, called antagonists, exert effects similar to those of surgical oophorectomy in premenopausal women and hypophysectomy or adrenalectomy in postmenopausal women. Tamoxifen, the initial antiestrogen of this type, was introduced for use in the United States in the mid-1970s. Early clinical observations indicated that this drug has an antiestrogen effect on breast tissue but an estrogen agonist effect on uterus, vagina, bone, pituitary gland, and liver.

Attempts to determine the divergent actions of tamoxifen led to an understanding of the complexity of ER-mediated transcriptional regulation and actions of the antiestrogens. Tamoxifen binds to both ER alpha and ER beta in the ligand binding domain (AF 2) of the receptor, which then facilitates the binding of the antiestrogen-ER complex to specific estrogen response elements on DNA. Conformational changes in the ER binding pocket at helix 12 induced by antiestrogens interfere with binding of coactivators to the ER but rather facilitate binding of corepressors to this area. The continued presence of corepressor in the complex is thought to explain the antiestrogenic properties of tamoxifen. The relative amounts of corepressor and coactivator in certain tissues and the presence of other unknown factors determine whether tamoxifen acts as an agonist or antagonist. Additional estrogenic effects are mediated by nongenomic actions at the cell membrane as well as by protein-protein interactions with c-Jun, specificity protein 1 (SP-1), insulin-like growth factor receptor, phosphatidylinositol-3-kinase (PI-3-kinase), HER2/neu, c-Src, and potentially other factors.

On the basis of the SERM concept, other agents have been introduced or are being developed to enhance breast antagonistic and bone agonistic properties. One of these, toremifene, is quite similar to tamoxifen although slightly less effective as an agonist to bone. On the basis of the SERM concept, other agents have been introduced or are being developed to enhance breast antagonistic and bone agonistic properties.

Pure Antagonist Antiestrogens

Clinical and experimental data suggested that long-term exposure to tamoxifen might induce tumors to undergo adaptive mechanisms to cause the agonistic properties of this SERM to predominate. On the basis of these observations, antiestrogens were developed that were devoid of agonist properties. Faslodex, the prototype drug, increases the rate of degradation of the ER and also inhibits estradiol-mediated transcription by favoring binding of corepressors to the Fastest-EOR complex.

Inhibitors of Estradiol Synthesis

Aromatase catalyzes the rate-limiting step in the conversion of androgens to estrogens. Aromatase has been a key target for development of inhibitors over the past 25 years. The first-generation inhibitor aminoglutethimide inhibited aromatase by 90% in postmenopausal women and was as effective as tamoxifen in causing breast tumor regressions. However, it blocked cortisol production and induced significant side effects. The efficacy of aminoglutethimide in postmenopausal women served as an impetus to develop second-generation and third-generation inhibitors.

Three third-generation agents (anastrozole, letrozole, and exemestane) are now approved drugs in the United States and Europe. These agents are 100-fold to 10,000-fold more potent than aminoglutethimide and are called selective aromatase inhibitors because they do not inhibit other enzymatic steps. The two major subclasses are nonsteroidal competitive inhibitors and steroidal enzyme inactivators. The competitive inhibitors bind to the active site of the enzyme with high affinity and compete with substrate. Aromatase inactivators bind covalently to the enzyme and permanently destroy its activity. Theoretically, the inactivators could have advantages over the competitive aromatase inhibitors because inhibition might continue if one missed one or more doses of medication and the degree of blockade could theoretically be superior.

Both classes of drugs reduce aromatase to 1% to 10% of baseline activity. Inhibitors substantially reduce plasma estradiol levels, and suppress tissue concentrations of this steroid in breast tumors. Because they lack estrogen agonistic properties, aromatase inhibitors do not increase the incidence of endometrial cancer, as occurs with tamoxifen. If they are used for long periods, one would expect acceleration of bone loss and perhaps cardiovascular disease as a result of total body aromatase inhibition. Studies evaluating these effects are ongoing. Preliminary data do not show adverse effects on lipid concentrations.
Chemotherapeutic agents destroy granulosa cells in the ovary and result in transient or permanent amenorrhea in premenopausal women. Complete ovarian destruction, as evidenced by the onset of amenorrhea, is more common in women older than 40 years as opposed to those younger (Fig. 39-9). Incomplete information is available regarding estradiol levels. The effect of chemical castration was not initially thought to be clinically important. However, there are data suggesting that adjuvant chemotherapy in premenopausal women exerts antitumor effects on receptor-positive tumors predominantly through an ovarian ablative effect.
Hormonal Therapies for Breast Cancer: Clinical Efficacy

Background

Nearly all patients with breast cancer undergo lumpectomy or mastectomy to remove the lesion and receive radiotherapy if it is warranted. High-risk patients then receive adjuvant therapy to destroy occult cancer cells at locoregional sites and distant micrometastases. Upon tumor recurrence, first-line, second-line, and third-line hormonal therapies or chemotherapies are then utilized for advanced disease.

Antiestrogens with Mixed Agonist-Antagonistic Actions

The overall benefit of tamoxifen is usually reported as the relative improvement in frequency of defined events such as tumor recurrence or survival. In both the adjuvant and advanced disease setting, tamoxifen results in a relative benefit of approximately 50%. However, a more useful parameter is the absolute benefit, defined as the number of women per hundred treated who benefit from taking tamoxifen. As illustrated in Figure 39-10, the more common the event, the more likely a woman is to benefit from tamoxifen.

When tamoxifen is used as adjuvant therapy, women with positive lymph nodes at the time of diagnosis experience approximately a 10% absolute reduction of tumor recurrence at 5 years and those with negative nodes experience a 5% reduction. Tamoxifen causes an absolute reduction of contralateral breast cancers of 2% at 5 years and prevents breast cancer in 1% to 5%, depending on the underlying risk (see Fig. 39-10). This antiestrogen is active in both premenopausal and postmenopausal women with breast cancer but only in those whose tumors are ER-positive or PR-positive, or both.

The optimal duration of use is 5 years. Direct comparative studies indicate superiority of 5 years versus 1 or 2 years of tamoxifen therapy, whereas 10 years of therapy provides no additional benefit over 5 years. Although tamoxifen is administered for only 5 years, benefit persists over the long term because disease-specific survival increases from approximately 6% at 5 years to 10% at 10 years in node-positive patients (Fig. 39-11).

Approximately 40% of women with advanced disease benefit from tamoxifen. Responses last 12 to 18 months on average before relapse. Efficacy is limited to those with tumors that are ER-positive or PR-positive, or both. Women with disease in soft tissue, bone, or viscera are the best candidates; chemotherapy is the more likely a woman is to benefit from tamoxifen. As illustrated in Figure 39-10, the more common the event, the more likely a woman is to benefit from tamoxifen.

Pure Antagonist Antiestrogens

Data regarding these agents are preliminary. A comparative study in 458 women with advanced disease demonstrated similar efficacy in patients treated with Faslodex and with anastrozole. Clinical benefit (i.e., complete objective response, partial objective response, and stable disease for 6 months) occurred in 44.5% receiving Faslodex and 45% receiving anastrozole. Further data are required to determine the relative place of Faslodex in the therapeutic armamentarium.

Aromatase Inhibitors

Previously, tamoxifen was considered the most effective hormonal agent for the treatment of breast cancer and progestins the next most effective. However, because of their potency and lack of side effects, the third-generation aromatase inhibitors (anastrozole, letrozole, and exemestane) are challenging these concepts for postmenopausal patients. Large and well-designed
trials have compared the relative efficacy of the third-generation inhibitors with that of other hormonal agents including tamoxifen.

Initial trials examined the efficacy of third-generation inhibitors in comparison with aminoglutethimide and demonstrated superior efficacy and side effect profiles with the newer agents. Next, head-to-head trials compared anastrozole and letrozole with megestrol acetate in the second-line setting. Anastrozole was associated with a greater duration of survival than megestrol acetate. Both anastrozole and letrozole were better tolerated than megestrol acetate. These findings moved the third-generation agents ahead of megestrol acetate into a role as second-line agents for advanced disease.

The next logical step involved a direct comparison with tamoxifen. Three large multicenter trials involving 1949 women compared anastrozole and letrozole with tamoxifen. All three followed a similar trial design and included postmenopausal patients with locally advanced or metastatic breast cancer. No women had received adjuvant tamoxifen within 12 months, and none were known to be ER-negative. In the North American trial, women experienced clinical benefit more frequently with anastrozole (59%) than with tamoxifen (45%) (P = .009). In the European trial, there were no differences in clinical benefit between therapies. The different results may be explained by the greater number of patients of unknown ER status in the European trial (95% versus 15%). When a post hoc, subset analysis of results for the ER-positive patients in the European trial was conducted, anastrozole again appeared superior to tamoxifen. The letrozole versus tamoxifen trial demonstrated superior effects in all.

parameters with objective response rates of 30% versus 20%, clinical benefit of 49% versus 38%, and median time to treatment failure of 9 versus 5.8 months (P < .0001 by log rank test). Pooled data indicate that tamoxifen was associated with DVT and pulmonary emboli (7.6% versus 4.5%) significantly more frequently than anastrozole. If one combines observations from all three trials, it appears that nausea, hot flashes, and gastrointestinal distress were comparable with either agent. A fourth but smaller study (238 patients) involving only ER-positive patients also demonstrated superiority of anastrozole over tamoxifen (P = .05).

Taken together, these trials provide evidence that the third-generation aromatase inhibitors are superior in efficacy and toxicity to tamoxifen in the advanced disease setting. Letrozole and anastrozole have now been approved in the United States as first-line therapy for breast cancer. These three large aromatase inhibitor trials showed for the first time in a head-to-head randomized trial that one endocrine therapy is superior to another. Before these studies, it was thought that the available endocrine therapies produced similar rates of response and could be distinguished only on the basis of side effects and cost. Direct comparisons of letrozole and anastrozole are now needed to determine whether one agent is superior to the other. This is particularly important because hormonal data suggest that letrozole may be more potent as an aromatase inhibitor than anastrozole.

At present, the choice of aromatase inhibitors over tamoxifen as first-line therapy for advanced disease is warranted but based on incomplete data. Comparisons of tamoxifen with the first-generation aromatase inhibitor aminoglutethimide demonstrated equal efficacy but a different pattern of cross-resistance. The aromatase inhibitors were efficacious if used after tamoxifen; in contrast, tamoxifen, when used after the aromatase inhibitors, appeared less effective. If this were true for the third-generation aromatase inhibitors, one might still wish to use tamoxifen as the first-line agent. However, preliminary data suggest that tamoxifen is effective after crossover from the aromatase inhibitor anastrozole. In a crossover comparison, 51% of 98 patients experienced clinical benefit from tamoxifen when used as a second-line agent after initial use of letrozole. By comparison, second-line letrozole after initial tamoxifen produced clinical benefit in 66% of 61 patients.

The aromatase inactivator exemestane is currently being tested as a first-line agent for treatment of advanced breast cancer. Preliminary data for 122 patients showed 44.6% objective responses (complete response and partial response) with exemestane and 14.3% with tamoxifen. Currently, exemestane is approved for clinical use in the United States as a second-line hormonal agent for treatment of breast cancer.

Androgens or estrogens are generally considered inferior to tamoxifen, aromatase inhibitors, and megestrol acetate. High-dose estrogen has been used sparingly since the advent of tamoxifen. However, a 20-year follow-up of a randomized comparison of tamoxifen with high-dose diethylstilbestrol (DES) reported a significant survival advantage in patients receiving the estrogen and lack of cross-resistance between these two therapeutic approaches. This surprising study suggests reconsideration of high-dose estrogen for selected patients.

Although still used in both the adjuvant and advanced disease settings, surgical oophorectomy may be replaced by use of GnRH agonists. Laparoscopic oophorectomy, however, may make this a more attractive approach. Prophylactic oophorectomy represented the first adjuvant endocrine therapy for breast cancer. Although it was initially thought to be ineffective, meta-analyses demonstrate a clear benefit with a 6% absolute survival advantage at 15 years for patients younger than 50 years of age. Because the receptor status of the patients in these trials was unknown, a large fraction of receptor-negative patients were probably included. One would expect better results if only hormone receptor-positive patients were so treated. In the advanced disease setting, surgical oophorectomy was clinically beneficial in approximately 50% of patients who were ER-positive or PR-positive, or both.

Emerging data support the use of medical oophorectomy in the adjuvant setting, and research emphasis focuses on this strategy. Demonstration of the effectiveness of prophylactic oophorectomy was surprising to the medical community and led to reconsideration of this approach, although with use of GnRH agonists rather than surgery. The goal of these studies has been to determine the efficacy of medical oophorectomy in patients who are exclusively ER-positive. Two studies in the advanced setting indicated that the GnRH analogues produce clinical effects similar to those induced by surgical oophorectomy. Accordingly, one would expect similar effects in the adjuvant setting. Preliminary results from a four-arm trial suggested that this may be the case. Benefit appeared to be greater in...
women younger than 45 years and with ER-positive tumors.

Further studies are required to determine whether the GnRH agonist approach is superior to the use of tamoxifen in this setting. Long-term studies have clearly shown the efficacy of tamoxifen as adjuvant therapy. Because four studies in the advanced disease setting suggested that tamoxifen is as effective as the GnRH agonists, these analogues may produce similar results in the adjuvant setting as well. Current studies are examining the use of GnRH agonists in comparison with chemotherapy in the adjuvant setting rather than with placebo (see "Chemical Castration"). Additional trials are needed to compare GnRH analogues with tamoxifen in this adjuvant setting.

Complete Estrogen Blockade

Medical oophorectomy alone has been compared with medical oophorectomy plus tamoxifen in the advanced disease setting. The rationale for the combined approach is to inhibit secretion of estradiol by the ovary and, at the same time, to block the action of residual estrogens of adrenal origin. This strategy, called complete estrogen blockade, appeared superior to tamoxifen alone in a single large trial and in a meta-analysis of four similar studies. The combined approach resulted in more frequent objective responses as well as improved progression-free and overall survival. This strategy has not yet been studied in the adjuvant setting and particularly in comparison with chemotherapy. Because premenopausal women in the United States are usually treated first with adjuvant chemotherapy, the approach of complete estrogen blockade needs to be explored in this setting.

Chemical Castration

Studies are ongoing to determine the mechanism of action of chemotherapy in premenopausal women. At least three possibilities exist: hormonal effects resulting from chemotherapeutic destruction of the ovary, direct cytotoxic effects on the tumor, or a combination of these two effects.

Eight trials have compared adjuvant chemotherapy with medical castration with and without tamoxifen in premenopausal women. The chemotherapy and hormonal therapies appeared to produce comparable antitumor effects. However, trends suggest that chemotherapy is more effective than medical castration in ER-negative patients and that medical castration is superior in ER-positive patients. These results suggest that at least part of the benefit of chemotherapy in premenopausal women results from chemical castration. These studies also provide further support for the benefit of medical oophorectomy in the adjuvant setting. The enhanced survival produced by chemotherapy compared with a placebo may be improved further by the effects of estriadiol deprivation.

Further support for the medical castration hypothesis comes from studies showing that chemotherapy is less effective in patients without complete cessation of menses. Although effects of varying chemotherapeutic regimens on ovarian function differ, nearly 100% of women older than 40 years experience permanent amenorrhea after adjuvant chemotherapy. From these data, it appears that comparisons of chemotherapy and endocrine therapy are not yet definitive. Further trials comparing medical oophorectomy alone with chemotherapy alone in receptor-positive, premenopausal patients are required. In the meantime, medical oophorectomy is a reasonable alternative to chemotherapy as adjuvant treatment for premenopausal women who are ER-positive or PR-positive, or both, and for various reasons are not candidates for chemotherapy.

Combination of Hormonal Therapy and Chemotherapy

On the basis of a meta-analysis, the combination of surgical or medical oophorectomy (GnRH agonists or tamoxifen) with chemotherapy provides additional benefit over chemotherapy alone in the adjuvant setting. The addition of the GnRH agonists might be particularly important for women younger than 40 years in whom estradiol remains in the premenopausal range after chemotherapy, as suggested by one trial.

New Approaches to Adjuvant Endocrine Therapy

Aromatase inhibitors may provide a means to block estrogen effectively without the emergence of estrogen agonistic effects in the adjuvant setting. Detrimental effects on vaginal mucosa, bone density, and cholesterol levels could result from deprivation of estriadiol.

A large ongoing trial is comparing tamoxifen alone, anastrozole alone, and the combination (Anastrozole and Tamoxifen Alone and in Combination [ATAC] trial). Preliminary data demonstrate statistically significant superiority of anastrozole with respect to recurrence-free survival and to number of new contralateral tumors. Subanalyses will compare potential detrimental effects on estrogen target organs such as bone, uterus, and liver. Other ongoing trials are examining sequential regimens with 2 to 5 years of tamoxifen and 2 to 5 years of an aromatase inhibitor with comparison of various inhibitors, dosages, and sequences. These results should be available in 3 to 5 years and should be helpful in suggesting optimal regimens.

Emerging Therapies

The concept of neoadjuvant chemotherapy has been adopted in clinical practice, and neoadjuvant hormonal therapy is now undergoing clinical trials. The rationale is to decrease tumor size before surgery to allow breast conservation. This type of therapy is selected for patients with tumors larger than 2 cm. A nonrandomized trial demonstrated an 81% reduction in tumor volume with letrozole versus 75% with anastrozole and 48% with tamoxifen. A randomized trial involving 324 patients compared letrozole at 2.5 mg daily with tamoxifen at 20 mg daily in women with ER-positive tumors larger than 2 cm. Letrozole resulted in a 55% rate of objective response (CR and PR) compared with 36% for tamoxifen (P < .001). Breast-conserving surgery was chosen by 45% of patients receiving letrozole and 35% receiving tamoxifen (P < .001). Approximately half of the patients experienced a sufficient reduction in size of the tumor to allow lumpectomy. Comparison of groups of patients treated conventionally and

with neoadjuvant endocrine therapy is required before recommendations regarding this approach are warranted.
Recommended Approaches to Hormonal Treatment of Breast Cancer

Adjuvant Therapy

Premenopausal Women

Current opinion recommends chemotherapy for premenopausal women with ER-positive tumors and the addition of tamoxifen for 5 years (Fig. 39-15) (Figure Not Available). A meta-analysis demonstrated statistically significant prolongation of survival with a combination of tamoxifen plus chemotherapy versus chemotherapy alone. Medical oophorectomy with or without tamoxifen could be considered as an alternative for women adverse to chemotherapy. 

Postmenopausal Women

Tamoxifen is the preferred adjuvant therapy for postmenopausal women with tumors larger than 1 cm that are ER-positive or PR-positive, or both. For some women, chemotherapy in combination with or followed by tamoxifen may be chosen. In younger postmenopausal women with more aggressive tumors, chemotherapy may be preferred but clinical trial data regarding this are incomplete. Information obtained from meta-analyses suggests that the combination of chemotherapy and tamoxifen may be preferable to use of tamoxifen alone, and this approach is often chosen. Early studies showed adverse drug-drug interactions between megalphan (L-Pam) and tamoxifen. However, this does not occur with CMF (cyclophosphamide [Cytoxan], methotrexate, fluorouracil) and CAF (Cytoxan, doxorubicin [Adriamycin], fluorouracil) regimens and chemotherapy and tamoxifen can thus be given concomitantly. The main consideration is whether the modest benefit accrued by adding chemotherapy is worth the additional toxicity encountered.

Small Tumors in Premenopausal and Postmenopausal Women

No randomized trial has examined the use of tamoxifen in women with tumors 1 cm or smaller in size. However, pooled data from four NSABP trials indicate a 4% absolute benefit in disease-free status and a 5% survival benefit for this group of women when given tamoxifen. Individual decisions are made on the basis of risk-benefit analysis, and not all women are offered this therapy.

Ductal Carcinoma in Situ

Adjuvant tamoxifen appears to provide benefit for women with noninvasive tumors (i.e., DCIS). In a large NSABP clinical trial, 13% of women treated with lumpectomy, irradiation, and placebo experienced a new tumor event over 5 years. One third of these new events involved appearance of a contralateral breast cancer, one third a new ipsilateral tumor, and one third local recurrence of the original tumors. Tamoxifen reduced these events in absolute terms by 5% with equal benefit in reducing new contralateral, ipsilateral, and original tumor events. Risk-benefit analysis is also required to advise patients appropriately; benefit is not judged to outweigh risks in all patients.

Decision Making

(See Fig. 39-15.) (Figure Not Available) An overview of these data suggests that all women with tumors that are ER-positive or PR-positive, or both, are potential candidates for tamoxifen as adjuvant therapy. In order to make a decision, one must determine whether the benefits of tamoxifen outweigh the risks in individual patients. From the NSABP prevention trial, the absolute risks of tamoxifen are known and include uterine cancer, cataracts, DVT, pulmonary emboli, and cerebrovascular accident. Presence of a uterus, a history of DVT or pulmonary embolus, or existing cataract would enhance these risks. Presence of larger or invasive tumors would enhance the absolute benefits of tamoxifen.

In general, most women with invasive tumors larger than 1 cm should be advised to take tamoxifen. Those with small or noninvasive tumors should be counseled on the basis of risk-benefit ratios. Those without a prior history of a thromboembolic event or cataract and with prior hysterectomy might be advised to take tamoxifen, particularly if they also have osteopenia. Raloxifene has not been used in the adjuvant setting and would not be advised for the patients discussed here.

Treatment of Advanced Disease

The majority of women who have advanced disease had previously received tamoxifen as adjuvant therapy. Those experiencing tumor recurrence while receiving tamoxifen or within 1 year of its cessation are considered resistant to this antiestrogen. For them, other hormonal therapies are chosen. In the other group of women, tumors recur more than 1 year after cessation of tamoxifen, and these patients are candidates for an additional course of tamoxifen. An algorithm can then be used to choose hormonal therapy for women considered candidates for tamoxifen (see Fig. 39-15) (Figure Not Available). For those resistant to tamoxifen, the next therapy in the sequential approach can be utilized.

Premenopausal Women

Aromatase inhibitors do not effectively inhibit ovarian estrogen production in premenopausal women, and tamoxifen (or toremifene) is usually considered first-line therapy for recurrent tumors. On the basis of current data, this might be combined with a GnRH analogue. If the initial therapy is tamoxifen alone, medical oophorectomy with use of a GnRH agonist analogue as second-line therapy would be recommended. Surgical oophorectomy can be substituted for the GnRH analogue. Aromatase inhibitors can be used if the GnRH analogue is continued. Megestrol acetate is then advised as additional therapy.

Postmenopausal Women

Data from trials comparing aromatase inhibitors with tamoxifen as first-line therapy for advanced disease suggest that aromatase inhibitors be considered the first choice of endocrine therapy. Responders would be treated with tamoxifen as second-line therapy upon relapse, but nonresponders would be treated with chemotherapy. Third-line therapy would utilize megestrol acetate and fourth-line therapy the aromatase inactivator exemestane. After this, high-dose estrogen or androgens might be chosen. If rapidly progressive disease develops at any time in this sequence, chemotherapy may be chosen instead of the next endocrine therapy.

Chemotherapy

The detailed management of breast diseases with chemotherapy has been addressed in the work of Ellis and colleagues.
Long-Term Quality of Life in Breast Cancer Survivors

With earlier diagnosis of breast cancer, an increasing percentage of women survive breast cancer in the long term. Two thirds of these women are menopausal at the time of diagnosis, and half of the premenopausal women undergo permanent ovarian failure as a result of chemotherapy. HRT is generally thought to be contraindicated in these women because estrogens may cause regrowth of residual tumor tissue after surgery or cause a second primary. Data from observational studies, however, do not provide evidence of a deleterious effect. Study of the safety of HRT in this setting is currently undergoing testing in a moderately large group of women in the Habits trial (a large, multi-institutional Scandinavian and European trial coordinated by Dr. Lars Hølmberg, Uppsala University, Sweden). Until the results are available, it is prudent to avoid estrogens in breast cancer survivors if alternatives to estrogen are effective.

As reviewed previously, effective agents are available to substitute for estrogen, including the following:

1. Bisphosphonates to prevent or treat osteoporosis.
2. Statins to prevent heart disease.
3. Selective serotonin reuptake inhibitors (SSRIs) to diminish the number and severity of hot flashes.
4. Low-dose vaginal estrogen for symptoms of urogenital atrophy.
5. SSRIs for depression thought to be related to estrogen deficiency.

Alternatives to HRT would not protect against Alzheimer’s disease or improve cognitive function; however, the benefits of estrogen to protect against Alzheimer’s disease have not been proven in randomized prospective studies and estrogens have not shown cognitive benefit in patients with Alzheimer’s disease. Yet, estradiol does provide cognitive benefit in non-Alzheimer’s patients. Some women, then, might wish to consider the use of HRT to prevent Alzheimer’s disease or improve cognitive function. If alternatives to estrogen are not satisfactory, women could receive HRT after a full discussion of the risks and benefits and with the informed consent of the patient.
Breast Cancer in Men

The incidence of breast cancer in men is 100-fold lower than in women with 1400 new cases in 2000 and 400 deaths. Risk factors include clinical disorders associated with reduced testosterone production or estrogen excess such as orchitis, undescended testes, testicular injury, Klinefelter's syndrome, chronic liver disease, and gynecomastia. The breast cancer susceptibility genes \textit{BRCA1} and \textit{BRCA2} are also believed to increase the risk in men. Breast cancer is suspected when a subareolar or upper outer quadrant, firm painless lesion is palpated. A diagnosis is then made by mammography and biopsy of the lesion. Most tumors are ER-positive, and patients are then treated by mastectomy and adjuvant tamoxifen. Later therapy may include aromatase inhibitors or progestins.
ENDOMETRIAL CANCER

Etiology

Approximately 36,000 new cases of endometrial cancer occur annually in the United States, and 6500 die of this disease. Relative risk factors are related to conditions causing overexposure to endogenous or exogenous estrogen. Enhanced tumor initiation and promotion mediated by the proliferative effects of estrogen occur as with breast cancer (see earlier). Progestins are antimitogenic to the endometrium, as opposed to breast tissue, and result in a decreased rate of cell proliferation. Thus, unopposed estrogen increases the risk of endometrial cancer and progestins reduce that risk.

Animal studies with raloxifene suggested that this SERM does not exert estrogen agonistic effects on the uterus, whereas tamoxifen does (see earlier). It is interesting that raloxifene, another SERM, did not cause an increase in endometrial cancer in a large study monitoring this as a safety parameter.

Use of tamoxifen as adjuvant therapy for breast cancer or prevention is associated with an increased risk of endometrial cancer as well as hyperplasia and polyps. Combination oral contraceptives containing both estrogen and a progestin decrease the risk. Multiple pregnancies also increase the duration of exposure to large amounts of estrogen, and use for less than 5 years, there was no increase in endometrial cancer risk. There is a trend, however, for an increase in endometrial cancer with longer term use of this continuous regimen. Beresford and co-workers reported a relative risk of 2.7 (95% CI 1.2 to 6.0) for current users of this regimen for more than 5 years, whereas Pike's group reported a relative risk of 1.09 (nonsignificant).

Caution should be advised concerning long-term use of such regimens. Sequential regimens in which a progestin is administered for less than 10 days are associated with an increase in risk of endometrial cancer. Pike and colleagues reported an increase in relative risk of 1.87 per 5 years of use (95% CI 1.33 to 2.65), and Beresford and co-workers reported a relative risk of 2.1 (0.9 to 4.7) for less than 5 years of use and 4.8 (95% CI 2.0 to 11.0) for more than 5 years of use. Taken together, these data suggest that progestins do not protect against estrogen-induced endometrial cancer unless taken continuously. In premenopausal women, oral contraceptives containing both estrogen and a progestin decrease the risk. Multiple pregnancies also increase the duration of exposure to large amounts of progesterone and decrease the risk of endometrial cancer.

Use of tamoxifen as adjuvant therapy for breast cancer or prevention is associated with an increased risk of endometrial cancer as well as hyperplasia and polyps (see earlier). It is interesting that raloxifene, another SERM, did not cause an increase in endometrial cancer in a large study monitoring this as a safety parameter. Animal studies with raloxifene suggested that this SERM does not exert estrogen agonistic effects on the uterus, whereas tamoxifen does.
Endocrinology of Endometrial Cancer

Most endometrial cancers contain appreciable levels of ER, whereas only differentiated ones generally have PRs. Tumors resulting from estrogen replacement therapy are generally well differentiated, have ERs and PRs, and are of low grade and stage. The diagnosis is suspected when unexplained vaginal bleeding is detected. Instillation of saline into the uterine cavity followed by ultrasonography can reveal an area of focal thickening that is shown by biopsy to be cancer. An associated finding is generalized thickening of the endometrial stripe to greater than 6 mm as a sign of concomitant endometrial hyperplasia. Any unexplained vaginal bleeding in a postmenopausal women requires such evaluation to rule out endometrial cancer.

Treatment

Treatment requires initial hysterectomy and bilateral oophorectomy in all patients. Patients at risk for vaginal spread are pretreated with brachytherapy implants in the vagina. Those with a poor prognosis (25% of patients) are treated postoperatively with brachytherapy implants or external beam radiotherapy. Upon recurrence, high doses of systemic progestagens may be used. Response to therapy is independent of age, site of metastasis, or previous or concurrent therapy. Two large gynecologic oncology studies reported that objective responses to a progestin occurred in 24% to 25% of patients.

The exact mechanism of tumor regression is unknown but may involve the following:

1. Direct effects on tumor cells.
2. Stimulation of the inactivating 17-hydroxysteroid dehydrogenase type I enzyme, which converts estradiol to estrone.
3. Inhibition of the production by the adrenal glands of androgenic estrogen precursors.
4. Down-regulation of estrogen receptors by progestagen.
5. Suppression of gonadotropin production in premenopausal women.

Experimental trials are ongoing to test the efficacy of aromatase inhibitors, antiestrogens, GnRH analogues, and combinations of these agents. Various chemotherapeutic regimens are also available for such patients.

Hormone Replacement Therapy in Endometrial Cancer Survivors

Patients with an excellent prognosis and disease-free survival for at least 1 year can be treated with HRT to relieve menopausal symptoms.
PROSTATE CANCER

Incidence

In the United States, 198,000 new cases of prostate cancer and 31,500 deaths were predicted in the year 2001. Black men have an age-adjusted relative risk of 1.73 (95% CI 1.23 to 2.45) compared with white men. The mortality of black men is nearly twice that of white men. Introduction of prostate-specific antigen (PSA) screening in the mid-1980s resulted in a tripling of prostate cancer detection rates between 1985 and 1997 from 96,000 per year to 334,500 per year. However, estimated case detection rates have gradually declined to 198,000 as the pool of previously undiagnosed patients has diminished. With the advent of PSA screening, 1 in 6.25 men will be diagnosed with prostate cancer during their lifetimes. As a result of PSA screening, 75% of patients present with organ-confined disease compared with only 25% before PSA screening.
Etiology

Environment, Diet

Environmental and dietary factors play a major role in prostate cancer etiology, as evidenced by epidemiologic data. Global incidence figures vary widely from 1.08 per 100,000 per year in China to 190 per 100,000 per year in the United States. Dietary or environmental explanations for this wide variance are likely. Japanese men who move to the United States have rates of prostate cancer that ultimately approach those of United States citizens. Differences in ingestion of high-fat diets, green tea, or soy products provide potential explanations for the divergent rates of prostate cancer among different populations. The incidence of premalignant prostate lesions and latent prostate cancer is the same in Japan and China as in the United States, whereas the rates of invasive cancer differ markedly. The additional mutations or promotional factors necessary to convert latent to invasive cancer apparently occur less commonly in Japanese and Chinese men living in Asia. In contrast, the initial mutation or mutations leading to latent cancer occur at the same rate. This observation suggests that environmental or dietary factors may influence the later additional mutations or tumor promotion more specifically.

Hormonal and Genetic Factors

Hormonal factors and particularly circulating androgens probably play a role in the initiation and promotion of prostate cancer. Prostate cancer rarely develops in men surgically orchiectomized before the age of 30 years, although this has been difficult to document in the published literature (P. Gann, Fang-Liu Gu, Peking University, personal communications). Genetic factors play an important role as well. Twin studies suggest a genetic component in 44% of patients with prostate cancer. Men with prostate cancer report a family history of this tumor 3.1 to 4.3 times more commonly than healthy men. The relative risk of prostate cancer is increased approximately twofold in men with one first-degree relative with prostate cancer. Specific genetic lesions resulting in prostate cancer are uncommon. The BRCA2 gene probably accounts for a small percentage of prostate cancer cases. An increase in the number of glutamine repeats in the variable region of the androgen receptor from germ line cells occurs in 10% of patients with prostate cancer. Data from a Utah pedigree identified another high-risk gene on chromosome 17p called ELAC2. Ross and associates suggested that the lower frequency of prostate cancer in Japanese and Chinese populations may be related to a lower level of 5-reductase activity than in their white counterparts, perhaps on a genetic basis. This may lead to lower levels of dihydrotestosterone (DHT) in prostatic tissue and less androgen-induced proliferation. However, direct isotopic kinetic measurements of 5-reductase activity in Chinese versus white men demonstrated no differences in the levels of this enzyme.
Early Case Detection

Until the introduction of PSA measurements, digital rectal examination (DRE), followed by biopsy, represented the standard screening procedure. Currently, PSA measurements detect cancers at a time when most are not palpable by DRE. The principle behind PSA screening is that tumors release more PSA into the bloodstream per gram of tissue than does tissue with benign prostatic hyperplasia or normal tissue. \[340\] PSA may be transiently increased by prostatitis, after endoscopic urethral manipulation, by prostatic biopsy, or to a lesser extent by ejaculation. Routine DRE has a minimal effect on PSA levels, but most physicians defer PSA testing until several days after this examination. \[341\]

The sensitivity of case detection with PSA is high but the specificity is low, and routine PSA screening has been controversial. In addition, there may be no advantage in detecting "latent" prostate cancers. \[347\] Attempts to increase specificity include use of age-related PSA normal ranges and PSA density (i.e., PSA divided by ultrasonographically determined prostate volume). Other methods to enhance specificity include PSA velocity (rate of increase in PSA over time), and percentage of free PSA. \[348\]

When free PSA was used in patients with borderline PSA values of 4.0 to 10.0 and a normal DRE, the rate of biopsy-proven prostate cancer increased from 8% to 20% to 56% with free PSA fractions of more than 25%, 15% to 20%, and 0% to 10%, respectively. \[351\]

Various professional societies have provided guidelines for PSA screening, but no general agreement exists. \[361\] Randomized trials to determine whether screening improves survival are not yet conclusive, \[363\] although one study demonstrated benefit. \[364\] A commonsense approach suggests screening only when test results would dictate diagnostic and therapeutic decisions. The patient should understand the consequences of a positive result and be willing to proceed with further diagnostic and therapeutic measures if the PSA findings are positive. \[364\] Accordingly, patients selected for screening should be those for whom definitive treatment or hormonal manipulations, and not watchful waiting, would be the likely choice if cancer were detected. \[364\] On the basis of this reasoning, informed consent of the patient is required before embarking on screening with PSA.
Evaluation of Abnormal Prostate-Specific Antigen or Digital Rectal Examination Findings

If the PSA is elevated to greater than 10 ng/mL or the DRE is abnormal, transrectal ultrasound-guided biopsy is indicated. An algorithm for those with PSA values between 4 and 10 has been described. Common practice involves obtaining biopsy specimens of ultrasound-detected hypoechoic lesions as well as "blind" biopsies.

Clinical staging of prostate cancer provides a means of determining the prognosis and is used for making treatment decisions. Two analogous systems have been used in the past, but the Tumor-Node-Metastasis (TNM) classification is now preferred.
Endocrinology of Prostatic Cancer Growth

DHT, the 5-reduced product of testosterone, binds to androgen receptors with 2.5-fold higher affinity than testosterone itself and serves as the major regulator of prostatic tumor growth. Direct effects of androgen, indirect effects induced by stimulation of growth factors, or a combination of these two mechanisms may mediate androgen-induced proliferation.\(^\text{[371]}\) Approximately 7000 µg of testosterone is secreted daily by the testes, of which 500 µg is converted into DHT in various peripheral tissues.\(^\text{[371]}\) The adrenal gland provides an additional 5% of the androgen produced in adult men.\(^\text{[372]}\) Testosterone as well as preandrogens such as androstenedione, dehydroepiandrosterone (DHEA or alternatively DHA), and DHEA sulfate (DHEAS) originate in the adrenal gland and are also converted in peripheral tissues into DHT. In addition to peripheral conversion, a large fraction of the DHT present in benign and malignant prostatic tissue is produced locally in the prostate gland from circulating precursors. Approximately 40% of prostatic DHT originates from steroids of adrenal origin.\(^\text{[373]}\)

Prostate cancer cells contain androgen receptors that bind DHT and transmit proliferative signals in androgen-dependent prostate cancer.\(^\text{[374]}\) As opposed to the use of receptor assays in breast cancer, measurement of the androgen receptor does not provide predictive information regarding hormone responsiveness. Mutations of the androgen receptor occur, but the frequency in primary prostate cancer is controversial.\(^\text{[375,380]}\) An early study found a 30% incidence of androgen receptor mutations in primary prostate cancers.\(^\text{[376]}\) Others have found a much lower frequency ranging from 0% to 5%. All investigators found mutations in metastatic disease ranging from 21% to 50%.\(^\text{[377]}\) The frequency and type of mutation appear to be influenced by the selective pressure exerted by antiandrogens. For example, 5 of 16 men receiving flutamide had mutations compared with 1 of 17 not receiving flutamide.\(^\text{[380A]}\) Most of these cases involved a mutation at amino acid 877, the site involved in LnCAP cells (a type of prostate cancer cell model named for Ln [lymph node] Ca [carcinoma] P [prostate]), which allows flutamide to become an androgen agonist. This finding may explain the occurrence of flutamide withdrawal responses in men with prostatic cancer (see later).\(^\text{[378,380]}\)

\[\text{Figure 39-17} \quad \text{Diagrammatic representation of the endocrinology of prostate cancer growth. Androgens are directly secreted by the testes into the circulation, are synthesized in peripheral tissues from steroidal precursors, and are formed directly in the prostate gland. AR, androgen receptor; DHA, dehydroepiandrosterone; DHT, dihydrotestosterone; HRE, hormone response element. (From Denis LJ, Griffiths K. Endocrine treatment in prostate cancer. Semin Surg Oncol 2000; 18:5254. Published with permission of Seminars in Surgical Oncology.)}\]
Decisions Regarding Treatment

One of three strategies can be chosen for treatment of prostate cancer: watchful waiting, hormonal manipulation, or definitive (curative) therapy. The first decision after a diagnosis of prostate cancer is whether to recommend therapy or to advise watchful waiting. The underlying principle in decision making is that some prostate cancers are latent. The term latent prostate cancer has been applied to small lesions, generally smaller than 0.5 cm in volume, that do not progress and do not result in cause-specific death.\(^3\)

In older men with other medical conditions, the risks of definitive therapy may outweigh the benefits, particularly if latent disease is present. Only a small percentage of men with limited disease die of prostate cancer each year after diagnosis (Fig. 39-18), and those with a limited expected life span because of other diseases may die with prostate cancer and not because of it. In general, watchful waiting is reserved for patients in older age groups (older than age 70 years) and those with a low stage and grade, localized disease, and a life expectancy of less than 10 years.\(^3\) A reasonable working construct is that the younger and more healthy the patient, the more likely the benefit from definitive therapy. Prognostic factors that can be used to assess aggressiveness include tumor stage and tumor grade, as assessed by the Gleason score, and PSA.

Several competing approaches have been developed for definitive therapy. Radical prostatectomy with nerve sparing is one option. A risk stratification tool developed to assess the likelihood of recurrence after this procedure includes the initial PSA level and Gleason score. Low-risk patients (initial PSA < 10 and Gleason score equal to or less than 6) had a 5-year relapse-free survival of 81% versus 40% for high-risk patients (PSA > 10 and Gleason score greater than or equal to 7). Conventional radiation, three-dimensional conformal radiation, or brachytherapy techniques provide other alternatives. Patients are informed of the various treatment modalities, and a participatory decision is made on the basis of the best expected outcome. The younger and healthier the patient, the more reasonable is the recommendation of radical prostatectomy. An experimental therapeutic option for locally advanced disease that avoids surgery and radiation involves monotherapy with an antiandrogen. The concept is that hormonal therapy shrinks the primary tumor and prevents metastatic spread. Preliminary reports of three exploratory studies of antiandrogens in the setting of previously untreated locally advanced disease have appeared. These studies compare antiandrogens not with prostatectomy or radiation but with castration. Pooled mature data suggested no survival difference between bicalutamide at 150 mg daily and surgical castration in this setting.\(^1\)

Another large study recruited 1453 men with confirmed metastatic disease or T3 to T4 disease with elevated PSA into two trials comparing castration with bicalutamide. Pooled data demonstrated decreased efficacy of bicalutamide with respect to survival (hazard ratio 1.3). Taken together, these data suggest that bicalutamide is not as effective as castration. Even though antiandrogen monotherapy is not as effective as castration, some men may choose this therapy in preference to watchful waiting.
Neoadjuvant Hormonal Therapy

Neoadjuvant therapy is defined as use of hormonal therapy before definitive treatment and involves (1) medical castration with GnRH analogues, (2) steroidal or nonsteroidal antiandrogens, or (3) a combination.

Radical Prostatectomy

Neoadjuvant strategies used for 3 to 6 months prior to radical prostatectomy resulted in an increase in organ-confined cancers and decrease in positive surgical margins. However, no differences in PSA-detectable relapse rates have been observed in several trials. Data are not yet sufficiently mature to assess overall survival rates. However, the efficacy of this maneuver is still undetermined because longer follow-up is needed before definitive conclusions regarding overall survival can be reached.

Radiation Therapy

Radiation Therapy Oncology Group (RTOG) trial 86-10 examined the role of neoadjuvant endocrine therapy prior to radiation. The incidence of local progression at 5 years was 46% for patients with T2 to T4 tumors receiving adjuvant hormonal therapy and 71% for those receiving radiation alone \((P = .001)\), but no difference in overall survival was noted.
Adjuvant Hormonal Therapy

Therapy given with the initial definitive treatment and before evidence of recurrence is termed adjuvant hormonal treatment. Two large multicenter trials provided evidence of the efficacy of this approach. An RTOG study involved patients with stage T3 and stage T4 disease given radiation therapy followed by medical castration with goserelin starting in the last week of radiation. Disease-free survival at 5 years was 60% for the goserelin arm and 44% for radiation alone (P < .0001). Only the subset of men with Gleason grade 8 to 10 tumors experienced a survival benefit from 55% to 66% (P = .03).

In a similar trial of the European Organization for Research and Treatment of Cancer (EORTC), the 5-year disease-free survival was 85% in the adjuvant hormone-treated group (goserelin) versus 48% in the group given radiation alone (P = .001). Overall survival was 79% for the adjuvant hormonal group versus 62% for the radiation-alone group, and this difference was statistically significant (P = .001).

The majority of data favor the use of adjuvant hormonal therapy in patients undergoing radiation therapy as definitive treatment for prostate cancer. An analysis of pooled data from the RTOG neoadjuvant and adjuvant trials concluded that the key consideration may be the necessity for long-term rather than short-term hormonal therapy in either the neoadjuvant or adjuvant therapy approach. In men receiving short-term hormonal therapy (goserelin and flutamide for 2 months before and 2 months after radiotherapy), statistically significant improvements were observed for two end points, distant metastasis-free and evaluable disease-free intervals, but not for overall survival. Only long-term adjuvant therapy improved overall survival and only in the subset of men with Gleason grade 7 to 10 tumors. These conclusions concerning long-term versus short-term adjuvant hormonal therapy were supported by a nonrandomized Canadian trial.

In a large ongoing study, men are randomly assigned to placebo or 150 mg of bicalutamide daily after radical prostatectomy or radiotherapy or during watchful waiting. A total of 8055 patients have been entered. Bicalutamide significantly reduced the incidence of disease progression by 42% (hazard ratio 0.58, 95% CI 0.51 to 0.66, P < .0001). This encouraging result needs to be confirmed with respect to overall survival.
**Immediate versus Delayed Hormonal Therapy**

Approximately 25% of men now present either with metastatic (M+ or stage D) disease or with local spread into lymph nodes (N+). Whether to initiate hormonal treatment immediately in these patients or to defer treatment until symptoms occur is a major question. Because cure with hormones is not possible, the two goals of treatment are to increase life expectancy and to relieve symptoms. If treatment is advocated for asymptomatic patients, it should improve their length of survival.

Rigorously controlled series from the Veterans Administration Cooperative Urological Research Group (VACURG) trials during the 1960s indicated no clear survival benefit from endocrine therapy in patients with stage C or D disease at the time of diagnosis. On the basis of these findings, standard practice usually involved withholding hormonal therapy until symptomatic metastatic (stage D) disease was present. However, two later studies provided evidence favoring therapy immediately upon detection of metastatic disease.

Messing and colleagues studied men with prostate cancer found to have positive pelvic lymph nodes at the time of radical prostatectomy. They randomly assigned these men to an immediate therapy group (medical or surgical orchiectomy) or to a group observed until an indication occurred. Death from prostate cancer occurred in 3 of 47 men in the immediate treatment group and 16 of 51 in the deferred group (P < .01).

In another study, 934 men with locally advanced prostate cancer or with asymptomatic metastatic disease were randomly assigned to immediate or deferred hormonal therapy. There were 203 deaths from prostate cancer in the immediate therapy arm and 257 in the deferred group (P = .02). There were fewer cases of spinal cord compression (9 versus 23) and pathologic fracture (11 versus 21) in the immediate treatment group.

Other evidence of efficacy comes from nonrandomized studies from the Mayo Clinic that demonstrate a statistically significant survival advantage for early endocrine therapy (castration) in men with stage D1 (T0 to T3, N1 and N2, M0) disease. Seventy-three patients underwent either radical retropubic prostatectomy or radical prostatectomy plus orchiectomy. An advantage of immediate adjuvant therapy was demonstrated in that 5-year survival rates were 93% in the immediate orchiectomy group and 80% in the deferred orchiectomy group. In another study, a benefit was seen only in patients with diploid and not in those with tetraploid or aneuploid tumors.

Finally, a subset analysis in the RTOG 85-31 study examined a group of 139 men with capsular penetration or seminal vesicle involvement. Seventy-one men received radiation therapy plus luteinizing hormone-releasing hormone (LHRH) agonist therapy and 68 received radiation therapy alone. A statistically significant improvement in progression-free survival and freedom from biochemical relapse (i.e., PSA) was observed in the LHRH group. Overall survival was not statistically significantly different in this relatively small subset of men with follow-up for only a median of 5 years.

Taken together, these data suggest, but do not prove, that immediate hormonal therapy may be efficacious for asymptomatic men presenting with lymph node spread. The data at present do not provide definitive evidence that early endocrine therapy is beneficial regarding survival of patients presenting with metastatic (M+) disease, but evidence from one study is suggestive.

Men who experience a rising PSA (biochemical failure) after an initial fall to undetectable levels following radical prostatectomy may also benefit from hormonal therapy. Data indicate that a rapid rise in PSA suggests the presence of metastatic disease, whereas a slow rise suggests local recurrence.

In a single small study, 68% of men progressed to detectable clinical disease upon observation for a median of 19 months. With adjuvant hormonal therapy or radiotherapy, the rate of progression to clinically detectable disease decreased to 21%. Use of the capromab pendetide (ProstaScint) radioisotopic scan has been advocated to identify patients with distant metastatic disease under these circumstances. Such patients would not be considered candidates for salvage radiotherapy to the pelvis.
Choice of Endocrine Treatment for Initial Disease

Which endocrine therapy to use initially is a major question. Surgical orchietomy produces a rapid decrease in serum androgen levels, does not require the patient's long-term compliance, and is effective in inducing tumor regression in nearly 90% of patients. The clinician can be assured that testicular androgens are completely suppressed. However, nearly 50% of men in the United States prefer medical castration \[434\] as a means to avoid surgery.

Highly potent agonist analogues of GnRH, called superagonist analogues, have been approved for use in prostate cancer and effectively produce a medical orchietomy. These compounds paradoxically inhibit LH secretion by the pituitary gland and thereby suppress testicular testosterone production. Clinically, the GnRH agonists stimulate LH threefold to fourfold and testosterone twofold for 1 to 2 weeks from initiation of therapy. Thereafter, LH is profoundly depressed and the plasma testosterone level falls from approximately 500 ng/dL to a castration level of 15 ng/dL. No escape from inhibition occurs for up to 2 years of continuous therapy.

The initial rise in testosterone causes a transient disease flare in 5% to 10% of patients. This flare produces an objective increase in tumor size in approximately 3% of patients and a subjective increase in bone pain in the remainder. Although tumor flare \[395\] \[397\] is usually transient, severe reactions with spinal cord compression or death have been observed in occasional patients. For this reason, it is necessary to administer an antiandrogen for the first month after starting GnRH agonist therapy. \[394\] GnRH agonist analogues are now in clinical trial as a means to abrogate this problem. \[435\]

The rationale for using GnRH agonists is to induce a medical castration selectively and without unwanted side effects. Testosterone and DHT levels in patients treated with GnRH agonists and with surgical orchietomy fall to a similar extent. \[429\] Hormonal effects with the GnRH agonists are selective with no alterations of adrenal, thyroid, parathyroid, or pancreatic function. Objective regressions occur as frequently as with orchietomy. The efficacy of medical castration is equal to that observed with surgical castration or use of DES. \[397\] \[430\]

Initially, a major problem with GnRH agonist therapy was the requirement for daily subcutaneous administration with the possibility of noncompliance with daily injections and incomplete androgen suppression. \[395\] Third-generation formulations are now available that allow injections at 3- or 4-month intervals. These biodegradable preparations appear highly effective and are well tolerated and acceptable to patients. The frequency of tumor regression or stabilization does not appear to differ in patients treated with orchietomy or long-acting GnRH agonists. \[398\] Because some androgen-responsive cells remain in prostate tumors after relapse with the GnRH agonists, continued GnRH agonist treatment is advocated when disease relapse occurs and preferably for the remainder of the patient's life.

Monotherapy with antiandrogens provides an alternative to surgical or medical castration. These agents can be subclassified as steroidal and nonsteroidal antiandrogens. The nonsteroidal agents, flutamide and bicalutamide, bind to androgen receptors and block the cellular effects of circulating testosterone and DHT on cell proliferation. Interruption of the androgen negative feedback system results in reflex increments in serum LH, testosterone, and DHT levels. Studies with antiandrogens suggested that these agents preserve erectile function. More recent data suggest that only 20% maintain morning erections and sexual activity, \[397\] but this is still higher than after castration (18% reduction with bicalutamide versus 37% with castration). \[395\] Side effects include hot flashes and diarrhea, particularly with flutamide. \[394\] Long-term effects on bone density require more detailed study.

Randomized trials have compared antiandrogens as monotherapy with medical or surgical castration. \[395\] \[397\] Initial studies showed that 50 mg of bicalutamide is not as effective as orchietomy. \[395\] \[397\] Later studies compared 150 mg of bicalutamide with medical or surgical castration in patients with locally advanced (T1 to T4, N+, M0) and metastatic (T1 to T4, N+, M0) disease. Bicalutamide was equivalent to medical or surgical castration for locally advanced \[399\] but not metastatic disease (hazard ratio for survival of 1.3). \[397\]

Another study examined whether 150 mg of bicalutamide was the equivalent of maximal androgen blockade (MAB) in the setting of locally advanced or M1 disease. In this study of Boccardo and colleagues, \[395\] bicalutamide was as effective as MAB for locally advanced (T1 to T4, N+, M0) disease but not for metastatic disease (M0). Taken together, these data suggest that antiandrogens as monotherapy may not block androgen effects as effectively as medical or surgical castration.

Administration of high-dose estrogen in the form of DES has been used as a treatment for prostate cancer since the 1940s. In the VACURG studies, 5 mg of DES decreased the rate of recurrence of prostate cancer but increased the cardiovascular death rate. \[395\] \[397\] After this observation, careful doseresponse studies indicated that 3 mg of DES daily minimizes the risk of cardiovascular disease acceleration and maximizes the beneficial effects on prostate cancer. Gynecomastia and impotence are the major side effects.

Comparison of these monotherapies was the subject of a meta-analysis involving 10 separate trials. It was concluded by the Technology Evaluation Center, an agency of the Agency for Health Care Policy and Research, \[394\] that orchietomy, GnRH agonist analogues, and DES produced equivalent survival in men and that no differences existed among the various GnRH analogues that are available.
Complete Androgen Blockade

A more comprehensive strategy for the endocrine treatment of prostate cancer has been proposed. The rationale for complete, or maximal, androgen blockade (CAB, complete androgen blockade; MAB, maximal androgen blockade) rests upon three considerations:

1. The adrenal glands contribute 5% of the total androgen pool.\(^{(372)}\)
2. The concentrations of DHT in prostate cancer tissue of patients fall by only 50% to 80% after surgical orchiectomy\(^{(373)}\) and DHT levels in that tissue after castration are still higher than in nonandrogen target tissues (Fig. 39-20).\(^{(427)}\)
3. In vitro systems demonstrate that some tumor cell clones are hypersensitive to the proliferative effects of androgen.\(^{(441)}\)

On the basis of these considerations, inhibition of both testicular and adrenal androgens (MAB) might be more effective than inhibition of testicular androgens alone (testicular androgen suppression).

Forty-six studies have examined the concept of MAB as initial endocrine therapy for advanced disease (stage D or Tx Nx M+ disease) in randomized trials, but the results and conclusions drawn from the studies are conflicting.\(^{(371)}\)\(^{(442)}\) However, a meta-analysis involving nearly 90% of men treated worldwide provided several definitive conclusions and reasons for discrepancies among prior results:\(^{(443)}\)

1. MAB regimens using nonsteroidal antiandrogens provide a 2.9% improvement in overall survival from 24.7% to 27.6% at 5 years (Fig. 39-21).\(^{(446)}\) This difference is statistically significant (\(P = .005\)) but of marginal significance clinically.
2. MAB regimens using the steroidal antiandrogen cyproterone acetate produce adverse survival effects compared with testicular androgen suppression.
3. Analyses of studies of MAB that pool patients treated with either cyproterone acetate or nonsteroidal antiandrogens demonstrate no survival benefit produced by MAB over testicular androgen suppression.
4. Men with disease limited to the axial skeleton did not gain more benefit from MAB than those with metastases to the appendicular skeleton. A large prior study had reported this to be the case.\(^{(444)}\)
5. Results were similar when subsets of men treated with surgical or medical castration or with flutamide or nilutamide were examined.

Past controversies about MAB can be explained by fact that some trials used cyproterone acetate, an antiandrogen that shortens survival in these patients.\(^{(443)}\) Other trials did not use a short-term antiandrogen in the medical castration group, and the disease flare may have compromised overall efficacy. Finally, early trials used daily GnRH injections, which may not have induced effective medical castration because of the need for compliance with daily injections for 5 years.

It should be noted that no properly designed study has compared MAB with sequential androgen blockade (defined as initial castration followed by antiandrogen upon relapse). In the largest study of MAB,\(^{(445)}\) the use of antiandrogen upon relapse in the placebo group was left to the investigator’s discretion and only 50% of men in the placebo arm later took an antiandrogen. For this reason, there is no information about whether antiandrogen therapy given at the time of relapse after medical or surgical orchiectomy (i.e., sequential androgen blockade) is as effective as MAB. This is an important issue because flutamide given upon relapse after castration (sequential androgen blockade) resulted in objective benefit in 23% of patients.\(^{(446)}\)

On the basis of the data presented and the additional considerations discussed, this author and others favor the sequential androgen blockade strategy and use of antiandrogens only after relapse from medical or surgical orchiectomy.\(^{(445)}\) However, if GnRH agonists are used to produce a medical orchiectomy, short-term use of an antiandrogen to prevent disease flare is necessary.\(^{(447)}\)\(^{(448)}\)

![Figure 39-20](image-url) Prostate tissue levels of dihydrotestosterone (DHT) in the normal prostate gland and after surgical orchiectomy alone and in combination with suppression of adrenal androgens. n.d., nondetectable. (From Denis LJ, Griffiths K. Endocrine treatment in prostate cancer. Semin Surg Oncol 2000; 18:5274. With permission of the publisher.)

![Figure 39-21](image-url) Results of complete versus partial androgen blockade in men with prostate cancer. Survival curves represent pooled data from a meta-analysis of multiple studies. (From Maximum androgen blockade in advanced prostate cancer: an overview of the randomised trials. Prostate Cancer Trialists’ Collaborative Group. Lancet 2000; 355:1491-1498. With permission of the publisher.)
Secondary Hormonal Therapy

Men who experience cancer recurrence after medical or surgical orchiectomy are treated with secondary hormonal therapies. Before the use of PSA measurements, the clinician had to rely on bone radiographs and soft tissue changes to document responses and these were rarely observed. Consequently, the efficacy of secondary hormonal therapies was controversial. Clinicians have now accepted a 50% decline in PSA as reflecting an objective response to therapy. Studies have shown that this PSA end point predicts a significantly prolonged median overall survival, objective progression-free survival, and time to pain progression. Using PSA measurements, it is now possible to demonstrate the efficacy of a number of secondary hormonal therapies. Responses to antiandrogens, ketoconazole, flutamide, bicalutamide, aminoglutethimide, DES, and glucocorticoids alone range from 14% to 60% (Table 39-2). Head-to-head comparisons are unavailable for these agents, and relative efficacy cannot be compared.
Experimental studies with LnCAP cells in vitro detected a mutation in the androgen receptor in this cell line that caused it to respond to flutamide with increased proliferation. As a result of this observation, clinical studies examined whether tumors in some patients might adapt to flutamide by developing mutations, allowing flutamide to behave as an androgen agonist. In support of this possibility, approximately 40% of men receiving flutamide experienced tumor regression after withdrawal of flutamide (so-called withdrawal responses). These data suggest that a first step in patients receiving flutamide, either as part of a complete androgen blockade regimen or as secondary hormonal therapy, is to stop flutamide. These responses occur with bicalutamide as well but with lower frequency.

### TABLE 39-2 - Prostate Cancer: Secondary Hormonal Therapies

<table>
<thead>
<tr>
<th>Modality</th>
<th>Responders</th>
<th>Total Patients</th>
<th>Responses (%)</th>
<th>Clinical Setting</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flutamide</td>
<td>23</td>
<td>100</td>
<td>23</td>
<td>First relapse after medical or surgical orchiectomy</td>
<td>[446]</td>
</tr>
<tr>
<td>Prednisone</td>
<td>21</td>
<td>101</td>
<td>21</td>
<td>First relapse after medical or surgical orchiectomy</td>
<td>[446]</td>
</tr>
<tr>
<td>Flutamide withdrawal</td>
<td>29</td>
<td>138</td>
<td>21</td>
<td>Relapse after combined androgen blockade or flutamide monotherapy</td>
<td>[385] [474]</td>
</tr>
<tr>
<td>AG/HC</td>
<td>14</td>
<td>29</td>
<td>49</td>
<td>After antiandrogen withdrawal</td>
<td>[475]</td>
</tr>
<tr>
<td>HC</td>
<td>20</td>
<td>48</td>
<td>40</td>
<td>After antiandrogen withdrawal</td>
<td>[476]</td>
</tr>
<tr>
<td>Ketoconazole/HC</td>
<td>43</td>
<td>72</td>
<td>60</td>
<td>After antiandrogen withdrawal</td>
<td>[478]</td>
</tr>
<tr>
<td>Megestrol acetate</td>
<td>17</td>
<td>119</td>
<td>14</td>
<td>After antiandrogen withdrawal</td>
<td>[479]</td>
</tr>
<tr>
<td>Diethylstilbestrol</td>
<td>71</td>
<td>243</td>
<td>29</td>
<td>After antiandrogen withdrawal</td>
<td>[480]</td>
</tr>
</tbody>
</table>

AG, aminoglutethimide; HC, hydrocortisone.
Other Secondary Hormonal Therapies

Prior to the PSA era, objective regression, stabilization, or symptomatic relief was reported with several agents. These included a high-dose formulation of DES called Stilphostrol and tamoxifen. A subsequent study using high-dose tamoxifen (160 mg/m² per day) observed only a 3.3% rate of objective response on the basis of PSA in heavily pretreated men with prostate cancer. Each of these agents could be considered for patients who have slow relapses after castration.
Experimental Hormonal Approaches

Experimental data for Shionogi tumors in mice suggest that intermittent androgen withdrawal might control tumor growth and delay development of hormonal resistance. Several pilot reports suggest the feasibility of this approach, and observations are ongoing. This approach cannot be recommended until sufficient data are available to support its efficacy. Another approach undergoing extensive testing of efficacy is PS SPES, an herbal supplement.
Development of Androgen-Independent Tumor Growth

At some point in the patient's course, the tumor becomes refractory to hormonally based therapies. Current theory holds that tumors escape androgenic regulation of growth and co-opt mechanisms utilizing ligand-independent activation of androgen receptors, up-regulation of growth factor pathways, or a combination of these mechanisms. Considerable experimental support for these concepts derives from the demonstration of increased activated MAP kinase activity (a protein that mediates mitogenesis) and enhanced phosphorylation of the androgen receptor, a marker of ligand-independent receptor activation.

A major decision at the time of prostate cancer relapse after medical or surgical castration is whether to use hormonal therapy or chemotherapy. No diagnostic test is available to make the distinction between hormone dependence and independence. Men with a rapid downhill course with widespread systemic metastases should probably receive chemotherapy. Others have already been treated with all available options of endocrine therapy. Chemotherapy or combination chemohormonal therapy is chosen at that point. The choice of agents is beyond the scope of this chapter, and the interested reader is referred to the comprehensive treatise of Carroll and colleagues.
Algorithm for Treatment of Prostate Cancer

To date, no standard approach to the treatment of prostate cancer has been universally accepted. To outline the options available, the algorithm in Figure 39-22 details decision branch points and currently accepted approaches with asterisks to indicate therapies that are controversial.

Initial Treatment of Nonmetastatic Prostate Cancer

No consensus exists regarding the selection of radical prostatectomy, radiation therapy, or watchful waiting. Patients with potentially curable, nonmetastatic (T1 to T3, N0, M0; stages A to C) prostate cancer are commonly treated by radical prostatectomy, radiation therapy, or watchful waiting. Individual considerations, including age, risk factors for recurrence, and overall health of the patient, influence these decisions, which are often based on personal choices. In general, the younger and healthier the patient, the more likely is a choice of radical prostatectomy. The older and more debilitated the patient, the more likely is the choice of watchful waiting. Radiotherapy is often selected for patients falling at neither extreme.

Patients with T1 to T3, N0, M0 (stages A to C) prostate cancer are commonly treated with radical prostatectomy. Neoadjuvant or adjuvant hormonal therapy is not usually recommended for these patients. Upon first recurrence, these men are then treated with either medical or surgical orchietomy.

In men undergoing radiation as definitive treatment, adjuvant hormonal therapy appears to be warranted. Neoadjuvant therapy has not been adequately tested in a trial in which this step is added to long-term adjuvant therapy. Nonetheless, neoadjuvant therapy would appear reasonable on the basis of the minimal additional toxicity and side effects associated with this therapy for an additional 3 to 6 months before radiation therapy.

Watchful waiting is selected for men whose other illnesses would be expected to result in demise before death related to prostatic cancer.

Initial Therapy for Men with Metastatic Disease at Time of Diagnosis

Men with metastatic disease (T1 to T4, N+, or M+) at presentation are not treated definitively with radical prostatectomy or irradiation but are frequently offered hormonal therapy before the development of symptomatic disease. This decision is based on incomplete data suggesting superior efficacy of early rather than late hormonal therapy for locally advanced disease. Randomized trials with a large number of patients are required before definitive advice can be given regarding management of PSA relapses.

Which hormonal therapy to use is a matter of choice. A meta-analysis of 10 trials suggested that medical castration, surgical castration, and DES all have equal efficacy, but individual trials indicated that antiandrogens are less effective. In this author's opinion, surgical orchietomy appears preferable for the initial treatment of these patients. The operation is relatively minor and can be performed under local anesthesia, if desired. Rapid and complete cessation of testicular androgen secretion ensues, and the patient's compliance after surgery is not a factor. Unwarranted cardiovascular or other toxic side effects, such as those occurring with DES, are unknown. The incidence of impotence and loss of libido is the same as with other available therapies with the exception of the antiandrogens.

In practice, nearly half of patients in the United States prefer a form of medical castration for psychological or other reasons. The greater safety of the GnRH agonists is an increasingly important consideration in prostate cancer therapy. Some patients may choose to continue the antiandrogens in the long term to gain the slight (2.9%) maximal benefit from MAB. An antiandrogen or DES, 3 mg/day, can be chosen as an alternative but is not preferred.

At present, medical castration with GnRH agonists can be achieved with a 3- or 4-monthly injection that is well tolerated and should not be associated with cardiovascular complications. Cost ($2700 for the 4-month injection) is the limiting factor. When concern about erectile dysfunction is a limiting factor, the persistence of libido and erectile function in a subset of patients treated with antiandrogens as primary therapy may be a decisive factor. Under these circumstances, some men may choose monotherapy with antiandrogens, even though this approach is not as effective as medical or surgical castration. This decision is related to quality of life, which is an increasingly important consideration in prostate cancer therapy.

Disease Recurrence

Antiandrogen therapy provides a reasonable next option unless it has previously been given as part of an MAB regimen. Patients receiving an antiandrogen as part of a complete androgen blockade regimen are observed for an antiandrogen withdrawal response. Regarding the choice of antiandrogen, bicalutamide is preferred to flutamide because of its better side effect profile and potentially superior efficacy. Upon relapse, men are observed for an antiandrogen withdrawal response and treated further only when progressive disease is documented.

A variety of therapies are available for men experiencing relapse after antiandrogen withdrawal. One group preferred ketoconazole and hydrocortisone because of their lack of toxicity and observed efficacy (i.e., 60% PSA response rate). The other agents listed in Table 39-2 can be used as alternatives or as third-line therapies. Further choices of chemotherapy alone or chemotherapy in combination with hormonal agents are a matter of individual preference.
References


82. Son HJ, Oh KK. Significance of follow-up mammography in estimating the effect of tamoxifen in breast cancer patients who have undergone surgery. AJR 1999; 173:905909.


References


null


References


1829


1831


Chapter 40 - Humoral Manifestations of Malignancy

Gordon J. Strewler

The syndrome of ectopic secretion of corticotropin (also known as ACTH) was first described by Brown in 1928 in a woman with bronchogenic carcinoma, 4 years before Cushing's description of the clinical syndrome of corticotropin excess and before the relationship between the hormone and the clinical syndrome was recognized. In 1941, Albright proposed the idea that tumors can cause endocrine syndromes by secreting hormones inappropriately; he suggested that hypercalcemia in a patient with renal carcinoma might be due to production of parathyroid hormone (PTH) by the tumor. Albright was led to this conclusion by the coexistence of hypercalcemia with hypophosphatemia (biochemical features of primary hyperparathyroidism) in a patient with a single bone metastasis.

In 1956, cases were reported in which hypercalcemia was cured by resection of the primary tumor, supporting Albright's hypothesis that the neoplasm produced humoral hypercalcemia. Subsequently, Schwartz and colleagues described the syndrome of inappropriate vasopressin (ADH) secretion in bronchogenic carcinoma. Meador and co-workers described the ectopic corticotropin syndrome in lung carcinoma, and Liddle and colleagues coined the term ectopic hormone syndrome to describe such situations.
INAPPROPRIATE HORMONE SECRETION IN MALIGNANCY

Common Features

Inappropriate secretion of peptide hormones is probably the most common cause of paraneoplastic syndromes. Although the manifestations vary widely, paraneoplastic hormonal syndromes have features that distinguish them from overproduction of hormones by endocrine glands (Table 40-1).

First, the secretion of hormones by extraglandular tumors is rarely suppressible. Tumor cells that secrete hormones typically do not possess the cellular machinery that allows regulation of hormone secretion. The most notable exception to this general rule is the secretion of corticotropin by carcinoid tumors of the lung or thymus. Corticotropin secretion by these tumors is often suppressible by glucocorticoids, with secretory dynamics that can be difficult to distinguish from those of corticotrope adenomas of the pituitary.

Second, because extraglandular tumors produce hormones relatively inefficiently, clinical syndromes of hormone excess become evident only in patients with advanced malignancies. It is probably for this reason that hormones are disappointing as tumor markers.

Third, tumors often lack the ability to process peptide hormones normally and secrete large, incompletely processed forms of hormones with reduced biologic activity.

Finally, some malignant tumors mimic syndromes of hormone excess not by secreting the expected hormone but by secreting related hormones that mimic the biologic actions of the expected hormone. For example, nonislet cell tumors can cause hypoglycemia not by secreting insulin but by secreting the related peptide insulin-like growth factor II (IGF-II). IGF-II does not ordinarily play a major role in glucose metabolism, but it has insulin-like activity and in large amounts causes hypoglycemia. Similarly, hypercalcemia in malignancy is a manifestation not of PTH excess but of an excess of parathyroid hormonereleated protein (PTHrP). PTHrP resembles PTH and in most circumstances is a local regulator but acts as a hormone when it is released into the circulation by malignant tumors.

| TABLE 40-1 -- General Characteristics of Paraneoplastic Hormonal Syndromes |
|-----------------|---------------------------------------------------------------|
| 1.              | Secretion of hormones is rarely suppressible.                |
| 2.              | Clinical syndromes are usually associated with advanced malignancies. |
| 3.              | Hormones are not useful as tumor markers for nonendocrine tumors. |
| 4.              | Tumors may mimic syndromes of hormone excess by secreting a related peptide (e.g., insulin-like growth factor II, causing hypoglycemia). |

become evident only in patients with advanced malignancies. It is probably for this reason that hormones are disappointing as tumor markers.

Finally, some malignant tumors mimic syndromes of hormone excess not by secreting the expected hormone but by secreting related hormones that mimic the biologic actions of the expected hormone. For example, nonislet cell tumors can cause hypoglycemia not by secreting insulin but by secreting the related peptide insulin-like growth factor II (IGF-II). IGF-II does not ordinarily play a major role in glucose metabolism, but it has insulin-like activity and in large amounts causes hypoglycemia. Similarly, hypercalcemia in malignancy is a manifestation not of PTH excess but of an excess of parathyroid hormonereleated protein (PTHrP). PTHrP resembles PTH and in most circumstances is a local regulator but acts as a hormone when it is released into the circulation by malignant tumors.
Ectopic versus Eutopic Secretion

The term ectopic hormone secretion is a misnomer. Ectopic means "out of place," implying secretion of a hormone by tissues that do not ordinarily do so, whereas hormones that are secreted by tumors are usually present in the nonmalignant precursor cells, albeit often in small amounts. For example, many of the hormones typically secreted by small cell lung carcinomas—vasopressin, calcitonin, and gastrin-releasing peptide (GRP)—are thought to be present in the neuroendocrine cells in the normal bronchial mucosa that are the probable precursors of the tumor. PTHrP is a normal product of the keratinocyte, the cell of origin of squamous carcinomas that cause humoral hypercalcemia by secreting PTHrP. Human chorionic gonadotropin (hCG), usually considered a placental hormone, is not under tight transcriptional control, and low levels of the hormone are detectable in a variety of other normal tissues, a finding that is in keeping with the occurrence of hCG secretion by many tumor types.

Thus, most endocrine manifestations of malignancy are caused by eutopic secretion of hormones by cells that were previously programmed to secrete them. This feature has implications for the pathogenesis of the humoral manifestations of malignancy that are discussed later. As for nosology, the term ectopic is firmly ingrained and will not soon be abandoned, even though true ectopic secretion of hormones is rare.
Peptide versus Nonpeptide Hormones

Most peptide hormones are secreted by nonendocrine malignant tumors (Table 40-2), with some exceptions. The glycoprotein hormones follicle-stimulating hormone (FSH), luteinizing hormone (LH), and thyrotropin (TSH) are rarely, if ever, produced by extrapituitary tumors, possibly because it is necessary to express the genes for two subunits, to glycosylate the subunits appropriately, and to assemble the complete dimer to produce a biologically active hormone. However, the glycoprotein hormone hCG is frequently secreted by nontrophoblastic tumors, illustrating that the requisite machinery can be present in nonpituitary cells.

Unlike the case of pituitary cells, synthesis of hCG by tumors is not tightly regulated. Pituitary glycoprotein hormone synthesis is tightly controlled by a series of pituitary-specific transcription factors, whereas hCG is normally expressed at low levels in a variety of nontrophoblastic cells. In the same vein, only one case of extrapancreatic secretion of insulin has been authenticated. Although a few copies per cell of insulin messenger ribonucleic acid (mRNA) can be detected in many cells by sensitive polymerase chain reaction (PCR) methods, physiologic expression of the insulin gene is driven by transcription factors that are specific to the pancreatic beta cell, and these factors are rarely, if ever, expressed in other cell types. Thus, the propensity of a peptide hormone for secretion by extraglandular tumors may be a function of the tightness of its transcriptional suppression in normal extraglandular tissues.

Steroid and thyroid hormones are not secreted by extraglandular tumors, although they are occasionally produced by teratomas that contain glandular elements. Their synthesis requires an extended series of enzymatic steps that are not present in nonsteroidogenic tissues. In contrast, 1,25-dihydroxycholecalciferol (1,25(OH)\(_2\)D, also called 1,25-dihydroxyvitamin D\(_3\)) is secreted by lymphomas as well as macrophages resident in granulomas of sarcoidosis and other granulomatous disorders. In these tissues, the synthesis of 1,25(OH)\(_2\)D requires a single enzymatic step, the 1-hydroxylation of the circulating precursor, 25-hydroxycholecalciferol.

### TABLE 40-2 – Hormones Produced by Tumors

<table>
<thead>
<tr>
<th>Hypercalcemia factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parathyroid hormone (PTH) related protein (PTHrP)</td>
</tr>
<tr>
<td>Tumor necrosis factor</td>
</tr>
<tr>
<td>1,25-Dihydroxycholecalciferol (1,25(OH)(_2)D)</td>
</tr>
<tr>
<td>Prostaglandins</td>
</tr>
<tr>
<td>PTH</td>
</tr>
<tr>
<td>Vasopressin</td>
</tr>
<tr>
<td>Corticotropin (ACTH)</td>
</tr>
<tr>
<td>Growth hormone-releasing hormone</td>
</tr>
<tr>
<td>Insulin-like growth factor II (IGF-II)</td>
</tr>
<tr>
<td>Calcitonin</td>
</tr>
<tr>
<td>Human chorionic gonadotropin (HCG)</td>
</tr>
<tr>
<td>Human placental lactogen (hPL)</td>
</tr>
<tr>
<td>Growth hormone</td>
</tr>
<tr>
<td>Corticotropin-releasing hormone</td>
</tr>
<tr>
<td>Erythropoietin</td>
</tr>
<tr>
<td>Oncogenous osteomalacia factor</td>
</tr>
<tr>
<td>Atrial natriuretic peptide</td>
</tr>
<tr>
<td>Endothelin</td>
</tr>
<tr>
<td>Renin</td>
</tr>
<tr>
<td>Other gut hormones</td>
</tr>
<tr>
<td>Gastrin-releasing peptide</td>
</tr>
<tr>
<td>Glucose-dependent insulinoemorphic peptide (gastrin inhibitory peptide)</td>
</tr>
<tr>
<td>Somatostatin</td>
</tr>
<tr>
<td>Pancreatic polypeptide</td>
</tr>
<tr>
<td>Vasoactive intestinal peptide (VIP)</td>
</tr>
<tr>
<td>Substance P</td>
</tr>
<tr>
<td>Motilin</td>
</tr>
</tbody>
</table>
Cellular Basis of Ectopic Hormone Secretion

Why is the secretion of hormones by malignant neoplasms so commonplace? The simplest idea is that random sets of genes are “derepressed” in the cancer cell, including genes that code for hormones. For example, epigenetic phenomena such as demethylation of deoxyribonucleic acid (DNA) may derepress hormone genes. However, the association of tumors and hormone secretion is nonrandom, with certain tumors (e.g., lung carcinoma) characteristically secreting certain hormones (e.g., corticotropin or vasopresin). Moreover, in a number of cases, the peptides that are secreted by tumors are the same peptides that are secreted by the normal cell of origin of the neoplasm. Derepression is again nonrandom and is a quantitative rather than a qualitative phenomenon.

The dedifferentiation hypothesis posits a retrograde movement of tumor cells along the pathway of differentiation, leading to the expression of fetal proteins (e.g., -fetoprotein and carcinoembryonic antigen) or hormones that are normally formed in immature cells. This hypothesis would account for both the nonrandom nature of ectopic hormone secretion and the propensity for secretion of hormones that play a critical role in development, for example, IGF-II, PTHrP, and possibly GRP and other peptides of neuroendocrine cells. In addition, tumors frequently secrete other fetal proteins (carcinoembryonic antigen, -fetoprotein). However, there is no compelling evidence for a generalized pattern of expression of primitive genes in tumor cells.

The dysdifferentiation hypothesis of Baylin and Mendelsohn holds that epithelial malignancy is the result of clonal expansion of a particular cell type that occurs along a complex pathway of epithelial differentiation. This process may give rise to overexpression of a hormone because of (1) clonal expansion of a normally rare population of committed cells or (2) clonal expansion of a primitive cell type not normally present in the mature epithelium.

Because the defining characteristic of neoplastic cells is uncontrolled growth, it is worth considering the possible relationship of disordered growth and the secretion of hormones. In some instances, an oncogenic event might directly activate transcription of a hormone gene. One example of direct gene activation is the secretion of PTHrP in adult T-cell leukemia, with associated severe hypercalcemia. The oncogenic event that gives rise to this form of leukemia involves integration of the human T-cell lymphotropic virus type I (HTLV-I), which can target the promoter of the PTHrP gene to induce its transcription using the trans-activating viral protein tax. Another clinical example is the association of the von HippelLindau suppressor gene VHL with the development of cerebellar hemangioblastoma and renal carcinoma. Evidence suggests that erythropoietin gene expression is directly up-regulated by this oncogenic event.

Secretion of a hormone might stimulate the growth of tumor cells by an autocrine or a paracrine mechanism, so that hormone secretion may provide a growth advantage, leading to selective outgrowth of cells that secreted high levels of the hormone. One of the characteristic products of small cell lung carcinoma (SCLC) is GRP, the mammalian counterpart of the amphibian hormone bombesin. GRP fulfills criteria for being an autocrine growth factor in SCLC: (1) it is secreted by tumor cells and can stimulate replication of the cells via specific receptors, and (2) blockade of its action by neutralizing antibodies to GRP or peptide antagonists inhibits cell replication in vitro and tumor formation in vivo. -Endorphin, one of the products of the pro-opiomelanocortin (POMC) gene, may also function as a growth factor in SCLC.

Endothelin receptors are often coexpressed on tumor cells with endothelin-1, and endothelin-1 is reported to have paracrine effects on tumor cell growth. Prolactin and its receptor are expressed by breast cancer cells, although rarely, if ever, at high enough levels to raise serum prolactin levels, and an autocrine pathway has been described in which prolactin induces constitutive phosphorylation of erbB-2 (Her/Neu), an oncogene that is important in growth of breast cancer. Another ectopic hormone with growth factor activity is IGF-II, the factor that is believed to cause hypoglycemia in nonislet cell tumors. However, there is no direct evidence for a role of IGF-II in the growth of these neoplasms.

Epigenetic events associated with tumorigenesis may activate the transcription of hormone genes. Altered methylation of CpG islands appears to play a role in the expression of PTHrP in renal carcinoma, with undermethylation of the PTHrP promoter in tumors that express the gene. Evidence suggests that demethylation of the POMC promoter may also be involved in expression of corticotropin in neuroendocrine tumor cells.

Thus, ectopic hormone production can be partly understood in the context of the determinants of tumor cell behavior generally. Nevertheless, why specific tumor types overexpress hormone genes and why overexpression occurs in some tumors of a given type (e.g., SCLC) and not others are questions that remain largely unanswered.
Neuroendocrine Cells and Hormone Secretion

Tumors that secrete corticotropin, vasopressin, calcitonin, gut peptides (GRP, somatostatin, vasoactive intestinal peptide [VIP]), and biogenic amines such as 5-hydroxytryptamine (5-HT) are characteristically of neuroendocrine cell origin. Neuroendocrine cells specialized for the production of peptide hormones and biogenic amines possess pathways for the rapid release of peptides or neurotransmitters in response to stimuli; such regulated pathways for protein secretion are distinct from the mechanism of constitutive secretion, which is ubiquitous in eukaryotic cells.

The most readily recognizable feature of the regulated pathway is the dense neurosecretory granule, which is designed for the secretion of peptides and amines. The neurosecretory granule, which is involved in both the storage of hormones in concentrated form and the rapid release of these stores in response to stimulation, is recognizable histologically because it is electron dense and intensely argyrophilic, reflecting the dense, nearly crystalline packing of its contents.

The neurosecretory granules bud from the trans-Golgi network after it is packed with its peptide or neurotransmitter contents. Proteins on the surface of the neurosecretory granule, in the vesicular compartments from which the granule buds, and on the plasma membrane of neurosecretory cells collectively determine the properties of the regulated pathway of hormone secretion. They are probably important in ectopic hormone secretion. In addition to stored hormones, the neuroendocrine granule contains one or more acidic proteins called chromogranins, which are released together with stored hormone and serve as additional neuroendocrine tumor markers, both in immunohistochemistry and in the circulation. The chromogranins are highly conserved in evolution and presumably play a role in the assembly, packing, or release of neurosecretory granules, but this role has yet to be fully clarified.

The neurosecretory granules contain serine proteases called prohormone convertases, which process precursor proteins to their mature forms. The prohormone convertase family is widely distributed in evolution, and several members of the family, such as furin, are localized in the trans-Golgi network and process a wide variety of proteins in the constitutive pathway. The two members of the family that occur mainly in neurosecretory granules, PC2 (SPC2) and PC1/PC3 (SPC3), have acidic pH optima and are dependent on calcium, suiting them for the environment of the neurosecretory granule. Both enzymes cleave their substrate peptides on the carboxyterminal side of polybasic residues, but they have slightly different specificities, and their different distribution in the pituitary accounts for the differences in processing of POMC in the anterior and intermediate pituitary lobes.

The prohormone convertases are important for our understanding of ectopic hormone secretion. Their levels in tumor cells account for the efficiency of precursor processing to mature and biologically active versions of peptide hormones, thus determining whether a given tumor produces a clinical syndrome of hormone excess. Further, they determine the pattern of peptides produced (e.g., from the polyhormone precursor POMC) and thus the nature of the clinical syndrome.

Neuroendocrine cells are scattered through the bronchial mucosa of the developing and mature lung. They occur singly and in distinct inervated corporcles referred to as neuroepithelial bodies. Subpopulations of the cells contain the peptides calcitonin, GRP, vasopressin, and leu-enkephalin, and some may also contain somatostatin, motilin, or pancreatic polypeptide. Neuroendocrine (enterochromaffin) cells are also scattered through the gastrointestinal mucosa and are found in other organs, such as the ovaries and the prostate gland. Some years ago, Pearce suggested that neuroendocrine cells, which he called APUD cells (amine precursor uptake and decarboxylation), although widely scattered in many tissues, have a common origin in the neural crest and represent a diffuse neuroendocrine system, a third branch of the nervous system. Not all APUD cells are of neural crest origin, however; some arise from primitive endoderm.

Neuroendocrine cells in the gut and lung are specified by a set of basic helix-loop-helix (bHLH) transcription factors that are involved in the determination of neuronal fate in mammals and Drosophila. Transient expression of the mouse achaete-scute homologue-1 (mASH1) is required for neurogenesis of autonomic and enteric neurons and adrenal chromaffin cells. Pulmonary neuroendocrine cells do not develop in mice deficient in mASH1 and forced expression of mASH1 induces metaplasia and cooperates with simian virus 40 (SV40) T antigen in tumorigenesis. Neuroendocrine tumor cells such as SCLC cells express the human orthologue hASH1.

Disruption of an inhibitory pathway may also lead to neuroendocrine cell expansion. Drosophila and vertebrate neurogenesis is characterized by lateral inhibition, a cell-cell interaction in which differentiating neuronal cells inhibit neuronal differentiation of their neighbors through actions of a transmembrane receptor, Notch, and the ligand Delta. One of the chief targets of the inhibitory pathway is hairy-enhancer-of-split-1 (Hes-1), which inhibits the proneural genes neurogenin, neuroD, and ASH. Ablation of Hes-1 leads to a marked increase in enteroendocrine cells, and SCLC cells are characteristically deficient in Hes-1. Collectively, these data delineate a pathway of origin of neuroendocrine tumor cells and identify transcription factors that may be directly involved in up-regulating expression in neuroendocrine cells. However, examples of direct regulation of hormonal expression by such genes have not yet been adduced.
Criteria for Diagnosis of Ectopic Hormone Secretion

Criteria for the diagnosis of ectopic hormone secretion, arranged in increasing order of stringency, are summarized in Table 40-3. The association of a clinical syndrome of hormone excess with a neoplasm provokes a search for inappropriate plasma or urinary hormone levels. In many cases, the clinician performs suppression tests because glandular hypersecretion of hormones is often suppressible whereas the secretion of hormones by neoplasms is typically autonomous and non-suppressible. In the usual clinical circumstance, the last step is to exclude other possible causal mechanisms for hormone excess.

The coincidental occurrence of an endocrine tumor and a cancer is not uncommon; for example, primary hyperparathyroidism can be present in a patient who also has cancer and can be detected with relative ease using modern assays for PTH. Occasionally, the presence of an ectopic hormone syndrome can be confirmed by showing that resection of the tumor reverses the clinical syndrome. Because such syndromes are typically late manifestations of widespread neoplasms, these opportunities are sadly rare.

The remaining criteria for ectopic hormone secretion are useful mainly for research purposes. The detection of a hormone in tumor tissue by immunoassay methods provides evidence that the tumor is a site of its production, although caution must be exercised because of the possibility of false-positive reactions in immunohistochemistry and radioimmunoassay. An additional theoretical concern is that the tumor may accumulate hormone from the circulation; no examples of this phenomenon have been reported, however. Detection of mRNA for the hormone confirms that the tumor is indeed a site of synthesis of the peptide. For this evidence to be compelling, hormone mRNA should be detectable in solution hybridization or RNA blotting assays. The technique of reverse transcription and PCR is so sensitive that a signal can be obtained from samples containing only a few molecules of hormone mRNA, a level that may be insignificant. In addition, identification of hormone mRNA without hormone protein leaves open the possibility that the mRNA is not translated for example, many normal tissues express a form of POMC mRNA that cannot be translated to protein.11

Demonstration of the presence of both hormone mRNA and protein provides strong evidence for synthesis in the tumor but does not directly establish that the hormone is secreted. The most rigorous criterion for ectopic hormone secretion is the demonstration of an arteriovenous gradient of the hormone across the tumor or of production and secretion of the hormone by tumor cells cultured in vitro. Unfortunately, selective catheterization to obtain a true arteriovenous gradient is often impossible, as many of the tumors are present in the pulmonary or splanchnic bed or are widely metastatic. Establishing tumor cells in culture provides an important research tool but requires an element of good fortune—many tumor cells, exuberant as their growth may be in the host, are difficult to propagate in cell culture.

---

**TABLE 40-3 -- Criteria for Diagnosis of Ectopic Hormone Secretion**

<table>
<thead>
<tr>
<th>Clinical criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. A clinical syndrome of hormone excess is associated with a neoplasm.</td>
</tr>
<tr>
<td>2. Serum or urine levels of the hormone are inappropriately elevated.</td>
</tr>
<tr>
<td>3. The hormone level is nonsuppressible.</td>
</tr>
<tr>
<td>4. Other possible causal mechanisms are excluded.</td>
</tr>
<tr>
<td>5. The syndrome is reversed by resection of the tumor (rare).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Research criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The hormone can be detected in tumor tissue.</td>
</tr>
<tr>
<td>2. Messenger RNA for the hormone is present in tumor tissue.</td>
</tr>
<tr>
<td>3. The hormone is secreted from tumor cells in culture.</td>
</tr>
<tr>
<td>4. There is an arteriovenous gradient for the hormone across the tumor.</td>
</tr>
</tbody>
</table>
MALIGNANCY-ASSOCIATED HYPERCALCEMIA

Clinical Features

Hypercalcemia is probably the most common endocrine complication of malignant tumors, occurring in as many as 5% of all cancers. The incidence of hypercalcemia in malignancy is 15 cases per 100,000 person-years, about one half the incidence of primary hyperparathyroidism, and malignant tumors are the most common cause of hypercalcemia in hospitalized patients (see Chapter 26).

Hypercalcemia in malignancy usually has a rapid onset and can cause confusion, stupor, nausea, vomiting, and dehydration. The offending neoplasm is almost always evident clinically, even when hypercalcemia is the initial manifestation. Thus, physical examination and a chest radiograph disclose the underlying tumor in about 98% of patients. Because hypercalcemia usually occurs in advanced malignancy, the prognosis is poor, with a median survival of only 4 to 8 weeks after the discovery of hypercalcemia. Exceptions are breast carcinoma and multiple myeloma, in which successful treatment of the underlying malignancy may allow long survival of the hypercalcemic patient.

The frequency of individual tumors in patients with hypercalcemia is shown in Table 40-4. Lung carcinoma, breast carcinoma, and multiple myeloma account for more than 50% of all cases of malignancy-associated hypercalcemia. Lung carcinomas that produce hypercalcemia have squamous or large cell histology, whereas small cell carcinoma almost never causes hypercalcemia. About two thirds of lung cancer patients have bone metastasis at the time when hypercalcemia develops. Among other solid tumors, the most common are squamous and renal carcinomas. Gastrointestinal tumors and prostate carcinoma are less common causes of hypercalcemia.

Hypercalcemia is uncommon in lymphomas and leukemia but occurs in two thirds of patients with adult T-cell leukemia syndrome, which is caused by the retrovirus HTLV-I. Another rare variety of leukemia in which hypercalcemia is common is the M7 variant of acute myelogenous leukemia.

Hypercalcemia is a common complication of multiple myeloma. Hypercalcemia in myeloma has been ascribed to a local osteolytic cause, but a substantial fraction of cases have increased PTHrP levels. Pheochromocytomas may produce hypercalcemia by secretion of PTHrP.

<table>
<thead>
<tr>
<th>Primary Site</th>
<th>No. (%) of Cases</th>
<th>Known Metastatic Disease (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lung</td>
<td>111 (25.0)</td>
<td>62</td>
</tr>
<tr>
<td>Breast</td>
<td>87 (19.6)</td>
<td>92</td>
</tr>
<tr>
<td>Multiple myeloma</td>
<td>43 (9.7)</td>
<td>100</td>
</tr>
<tr>
<td>Head and neck</td>
<td>36 (8.1)</td>
<td>73</td>
</tr>
<tr>
<td>Renal and urinary tract</td>
<td>35 (7.9)</td>
<td>36</td>
</tr>
<tr>
<td>Esophagus</td>
<td>25 (5.6)</td>
<td>53</td>
</tr>
<tr>
<td>Female genitalia</td>
<td>24 (5.2)</td>
<td>81</td>
</tr>
<tr>
<td>Unknown primary</td>
<td>23 (5.2)</td>
<td></td>
</tr>
<tr>
<td>Lymphoma</td>
<td>14 (3.2)</td>
<td>91</td>
</tr>
<tr>
<td>Colon</td>
<td>8 (1.8)</td>
<td></td>
</tr>
<tr>
<td>Liver, biliary</td>
<td>7 (1.6)</td>
<td></td>
</tr>
<tr>
<td>Skin</td>
<td>6 (1.4)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>25 (5.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>444 (100)</td>
<td></td>
</tr>
</tbody>
</table>

Data from references 63, 64, 65. Data on metastatic disease from references 64 and 65.

Laboratory Features

Overall, about 80% of cases, including most patients with solid tumors, have increased serum levels of PTHrP, which can be measured in two-site, amino-terminal or midregion assays (Fig. 40-1). Hypophosphatemia is common because of the phosphaturic effect of PTHrP. Although the combination of hypercalcemia and hypophosphatemia is consistent with the presence of primary hyperparathyroidism, the level of intact PTH is suppressed to less than 2 pmol/L (20 pg/mL) in patients with malignancy-associated hypercalcemia. The serum level of 1,25(OH)₂D is also suppressed in hypercalcemic patients, except in lymphoma, in which 1,25(OH)₂D levels are often high. Renal function may be impaired by hypercalcemia; the decreased glomerular filtration rate may lead to normalization of blood phosphate in patients with PTHrP-mediated hypercalcemia.
Pathogenesis

Hypercalcemia in malignancy is caused by excessive bone resorption. Multiple myeloma and some breast cancers induce hypercalcemia by local osteolytic mechanisms, but in most patients bone resorption is induced by humoral factors. The most common humoral factor is PTHrP, but 1,25(OH)2 D is a hypercalcemic factor in lymphomas and in some instances PTH is secreted ectopically by nonparathyroid tumors.

Parathyroid Hormone-Related Protein

PTHrP is related to PTH structurally (see Fig. 24-12) and shares a common receptor with PTH1R (see Chapter 26 for a discussion of the chemistry of PTHrP). Because PTH and PTHrP share a receptor, their biologic actions are similar. PTHrP produces hypercalcemia by increasing resorption of bone throughout the skeleton and by increasing the renal resorption of calcium and causes hypophosphatemia through a phosphaturic effect at the kidney. The hypercalcemic effect of PTHrP probably plays a significant role in the pathogenesis of hypercalcemia, albeit secondary to the role of bone resorption. In some studies, the effects of bisphosphonate treatment of hypercalcemia, the effect was not negatively correlated with the serum level of PTHrP, suggesting that increased renal calcium reabsorption may limit the response of some patients to treatment with inhibitors of bone resorption.

PTHrP functions in normal physiology as a tissue factor that regulates cellular proliferation and differentiation in fetal development and in tissues such as the breast, skin, and hair follicle in the adult (see Chapter 26). Remarkably, PTHrP locally regulates development, in part acting via the same receptor that PTH uses systemically to regulate its target tissues, bone and kidney. However, when PTHrP is produced by a tumor of sufficient mass, it enters the systemic circulation, where it activates PTH-PTHrP receptors in bone and kidney and produces hypercalcemia.

PTHrP can either produce humoral hypercalcemia or cause local osteolytic hypercalcemia by direct activation of osteoclasts in the vicinity of bone metastases. In lung and renal carcinoma, PTHrP often acts as a humoral factor because hypercalcemia can occur without evidence of bone metastasis (see Table 45-1). Even when bone metastases are present, hypercalcemia is predominantly humoral because the serum calcium level correlates better with the level of PTHrP than with the number or size of bone metastases.

There is experimental evidence that PTHrP, when expressed in bone metastasis, can also be a local osteolytic factor. In an experimental model, transfusion of PTHrP complementary DNA into breast carcinoma cells increased their propensity for bone metastasis, and bone metastasis induced local osteolysis without an increase in the circulating level of PTHrP. This result is in agreement with the biology of human breast carcinoma, in which about 50% of hypercalcemic patients have extensive bone metastases without increased serum levels of PTHrP and are presumed to have local osteolytic hypercalcemia (see Table 45-1).

In humans, metastases of breast cancer to bone are immunohistochemically positive for PTHrP in 92% of cases, compared with 17% for nonosseous metastases. Tumor cells that secrete PTHrP have a selective advantage in bone, probably because they induce local resorption. Moreover, transforming growth factor-β (TGF-β) released from bone matrix as a result of bone resorption induces the expression of PTHrP, setting up a positive feedback loop to perpetuate the process.

At least two aspects of the hypercalcemia syndrome associated with PTHrP are paradoxical. First, plasma levels of 1,25(OH)2 D tend to be low in malignancy, despite the acute effect of PTHrP to suppress renal synthesis of 1,25(OH)2 D. This finding contrasts with normal to high levels of 1,25(OH)2 D in primary hyperparathyroidism. This difference may be due to the capacity of hypercalcemia itself to suppress production of 1,25(OH)2 D, tending to counteract the acute stimulatory effect of PTH or PTHrP. Thus, a continuous chronic infusion of PTH or PTHrP, in contrast to the effects of primary hyperparathyroidism, suppresses 1,25(OH)2 D levels, but it does not suppress the serum calcium level. In some patients, normocalcemic with malignancy have a marked increase in 1,25(OH)2 D to supranormal levels when hypercalcemia is treated with bisphosphonates. It is possible that the parallel secretion of PTHrP (continuous?) or associated ancillary factors increases the susceptibility of 1,25(OH)2 D synthesis to suppression by hypercalcemia.

A second paradox concerns bone turnover in the setting of high PTHrP levels. Despite avid bone resorption, bone formation is reduced in postmortem bone biopsy specimens from patients with malignancy and hypercalcemia. This uncoupled state contrast with primary hyperparathyroidism and most other resorptive states, which are characterized by coupled increases in bone formation, and also contrasts with the results of a PTHrP infusion. It is possible that immobilization, anorexia, illness, or other cytokines secreted by neoplasms depress osteoblastic activity.

Mechanisms involved in the activation of PTHrP gene expression in malignant tumors may include trans-activation by tumor-specific factors and differential methylation. The best example of trans-activation is adult T-cell leukemia, which is commonly characterized by PTHrP-dependent hypercalcemia.

A specific trans-activating protein, called tax, in the genome of HTLV-I is capable of direct activation of PTHrP transcription, acting primarily at an Ets-1 site near the most downstream promoter. The keratinocyte is a prominent site of normal PTHrP expression; thus, secretion of PTHrP by squamous carcinomas can be regarded as ectopic. However, only a fraction of patients with squamous carcinomas have hypercalcemia. When PTHrP promoter constructs were fused to a reporter gene and transfected into squamous carcinoma cell lines, the relative level of expression of the reporter gene correlated with the intensity of endogenous PTHrP gene expression in the same cell lines, suggesting that in this circumstance also, the expression of PTHrP is regulated in trans by a factor or factors that are differentially expressed in different squamous carcinomas.

PTHrP expression in human squamous carcinoma cells is repressed by mutant forms of the tumor suppressor gene p53. Most squamous carcinomas express the PTHrP gene at some level, but renal carcinomas are more sharply divided between PTHrP expressors and nonexpressors. The PTHrP gene is undermethylated in renal carcinomas that express PTHrP compared with those that do not. This finding suggests that hypomethylation may be a mechanism by which the gene is expressed in some renal carcinomas.

About 50% of lymphoma patients who become hypercalcemic have inappropriately high serum 1,25(OH)2 D levels. In a few cases, lymph node tissue from...
such patients has been shown to produce 1,25(OH)\(_2\) D in vitro from 25-OHD. \[120\] Challenge of normocalcemic lymphoma patients with the precursor sterol 25-OHD resulted in increased serum 1,25(OH)\(_2\) D levels, increased serum calcium levels, and suppression of PTH. \[120\] This response is in marked contrast to that of normal individuals, who regulate the conversion of substrate to 1,25(OH)\(_2\) D tightly.

The enhanced responsiveness of normocalcemic lymphoma patients to vitamin D indicates that the fundamental abnormality in lymphoma, unregulated extrarenal production of 1,25(OH)\(_2\) D, is more common than hypercalcemia. As would be expected from this interpretation, hypercalcemia is also more common than hypercalcemia in lymphoma patients \[13\] and presumably compensates at least in part for the inappropriate synthesis of 1,25(OH)\(_2\) D. This syndrome resembles the hypercalcemia of sarcoidosis, which is also due to enhanced extrarenal production of 1,25(OH)\(_2\) D. \[89\] As in sarcoidosis, hypercalcemia in lymphoma is frequently responsive to administration of glucocorticoids. Hypercalcemia may not respond well to treatment with hydroxychloroquine. \[129\]

**Parathyroid Hormone**

Secretion of PTH from extraparathyroid tumors is extremely rare, \[71\] \[91\] \[92\] \[93\] although one case fulfilled the most rigorous criterion: demonstration of an arteriovenous gradient for PTH across the tumor. \[124\] Most nonparathyroid tumors that secrete PTH are neuroendocrine tumors, although one was an ovarian adenocarcinoma and one a hepatocellular carcinoma. The diagnosis should be considered in patients with malignant tumors (particularly small cell tumors), hypercalcemia, and elevated PTH levels. However, most patients with these findings have a malignant tumor with coincident primary hyperparathyroidism because this coincidence is more likely than the truly rare syndrome of ectopic PTH secretion. Consequently, exploration of the parathyroid glands may be indicated in patients who require treatment for hypercalcemia.

**Local Osteolytic Hypercalcemia**

Osteolytic lesions cause hypercalcemia by activation of osteoclasts and secretion of bone-resorbing cytokines. Cytokines with osteoclast-activating activity include interleukin-1, tumor necrosis factor (TNF-\(\alpha\)), interleukin-6, TGF-\(\beta\), and PTHrP. \[126\] \[127\] \[128\] \[129\] As discussed earlier, PTHrP appears to be the local osteolytic factor that causes osteolytic hypercalcemia in breast carcinoma. \[126\] \[127\] \[128\] \[129\] The other classic example of local osteolytic hypercalcemia is multiple myeloma. Although at least one third of myeloma patients have hypercalcemia at some time during their disease, the offending cytokine has not been identified with certainty. Cultured human myeloma cell lines produce bone-resorbing factors that can be neutralized with antisera to interleukin-1 \[126\] or TNF-\(\alpha\) (lymphotoxin), \[127\] but it is likely that other factors are responsible for the hypercalcemia.

A preliminary report suggested a strong association of hypercalcemia with secretion of the chemokine macrophage inflammatory protein 1. \[129\] It is clear that a fraction of hypercalcemic patients with multiple myeloma have high serum PTHrP levels and thus humoral rather than local osteolytic hypercalcemia. \[129\] \[130\] \[131\] \[132\] This finding raises the question of whether PTHrP, which may be expressed commonly in marrow myeloma cells, \[130\] is also a dominant local osteolytic factor in multiple myeloma and lymphoma.
Diagnosis

The diagnosis of malignancy-associated hypercalcemia is usually not difficult because the offending neoplasm is clinically evident. It is important to exclude intercurrent primary hyperparathyroidism by showing that the level of intact PTH is suppressed below 2 pmol/L (20 pg/mL). A low serum phosphorus level in conjunction with suppressed PTH levels suggests that the causative factor is PTHrP. Demonstration of elevated levels of PTHrP confirms the diagnosis in patients with solid tumors, but this is often unnecessary clinically. PTHrP is processed to amino-terminal, midregion, and carboxyl-terminal peptides, and similar assay performance has been achieved with assays of the amino terminus and mid-region and two-site immunoradiometric assays (see Fig. 40-1). Two-site assays for PTHrP have become the standard.
Treatment

The treatment of hypercalcemia is discussed in Chapter 26. The mainstays of therapy for tumor patients, in whom hypercalcemia is often acute and severe, are rehydration, institution of a saline diuresis, and institution of chronic treatment. In general, the treatment of choice is the second-generation bisphosphonate pamidronate, 60 to 90 mg by intravenous infusion. Patients with multiple myeloma or lymphoma often respond to glucocorticoid treatment.
SYNDROME OF INAPPROPRIATE VASOPRESSIN SECRETION (See Chapter 9)

Clinical Features

The syndrome of inappropriate vasopressin secretion (commonly termed the syndrome of inappropriate antidiuretic hormone [SIADH]) is probably the second most common endocrine complication in cancer patients. The secretion of vasopressin impairs the ability to dilute the urine, leading to a state of water intoxication with hypotonicity and hyponatremia. Patients with hyponatremia may be asymptomatic if the condition has developed gradually. Patients may experience weight gain because of water retention, but because the retained water is distributed among both extracellular and intracellular spaces, there is no edema. However, when the serum sodium level falls rapidly to below 120 mmol/L, somnolence, coma, and seizures can occur. Symptomatic hyponatremia carries a mortality rate of 10% to 15%, but this rate is higher when the serum sodium level is below 110 mmol/L.

By far, the most common tumor that causes SIADH is SCLC. SIADH occurs in 5% to 15% of patients with SCLC and in fewer than 1% of patients with non-small cell lung cancer (non-SCLC). Other neuroendocrine tumors, including carcinoids and small cell carcinomas of the prostate and cervix, can also cause SIADH. SIADH also occurs occasionally in a wide range of carcinomas, including as many as 2% of squamous carcinomas of the head and neck, adenocarcinoma of the colon, Hodgkin’s disease, non-Hodgkin’s lymphomas, and several varieties of brain tumors. The finding of hyponatremia in cancer is nonspecific, and central secretion of vasopressin from various nonosmotic stimuli may be at fault in some of these patients, for example, patients with thoracic or intracranial tumors. However, vasopressin and associated neurophysins have been found in non-neuroendocrine tumor cells, and some epithelial tumors have the ability to secrete vasopressin.
Laboratory Features

The cardinal features of SIADH are hyponatremia and an inappropriately concentrated urine. It is often unnecessary to measure serum osmolality directly in a hyponatremic patient because the effective serum osmolality (serum sodium (mmol/L) × 2 + glucose (mmol/L)) closely approximates direct measurements. A urine osmolality greater than 50 to 60 mmol/L water is inappropriate in the setting of serum hypotonicity, which should inhibit the release of vasopressin and permit the excretion of a maximally dilute urine. The urine osmolality in SIADH is often higher than the serum osmolality, but that is not a necessary feature of the syndrome. If measured directly, vasopressin levels are inappropriately elevated, as are the levels of the associated neurophysins. However, it is rarely necessary to measure vasopressin. As discussed later, other causes of hypotonicity can generally be excluded with reasonable certainty by reliance on clinical and biochemical criteria.

Several other laboratory features of SIADH are helpful in diagnosis. The urinary sodium concentration is typically high, reflecting the natriuresis induced by expansion of extracellular fluid volume. However, the ability to conserve sodium in SIADH is usually unimpaired, and the urinary sodium excretion can fall to low levels in the setting of reduced dietary sodium intakes. The blood urea nitrogen and serum uric acid levels are low, again reflecting expanded extracellular fluid volumes and decreased tubular resorption of these solutes. Other electrolytes are diluted in proportion to the serum sodium, except for serum bicarbonate, which is normal.
Pathogenesis

Vasopressin is synthesized as a prohormone of 166 amino acids that is processed to produce three peptides: the mature octapeptide hormone, a midregion peptide of molecular weight 10,000 with vasopressin-binding activity called neurophysin II, and a C-terminal glycopeptide. Vasopressin and its neurophysin are packaged together in neurosecretory granules, stored in nerve terminals in the posterior pituitary, and released in response to hypotonicity or nonosmotic stimuli (baroreceptor stimulation, pain, nausea). Vasopressin is similarly processed in neuroendocrine tumor cells, but these cells frequently secrete not only vasopressin and neurophysin II but also vasopressin’s sister peptide oxytocin together with its binding protein, neurophysin I.

The molecular basis for inappropriate secretion of vasopressin from tumor cells is poorly understood. Immunoreactive vasopressin is identifiable in a portion of bronchial neuroendocrine cells, the presumed precursors of SCLC. Thus, like other hormonal products of neuroendocrine neoplasms, vasopressin may be regarded as secreted ectopically from neuroendocrine tumors. The usual coordinate expression of vasopressin and oxytocin precursors in tumor cells may be related to the physical linkage of the two genes, which are found within 12 kb of each other in the human genome with an inverted arrangement of the coding strands. Either DNA rearrangements or trans-acting factors expressed in malignant tumors could simultaneously activate both promoters.

An E-box in the vasopressin promoter, a binding motif for bHLH transcription factors, is necessary for increased vasopressin gene expression in high-expressing SCLC cell lines. This motif appears to bind the transcription factor USF (upstream transcription factor); a nearby sequence appears to interact and is a candidate for binding other bHLH factors. A neuron-specific silencer element has also been identified in the arginine vasopressin promoter. Vasopressin receptors (V1α and V2 receptors) are present on SCLC cells, and vasopressin can have paracrine effects on the growth of SCLC cells to promote their growth, but increased proliferation in response to vasopressin has not been observed.

It is not known whether the secretion of neurohypophyseal peptides from tumor cells is under regulatory control. Four patterns of vasopressin release have been identified in SIADH. Most commonly (37%), vasopressin levels fluctuate widely and independently of the serum osmolality. In a second group (33%), vasopressin is released in response to changes in osmolality, but the osmotic threshold for vasopressin release is decreased (reset osmostat). Other patients with SIADH manifest a constant leak of vasopressin or have no demonstrable abnormality in vasopressin secretion and could conceivably produce a different antidiuretic substance.

Patients with cancer and SIADH fall into all four categories. It is conceivable that those with a reset osmostat express both the vasopressin gene and an osmoreceptor in their tumors; more likely, however, vasopressin is released centrally in these patients because of stimulation of baroreceptors in the pulmonary bed or periphery, invasion of the vagus nerve, or metastasis to regulatory centers, for example, the hypothalamus.

Expression and secretion of vasopressin are more common than hyponatremia in SCLC. More than 50% of SCLC patients have elevated plasma vasopressin levels, and plasma levels of the neurophysins are increased in 44% to 65% of untreated patients with SCLC. Some tumors also express the gene for oxytocin but no clinical syndrome of inappropriate oxytocin release has been reported in tumor patients.

In patients with elevated vasopressin levels who do not have hyponatremia, abnormalities in water metabolism can be elicited by water loading, which discloses an impaired diuretic response in 47% of patients with limited SCLC and 86% of patients with extensive disease. Patients with milder degrees of vasopressin excess may secrete for reduced free water excretion by reducing fluid intake; only if free water intake exceeds the maximal excretion of free water does hyponatremia result. Thus, the development of hyponatremia is a function not only of the level of vasopressin but also of fluid intake. Although some tumors may secrete abnormally processed forms of vasopressin with reduced biologic activity, it is likely that compensatory mechanisms of this type account for the disparity in the frequency of biochemical and clinical abnormalities in SIADH.

When water is retained, extracellular and intracellular volumes are expanded. The expansion of extracellular fluid volume, probably by causing suppression of aldosterone and an increase in atrial natriuretic peptide (ANP), induces the natriuresis that is characteristic of patients with SIADH who have an adequate intake of sodium. Plasma levels of ANP are normal or high in SIADH; it is not clear whether high ANP levels are a compensatory response to extracellular fluid volume expansion or a consequence of release of ANP from the tumor. Many tumors that express the vasopressin gene also express the gene for ANP. Restriction of sodium intake in patients with SIADH causes weight loss, however, and as the extracellular fluid volume returns to normal, natriuresis is reversed and sodium is conserved appropriately. This suggests that the secretion of ANP is compensatory and that an ANP-induced natriuresis does not contribute significantly to the genesis of hyponatremia.

Acute water retention causes neurologic symptoms by rapidly increasing the intracellular volumes of brain cells and thus inducing cerebral edema. Chronic hyponatremia is probably less symptomatic because there is time for activation of compensatory volume-regulatory mechanisms in the central nervous system. Brain cells compensate for volume gain by activating ion transport processes that pump out intracellular potassium chloride (KCl) and sodium chloride (NaCl). This compensation has therapeutic importance because rapid correction of hyponatremia by infusion of hypertonic saline produces a transient hypertonic encephalopathy as water is drawn out of the already contracted intracellular space. This can cause permanent neurologic damage (e.g., central pontine myelinolysis) and death.
Diagnosis

The diagnosis of SIADH is usually made on clinical grounds. The first step is to establish that the urine is inappropriately concentrated in the presence of hypotonicity. In a hyponatremic patient, a urine osmolality higher than 50 to 60 mmol/L water is inappropriate, and many patients with SIADH have urine osmolalities higher than the plasma osmolality. Next, other causes of hypotonicity must be excluded. The differential diagnosis of hyponatremia includes states of true volume contraction; edematous states such as congestive heart failure and hepatic failure, in which the effective central plasma volume is diminished; adrenal insufficiency; hypothyroidism; drug effects; and SIADH (see Chapter 9).

Measurements of vasopressin are of little value in the differential diagnosis because vasopressin levels are increased in most hyponatremic states. True volume contraction and edematous states can usually be excluded on clinical grounds. It is appropriate to exclude adrenal insufficiency with a corticotropin stimulation test in patients with malignant tumors and SIADH, particularly because patients with bilateral adrenal metastases are at risk for adrenal insufficiency. Thyrotropin should be measured to exclude hypothyroidism. Among the drugs that stimulate the nonosmotic release of vasopressin are the cancer chemotherapeutic agents vincristine, vinblastine, and cyclophosphamide. Normovolemic patients without hormonal disorders or drug causes are presumed to have SIADH.

SIADH is a common cause of hyponatremia in hospitalized patients, but only a small minority of these patients develop the syndrome as the result of inappropriate secretion of vasopressin by a tumor. Furthermore, not all patients with cancer who meet the criteria for SIADH have ectopic secretion of vasopressin from a tumor. Not only are benign forms of SIADH more common, but the response of vasopressin to osmotic stimuli in some patients with malignant tumors is consistent with eutopic secretion of the hormone from the pituitary.

The uncertainty regarding the etiology in an individual patient may be intellectually unsatisfying but is not of great practical importance. Patients with severe and symptomatic hyponatremia are more likely to have a true ectopic source of vasopressin. The treatment of hyponatremia in SIADH is similar regardless of the cause of the syndrome.
Treatment

Symptomatic hyponatremia in patients with a serum sodium level below 120 mmol/L requires immediate treatment (see Chapter 9). The therapeutic options include infusion or administration of hypertonic saline (3% or 5% saline) or of saline and furosemide. The latter regimen has the advantage of not rapidly expanding extracellular fluid volume in an already volume-expanded patient. The goal of acute treatment is to raise the serum sodium level above 125 mmol/L. Such an increase takes the patient out of immediate danger, and further correction can be accomplished in a more leisurely fashion.

How rapidly the initial phase of correction should be carried out is controversial; one school argues for rapid correction in view of the high mortality of the untreated syndrome, and the other suggests that too rapid correction of hyponatremia can predispose to central pontine myelinolysis and other neurologic sequelae. As discussed earlier, the risk of rapid correction probably has to do with brain shrinking in the presence of high concentrations of extracellular sodium and intracellular dehydration, which is exacerbated by prior loss of cell solute, the adaptive response of the central nervous system to hyponatremia (see Chapter 9). Under most circumstances, it seems best to correct hyponatremia at a rate of 0.5 mmol/L per hour until the serum sodium concentration reaches 120 to 125 mmol/L.

In asymptomatic patients or after acute correction of hyponatremia in symptomatic patients, the mainstay of chronic therapy is water restriction. Moderate fluid restriction may be reasonably well tolerated. The goal is to establish a fluid intake at which the intake of free water does not exceed the maximal free water clearance, which is determined by the circulating vasopressin level. If necessary, most patients can be maintained with severe restrictions to 800 to 1000 mL of fluid intake daily. At this level, the free water intake is actually negative because the patient is ingesting osmoles from food in excess of water. Therefore, even a patient who is obliged to excrete a concentrated urine and thus has negative free water excretion may be maintained in zero net water balance. However, severe fluid restrictions are onerous and difficult to maintain.

As an adjunct to water restriction, it can be beneficial to interfere with vasopressin action. The drug of choice for this purpose is demeclocycline, an antibiotic that blocks the action of vasopressin and produces nephrogenic diabetes insipidus. At a dose of 150 to 300 mg four times a day, demeclocycline has a reproducible effect on the urine-concentrating mechanism but up to 2 weeks may be necessary for the full effect. Side effects include photosensitive rashes and liver toxicity.

An alternative, fludrocortisone in doses of 0.1 to 0.3 mg/day, corrects hyponatremia partially but at the risk of causing edema and congestive heart failure. Lithium also produces nephrogenic diabetes insipidus, but its effects are less predictable than demeclocycline or fludrocortisone and the drug should be given only in refractory cases. SCLC is now treated with aggressive combination chemotherapy, and SIADH often remits in responders to chemotherapy.
ECTOPIC CORTICOTROPIN SYNDROME AND ECTOPIC SECRETION OF CORTICOTROPIN-RELEASING HORMONE

Clinical Features

The ectopic corticotropin syndrome accounts for 10% to 20% of cases of Cushing's syndrome. Unlike Cushing's disease, with an 8:1 female preponderance, this syndrome is more common in men than in women. The typical presentation also differs; the onset is sudden, and progression is rapid. Patients complain of proximal myopathy and peripheral edema. Hypertension, hypokalemia, and severe glucose intolerance are often present. Hyperpigmentation may occur, but hirsutism is unusual. Other manifestations of cancer, such as anorexia, weight loss, and anemia, are common.

The somatic features of Cushing's syndrome are notably absent in the typical patient, perhaps because of the rapid evolution of the clinical picture. Patients with slowly growing carcinoid tumors of the bronchus or thymus have a more indolent disease and often present with the classic habitus of Cushing's syndrome—moon facies, centripetal obesity, proximal myopathy, polydipsia, and polyuria. Hyperpigmentation is common in these patients, as is hirsutism in women.

The tumors that produce the ectopic corticotropin syndrome are primarily of neuroendocrine cell origin. In published series, approximately 45% are SCLC, 15% are thymic carcinoids, 10% are bronchial carcinoids, 10% are islet cell tumors, 5% are other carcinoid tumors, 2% are pheochromocytomas, and 1% are ovarian adenocarcinomas. However, adenocarcinoma and squamous carcinoma are also occasionally associated with the syndrome. It appears that SCLC is greatly underrepresented in these referral series; it probably accounts for well over 50% of unselected cases.
Laboratory Features

Both the level of cortisol secretion and the level of corticotropin tend to be higher in ectopic than in pituitary Cushing's syndrome, although there is some overlap. In most ectopic cases, both cortisol and corticotropin levels are elevated to two to four times the normal morning values and the normal diurnal variation in their levels is lost. Urinary excretion of adrenal steroid metabolites is increased correspondingly. Two-site immunoradiometric assays, which are coming into common use for detection of corticotropin, give lower values for corticotropin in ectopic cases than older radioimmunoassays, probably because they do not detect partially processed forms that are common in the ectopic syndrome.

Despite the abnormal processing of corticotropin by nonpituitary tumors, there is no other POMC peptide in serum whose presence is decisive in the diagnosis of the ectopic syndrome. More than one half of nonpituitary tumors that secrete corticotropin also secrete other peptides, including carcinoembryonic antigen, GRP, calcitonin, somatostatin, and corticotropin-releasing hormone (CRH), and the presence of these peptides is suggestive of the ectopic corticotropin syndrome.

Hypokalemia occurs in 80% to 100% of cases in various series, and potassium wasting is more severe than in pituitary Cushing's disease. The hypokalemia is probably explained by the mineralocorticoid effects of cortisol, which are more evident both because cortisol levels tend to be higher in ectopic than in pituitary Cushing's syndrome and because 11-hydroxysteroid dehydrogenase activity appears for unknown reasons to be decreased in patients with ectopic corticotropin secretion. A deficiency of 11-hydroxysteroid dehydrogenase activity impairs the inactivation of cortisol in the renal tubule, leading to increased exposure of mineralocorticoid receptors to cortisol. In disorders such as congenital deficiency of 11-hydroxysteroid dehydrogenase and licorice intoxication, in which the activity of the enzyme is inhibited, normal levels of cortisol produce a state of pseudohyperaldosteronism.
Pathogenesis

Although many nonpituitary tissues contain POMC mRNA, most are short transcripts (800 nucleotides) that are initiated by a downstream promoter at the third exon of the POMC gene and do not include coding sequences for the signal peptide that is necessary for direction of POMC into the secretory pathway. Thus, nonpituitary POMC transcripts probably do not generate bioactive POMC products that can be secreted. In contrast, nonpituitary tumors that secrete corticotropin contain a 1150-nucleotide mRNA similar to the predominant pituitary species, and many nonpituitary tumors also contain 1350-nucleotide transcripts initiated from an upstream promoter that is largely quiescent in pituitary cells.

Two regions of the POMC promoter contribute to activity in SCLC; the region that confers high POMC promoter activity on pituitary cells is not active. One region binds the transcription factor E2F in a methylation-sensitive fashion, and E2F is inactivated by the tumor suppressor gene Rb, which is inactive in 90% of cases of SCLC. Thus, both loss of Rb and differential methylation of the POMC promoter are potential mechanisms of POMC expression in SCLC.

Consistent with the nonsuppressibility of most nonpituitary tumors by glucocorticoids, the sensitivity of POMC gene expression to inhibition by glucocorticoids is reduced in SCLC cell lines. In some cell lines, glucocorticoid receptors are absent; in others, glucocorticoid receptor action appears to be defective. Negative glucocorticoid regulatory elements are typically composite elements that require binding of regulatory factors in addition to the glucocorticoid receptor, leading to the possibility that such accessory factors may be abnormal in transformed cell lines. However, glucocorticoids also fail to stimulate transcription from classic glucocorticoid regulatory elements in SCLC cell lines, and this defect is overcome by overexpression of the wild-type glucocorticoid receptor.

POMC processing in nonpituitary tumors is often incomplete, with the release into blood of POMC fragments with reduced biologic activity. These incompletely processed forms are larger than corticotropin by gel filtration and were first described in the serum of cancer patients with the ectopic corticotropin (ACTH) syndrome as big ACTH. As noted earlier, some incompletely processed POMC peptides can be detected by radioimmunassay techniques for corticotropin but not by two-site immunoradiometric assays. In one study using a precursor-specific assay, the ratio of corticotropin precursors to corticotropin in plasma was 58:1 in the ectopic corticotropin syndrome and 5:1 in pituitary Cushing's disease.

Unusual small peptides are also produced from POMC in nonpituitary tumors. In anterior pituitary corticocyte cells, four of the six dibasic sites in POMC are cleaved by the prohormone convertase PC1/PC3, and the predominant products are six peptides:

- An NH$_2$-terminal peptide
- A joining peptide
- Corticotropin
- Lipotropin (-LPH)
- Smaller amounts of -LPH and -endorphin

Additional products that are detected routinely in extracts of nonpituitary tumors include the corticotropin-like intermediate lobe polypeptide (CLIP) and melanocyste stimulating hormone (MSH) (522). Both peptides are present in the intermediate lobe in the rodent pituitary gland, and their presence in nonpituitary tumors indicates that nonpituitary tumors contain the PC2 convertase, which is normally present in intermediate but not anterior pituitary cells. These peptides are not secreted in large amounts and are not useful as tumor markers in blood, but the serum LPH/corticotropin ratio in the ectopic corticotropin syndrome is higher than in pituitary tumors, possibly reflecting the increased PC2 activity in nonpituitary tumors.

Corticotropin-like activity can be identified in extracts of many nonSCLCs and in virtually all SCLCs, and at least one third of all SCLCs show POMC mRNA by in situ hybridization. Yet only 1% to 3% of patients with SCLC have clinical evidence of corticotropin excess. Thus, the ectopic corticotropin syndrome is a good example of the principle that ectopic production of hormones is more common than clinical syndromes of hormone excess. Differences between tumors in trans-acting nuclear factors or epigenetic regulation of the POMC promoter by DNA methylation may account for differential expression of the POMC gene.

Patients with corticotropin-producing neoplasms are protected from the consequences of hormone excess by several mechanisms. Malignant tumors contain much smaller quantities of POMC mRNA and peptides than the pituitary and are thus inefficient in producing corticotropin. Tumor cells are much poorer in neurosecretory granules than pituitary corticotropes and are relatively deficient in the ability to process POMC efficiently and secrete the peptide products. Inefficient cleavage of POMC leads to incompletely processed forms of corticotropin with little biologic activity. Processing of POMC by tumors can also lead to production of biologically inactive products. For example, some tumors produce significant amounts of corticotropin but cleave it to the CLIP.
Laboratory Diagnosis

The diagnosis of the ectopic corticotropin syndrome is described in Chapter 8 and Chapter 14.

The first step consists of determining whether cortisol excess is present and whether it is corticotropin-dependent. Increased basal cortisol secretion can often be shown by measurement of serum cortisol or urinary free cortisol, both of which are increased in the ectopic corticotropin syndrome. When the basal levels are not markedly increased, the presence of cortisol excess can be established with a low-dose dexamethasone suppression test, for example, the 1-mg overnight dexamethasone suppression test. The corticotropin dependence of cortisol excess can be established by measurement of corticotropin in the same sample in which cortisol is measured.

Corticotropin-dependent Cushing's syndrome results from either pituitary or ectopic secretion of corticotropin. In the classic form of the syndrome (e.g., corticotropin-secreting SCLC), secretion is nonsuppressible and there is little or no response of serum or urinary cortisol to the administration of high-dose dexamethasone, whereas the secretion of corticotropin by pituitary adenomas is dexamethasone-responsive. In a patient with a recognized malignancy and clinical features suggestive of the ectopic syndrome, the finding of nonsuppressible hypercortisolism usually suffices to make the diagnosis.

In occasional lung cancers and in about 50% of bronchial or thymic carcinoid tumors, the secretion of corticotropin can be suppressed with high-dose dexamethasone. This circumstance has been called the occult ectopic corticotropin syndrome and presents a major diagnostic challenge because the clinical presentation and secretory dynamics may be identical to those of pituitary Cushing's syndrome and because neither these small tumors nor pituitary corticotroph adenomas may be evident on routine radiologic studies.

Although a number of noninvasive methods are used in this circumstance, none is definitive. Because nonpituitary tumors are not as well suppressed by glucocorticoids as corticotrope adenomas of the pituitary gland, it is useful to apply stringent criteria for glucocorticoid suppressibility, namely suppression of urinary free cortisol by more than 80% after administration of high-dose dexamethasone, which has a sensitivity of 81% and a specificity of 92% for pituitary Cushing's syndrome. Stimulation with the ovine CRH test is valuable because nonpituitary tumors do not respond well to CRH. An increase in plasma corticotropin of 35% after administration of ovine CRH (1 µg/kg body weight) was reported to have a sensitivity of 93% and a specificity of 100% for the diagnosis of pituitary Cushing's syndrome. Metyrapone testing in combination with high-dose dexamethasone may also improve diagnostic accuracy.

The definitive study for distinguishing pituitary from nonpituitary forms of hypercortisolism is inferior petrosal sinus sampling with administration of ovine CRH. The ratio of corticotropin in the inferior petrosal sinus to that in peripheral blood after administration of CRH is greater than 3 in patients with pituitary tumors and less than 2 in patients with corticotropin-secreting nonpituitary tumors. Cavernous sinus sampling has also been successful. Localization of bronchial and thymic carcinoids may also be difficult. Thin-section computed tomography of the chest and scanning with labeled octreotide have sometimes been of use but have a high failure rate.
Treatment

The management of Cushing's syndrome is discussed in Chapter 8 and Chapter 14.

When possible, the treatment of the ectopic corticotropin syndrome is surgical. With slow-growing carcinoid tumors of the bronchus, thymomas, or pheochromocytomas, surgical resection can be curative. If the tumor cannot be identified, it is necessary to block cortisol secretion with adrenolytic agents. Some patients ultimately require surgical adrenalectomy to control hypercortisolism.

Malignant nonpituitary neoplasms that secrete corticotropin are rarely amenable to resection because the tumor is usually advanced and inoperable by the time the clinical syndrome appears. With malignant neoplasms the aim is to palliate hypercortisolism by medical adrenalectomy using adrenolytic drugs, such as aminogluthethimide (250 mg three times a day) or metyrapone (250 to 500 mg three times a day). Ketoconazole (200 to 400 mg twice a day) has also been useful for the treatment of ectopic corticotropin syndrome. A replacement dose of hydrocortisone should be administered with these drugs to avoid adrenal insufficiency. Some patients respond to the long-acting somatostatin agonist octreotide, and the glucocorticoid antagonist mifepristone has also been used.
Ectopic Secretion of Corticotropin-Releasing Hormone

Nonendocrine tumors rarely cause Cushing's syndrome by secretion of CRH. Patients have increased CRH levels in tumor tissue or in plasma and high plasma corticotropin levels. It is important to document that the site of corticotropin secretion is the pituitary gland because many nonendocrine tumors that secrete CRH also secrete corticotropin itself. Presumptive evidence of a pituitary source of corticotropin may come from demonstration that the gradient of corticotropin between the inferior petrosal sinus and peripheral blood is more than 3:1, from finding pituitary corticotropic hyperplasia in patients who underwent pituitary surgery for a presumed corticotropic adenoma, or from the failure to detect corticotropin in the nonendocrine tumor. When the nonendocrine tumor secretes both CRH and corticotropin, the true role of CRH in the clinical syndrome may be indeterminate.

Cushing's syndrome resulting from ectopic secretion of CRH does not have a distinctive presentation. In most cases, the hypercortisolism is unresponsive to dexamethasone suppression, but a normal response to high-dose dexamethasone has also been reported. The response to metyrapone is also variable. Tumors that secrete CRH include small cell carcinomas of the prostate and lung, medullary thyroid carcinoma, carcinoids, and a hypothalamic gangliocytoma. These neuroendocrine tumors are similar to the tumors that cause Cushing's syndrome by direct secretion of corticotropin.

The diagnosis of ectopic CRH secretion as the cause of Cushing's syndrome is usually made retrospectively. In view of the rarity of the disorder, it is probably inappropriate to measure CRH routinely in Cushing's syndrome. However, it may be worthwhile to determine the plasma CRH level when pituitary surgery has disclosed diffuse corticotropic hyperplasia in a patient with Cushing's syndrome.
HYPOGLYCEMIA WITH NONISLET CELL TUMORS

Clinical Features

Fasting hypoglycemia produced by nonislet cell tumors typically causes neuroglycopenic symptoms of obtundation, confusion, or behavioral aberrations, which may have been present for some time before the diagnosis is made. Nonislet cell tumors rarely secrete insulin, but a case of small cell cervical carcinoma with high levels of insulin, proinsulin, and C peptide has been reported; the tumor was found to contain insulin mRNA by in situ hybridization and immunoreactive insulin by immunohistochemical methods.

Most extrapancreatic tumors that cause hypoglycemia do so by secreting IGF-II. The offending neoplasms are usually bulky, slow-growing mesenchymal tumors. Fibrosarcomas, rhabdomyosarcomas, leiomyosarcomas, mesotheliomas, and hemangiopericytomas account for more than 50% of cases. Hepatocellular carcinomas (hepatomas), carcinoid tumors, and adrenocortical carcinomas account for about 25% of cases, and the remainder are made up of various carcinomas, leukemias, and lymphomas. More than one third of the tumors are retroperitoneal, about one third are intra-abdominal, and the remainder are intrathoracic.
Pathogenesis

Fasting hypoglycemia produced by nonislet cell tumors results from increased peripheral utilization of glucose, primarily in skeletal muscle, coupled with decreased hepatic glucose output.\textsuperscript{220,221,222} Lipolysis is inhibited, and free fatty acid levels are low. Although it had been suspected that bulky tumors themselves, sometimes weighing many kilograms, might metabolize enough glucose to exceed the capacity for hepatic glucose production, this phenomenon has not been documented. Despite insulin-like effects on glucose utilization, hepatic glucose production, and lipolysis, fasting insulin levels during hypoglycemia are appropriately suppressed. For this reason, it has appeared that an insulin-like factor is probably responsible for hypoglycemia.

Sera from patients with nonislet cell tumors contain elevated levels of an insulin-like activity by radioreceptor assay.\textsuperscript{223} The level of IGF-II mRNA in nonislet cell tumors is often increased, even in patients with normal serum IGF-II levels.\textsuperscript{224,225,226} IGF-II levels are sometimes elevated during hypoglycemia but may be normal,\textsuperscript{227} and the levels of IGF-I are typically suppressed.\textsuperscript{228} Although not all patients have high IGF-II levels, it appears that IGF-II is in fact the causative agent of hypoglycemia and can cause hypoglycemia at normal total serum levels as a consequence of altered processing and increased bioavailability.

Altered binding of IGF-II in the tumor-hypoglycemia syndrome increases its bioavailability to peripheral receptors. In normal serum, IGFs are bound largely in one of two complexes. Most IGF is normally bound to a heterotrimeric 150-kd complex consisting of the IGF, the binding protein IGFBP3, and an acid-labile glycoprotein. The large complex is retained in the circulation, and as a result the half-life of the IGF-II complex is relatively long, 12 to 15 hours. A minority of IGF circulates in a smaller 50-kd complex that contains mainly IGF and a different binding protein, IGFBP2. The small complex can cross capillaries and deliver IGF to tissue receptors, and IGF-II bound to this complex has a half-life of only about 30 minutes.\textsuperscript{230} In sera from patients with nonislet cell tumors and hypoglycemia, the fraction of IGF-II bound to the small, bioavailable complex is increased, on average by threefold.\textsuperscript{231,232,233} Presumably increasing the access of IGF-II to the receptor, even in the setting of normal total IGF-II levels.

A substantial fraction of IGF-II in both tumors and sera is present in a high-molecular-weight form, big IGF-II,\textsuperscript{225,226,227} a partially processed form that contains a 21-amino-acid carboxyl-terminal extension from the E domain.\textsuperscript{227} Big IGF-II was reported to lack O-linked glycosylation,\textsuperscript{234} which may give it increased bioactivity,\textsuperscript{235} but O-linked glycosylation was reported to be normal in another study.\textsuperscript{236} Whether a decreased ability of big IGF-II to form a normal ternary complex contributes directly to its increased bioavailability remains to be determined.

Current concepts of the alteration in IGF-II binding are summarized in Figure 40-3.\textsuperscript{8,216} Oversecretion of big IGF-II suppresses the secretion of insulin, growth hormone (GH), and IGF-I.\textsuperscript{237} In turn, suppression of GH and IGF-I down-regulates the synthesis of IGFBP3 and the acid-labile subunit, both of which are GH-dependent,\textsuperscript{238} and up-regulates the synthesis of IGFBP2. Consistent with this proposal regarding the role of GH is the response of a patient to GH therapy.\textsuperscript{239} Thus, IGF-II oversecretion leads to altered binding and increased bioactivity of IGF-II and can cause hypoglycemia even when total IGF-II levels are normal. The level of free IGF-II in serum is also increased.\textsuperscript{240}
Laboratory Diagnosis

The fasting levels of insulin and C peptide are appropriately suppressed in samples obtained during hypoglycemia (insulin < 36 pmol/L [6 µU/mL], C peptide < 0.2 nmol/L [0.6 ng/mL]). The IGF-II level may be normal or increased. In patients with normal levels of IGF-II, the diagnosis is supported by finding low levels of IGF-I, GH, and IGFBP3.
Treatment

The mainstay of treatment is resection of the tumor. Even partial debulking may ameliorate hypoglycemia. In patients with unresectable tumors, several maneuvers based on the pathogenetic scheme have been attempted. Therapy with GH, glucagon, glucocorticoids, or somatostatin has been effective in individual patients with unresectable tumors. These measures are temporary, however, until the unresectable tumor can be treated with chemotherapy.
SYNDROMES CAUSED BY GROWTH HORMONERELEASING HORMONE, GROWTH HORMONE, AND HUMAN PLACENTAL LACTOGEN

Since 1980, more than 40 cases of acromegaly have been associated with nonpituitary tumors. There are only two well-documented cases of acromegaly resulting from secretion of GH by a nonpituitary tumor: the other tumors caused acromegaly by secreting growth hormonereleasing hormone (GHRH), which was first isolated from extracts of pancreatic tumors.

Overall, secretion of GHRH accounts for less than 1% of cases of acromegaly. The clinical findings, aside from the presence of a nonpituitary tumor, do not differ from those in acromegaly caused by somatotropin adenomas. The mean duration of acromegalic features before diagnosis is 7.9 years, about the same as in pituitary acromegaly. Diabetes mellitus, amenorrhea, and galactorrhea are common. In about 50% of cases, the extrapituitary neoplasm is symptomatic. Other syndromes of hormone excess, including Cushings syndrome, primary hyperparathyroidism, and the Zollinger-Ellison syndrome, may occur in conjunction with acromegaly.

Carcinoids are the most common extrapituitary tumors that produce acromegaly (69% of cases), followed by islet cell tumors (23%), pheochromocytoma, and paraganglioma. GHRH immunoreactivity can frequently be demonstrated in neuroendocrine tumors from patients without acromegaly, usually in smaller amounts than in tumors associated with acromegaly. However, high plasma levels of GHRH have been reported in SCLC patients without acromegaly. Some of these patients have abnormal GH secretory dynamics, such as a paradoxical GH increase after administration of thyrotropin-releasing hormone (TRH), suggesting that subclinical or incomplete forms of acromegaly may be present. All three isoforms of GHRH have been identified in nonpituitary tumors, but the predominant species in most tumors is GHRH(140), whereas the dominant hypothalamic form is GHRH(144).

Serum levels of GHRH in acromegaly caused by extrapituitary tumors are markedly elevated, from 0.3 to 5 µg/L (0.3 to 50 ng/mL). Normal fasting GHRH levels are less than 60 ng/L (0.06 ng/mL), and the peripheral level is less than 200 ng/L (0.2 ng/mL) in typical acromegaly.

The dynamics of GH secretion in acromegaly induced by nonpituitary secretion of GHRH are not distinctive. GH and IGF-I levels are high, and the normal circadian rhythm of GH secretion is lost. Prolactin levels are elevated in 80% of patients. Virtually all patients display a paradoxical increase of GH after administration of TRH, compared with approximately 40% of patients with classical acromegaly. Many patients with GHRH-induced acromegaly do not respond to exogenous GHRH, but this is not a uniform finding and cannot be used diagnostically. In most cases, the nonpituitary tumor can be identified by imaging studies of the chest and abdomen. About 90% of carcinoid tumors that cause acromegaly are located in the chest (see Chapter 8).

The primary therapy is surgical. About half of patients have resectable tumors. For patients with nonresectable disease, the therapy of choice is octreotide, the somatostatin agonist. In about half of patients, GH levels return to normal with octreotide or lanreotide treatment, and most of the remainder have a partial response. The level of GHRH is often reduced less than that of GH, which suggests that the drug affects primarily the pituitary response to GHRH.

One case of extrapituitary acromegaly caused by nonpituitary secretion of GH itself involved a pancreatic islet cell tumor that contained both GH and GH mRNA. At surgery, an arteriovenous GH gradient was demonstrated across the tumor, and tumor cells in culture secreted immunoreactive GH. GH was not detectable in plasma. The other well-documented case was a follicular non-Hodgkin’s lymphoma in which the cells expressed the GH gene, but not the GHRH gene, and secreted large amounts of GH in culture. Hypersecretion of GH was abolished by successful chemotherapy of the lymphoma.

The propensity of other members of the GH family for secretion by nonendocrine tumors is variable. Prolactin has not been conclusively shown to be secreted into the blood by nonpituitary tumors. However, prolactin and its receptor are expressed in human breast carcinoma cells, and an autocrine pathway has been described in which prolactin induces constitutive phosphorylation of erbB-2 (Her/Neu). Neoplastic production of human placental lactogen (hPL) (or chorionic somatomammotropin) appears to be relatively common. In large series, hPL was detectable in plasma in 9% of patients with malignant disease, most commonly lung carcinoma but also carcinoma of the thyroid, breast, stomach, pheochromocytoma, carcinoids, and leukemia; 14% of patients with breast carcinoma had increased blood hPL levels. hPL has weak GH activity but substantial lactotrophic activity. Patients with elevated hPL levels in blood do not have galactorrhea, however, because the circulating hPL levels in such patients are lower than the equivalent levels of prolactin that produce galactorrhea. It is also possible that neoplasms secrete hPL in biologically inactive forms.
SYNDROMES CAUSED BY HUMAN CHORIONIC GONADOTROPIN

hCG is produced eutopically by trophoblastic and germ cell tumors, including testicular embryonal carcinoma and extragonadal germinomas. For these tumors, hCG is a useful tumor marker. Gynecomastia has been reported in a few adult patients with nontrophoblastic malignancies, and isosexual precocious puberty has occurred in children (see Chapter 21). The associated tumors include lung carcinoma, hepatocellular carcinoma, adrenocortical carcinoma, and renal carcinoma.

Although hCG is secreted by approximately 18% of nontrophoblastic tumors (Table 40-5), it is not useful as a tumor marker for screening for nontrophoblastic neoplasms. A significant number of normal individuals, including 2.4% of blood donors and 3.6% of patients with various benign diseases, also have detectable levels of hCG (assayed as hCG) in serum, decreasing the positive predictive value of a detectable hCG level.

Some tumors secrete free subunits of the glycoprotein hormones. For example, about 50% of malignant islet cell tumors secrete the subunit of glycoprotein hormones, whereas benign islet cell tumors rarely do so. Immunohistochemical staining for the subunit has also been used to distinguish benign from malignant islet cell tumors histologically. A degradation product of hCG in urine, the -core fragment, has been reported in about 50% of patients with lung cancer, but the fragment is also detected at low levels in 6% to 14% of individuals with benign disorders; its clinical value remains to be established.

### Table 40-5: Serum Human Chorionic Gonadotropin Levels in Patients with Cancer

<table>
<thead>
<tr>
<th>Tumor or Site</th>
<th>Positive (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Islet cell</td>
<td>39.4</td>
</tr>
<tr>
<td>Gynecologic</td>
<td>28.9</td>
</tr>
<tr>
<td>Carcinoid</td>
<td>26.8</td>
</tr>
<tr>
<td>Gastrointestinal</td>
<td>18.0</td>
</tr>
<tr>
<td>Lung</td>
<td>17.4</td>
</tr>
<tr>
<td>Breast</td>
<td>16.8</td>
</tr>
<tr>
<td>Melanoma</td>
<td>23.9</td>
</tr>
<tr>
<td>Genitourinary</td>
<td>11.8</td>
</tr>
<tr>
<td>Sarcoma</td>
<td>11.8</td>
</tr>
<tr>
<td>Hematopoietic</td>
<td>6.1</td>
</tr>
</tbody>
</table>

ONCOGENOUS OSTEOMALACIA

More than 50 cases of hypophosphatemic osteomalacia or rickets have been reported in patients with tumors of mesenchymal origin, usually small, benign skeletal tumors of the extremities or head. Histologically, the causative tumors include hemangiopericytoma, ossifying and nonossifying fibroma, and giant cell tumors. Hypophosphatemic osteomalacia has also been reported in patients with disseminated prostatic carcinoma, a situation in which the disorder may be due to phosphate uptake by osteoblastic metastases.

Most patients with oncogenous osteomalacia are middle-aged and present with bone pain and proximal myopathy, which may have been present for years before diagnosis. However, the disorder has been described in children. The serum phosphorus level is markedly reduced because of renal phosphate wasting. The serum alkaline phosphatase is increased, but the serum calcium and PTH levels are normal. The level of 1,25(OH)₂ D is typically low and the level of 25-hydroxyvitamin D is normal. Osteomalacia is present in bone biopsies. The fact that the syndrome is reversed by resection of the tumor indicates that it has a humoral basis.

Although rare, oncogenous osteomalacia is probably the most common cause of acquired hypophosphatemic osteomalacia. The manifestations are similar to those of hereditary phosphate-wasting disorders such as X-linked hypophosphatemic rickets, and the primary event is probably the induction of severe phosphaturia by a humoral factor, phosphatonin, secreted by the tumor.

The mutation responsible for one of the hereditary forms of renal phosphate wasting, autosomal dominant hypophosphatemic rickets, has been localized to a new member of the fibroblast growth factor family, FGF-23. Each of six tumors from patients with oncogenous osteomalacia expressed high levels of FGF-23 mRNA or protein, and FGF-23 induces phosphate wasting. The gain-of-function mutations in FGF-23 that cause autosomal dominant hypophosphatemic rickets prevent its degradation. It is likely that the responsible protease is Phex; loss of this protease produces a third form of phosphate wasting, X-linked hypophosphatemia.
SYNDROMES CAUSED BY OTHER HORMONES

Erythropoietin and Erythrocytosis

Erythrocytosis occurs in 1% to 4% of renal carcinomas, 5% to 10% of hepatocellular carcinomas, and 10% to 20% of cerebellar hemangioblastomas. It has been observed in patients with uterine fibromyomas, adrenocortical carcinomas, or ovarian tumors. Renal and hepatocellular carcinomas account for 71% of cases; thus, erythrocytosis is most common in tumors arising from the tissues that normally secrete erythropoietin (the fetal liver and by the adult kidney).

The association of erythropoietin and vascular endothelial growth factor secretion from cerebellar hemangioblastoma and renal carcinoma raised the possibility that both of these hypoxia-responsive genes are specifically activated by loss of the von Hippel-Lindau tumor suppressor gene VHL. The VHL protein appears to stabilize mRNA for hypoxia-inducible proteins. VHL-negative renal carcinomas display constitutively high levels of the hypoxia-inducible factor-1 (HIF-1), a heterodimeric member of the bHLH PAS family of transcription factors, composed of \( \alpha \) and \( \beta \) subunits. The action of the VHL protein to increase the proteasome-mediated degradation of the HIF-1 protein and other proteins may prove central to many manifestations of VHL gene alteration.

The VHL gene also displays loss of heterozygosity in some hepatocellular carcinomas, and tissue-specific inactivation of the VHL gene in hepatocytes gives rise to cavernous hemangioma of the liver and up-regulation of the erythropoietin gene. It is thus likely that inactivation of VHL plays a central role in the induction of erythropoietin synthesis in each of the tumors classically associated with erythropoietin-dependent polycythemia, by either transcriptional or post-transcriptional mechanisms.

Early studies reported that erythropoietic bioactivity was frequently present in tumor extracts from polycythemic patients; erythropoietin mRNA has been demonstrated in extracts of renal carcinomas, hepatocellular carcinoma, and cerebellar hemangioblastoma. Erythrocytosis has been produced in nude mice by transplantation of erythropoietin-positive renal carcinoma cells and hepatocarcinoma cells. Some patients with tumors and erythrocytosis have increased serum erythropoietin levels. However, in the best studied group, patients with hepatocellular carcinoma, it has been difficult to demonstrate a consistent relationship between the red blood cell mass and serum levels of erythropoietin. Increased serum levels of erythropoietin are common in patients with hepatocellular carcinoma; however, few of the patients with high erythropoietin levels have erythrocytosis, and some patients with erythrocytosis have normal levels of erythropoietin. Absence of erythrocytosis in the presence of high erythropoietin levels could reflect secretion of biologically inactive (e.g., precursor) forms of erythropoietin.
Calcitonin

Calcitonin is present in neuroendocrine cells of the normal bronchial epithelium and is frequently secreted by neuroendocrine tumors, including 18% to 60% of SCLCs. Calcitonin is also secreted by other lung carcinomas, breast cancers, leukemias, and a broad spectrum of other neoplasms. Estimates of the frequency of calcitonin secretion are lower in studies that rigorously control for assay artifacts, but it is clear that some tumors express the gene for calcitonin/calcitonin generelated peptide (CGRP) and secrete calcitonin in vitro.

Tumors frequently secrete large forms of calcitonin and are less sensitive to stimulation than in patients with hypercalcitoninemia resulting from medullary thyroid carcinoma. CGRP, which is derived from alternative splicing of the calcitonin gene, is expressed in normal bronchial epithelium and has been detected in tumor extracts and serum. The levels of calcitonin in the sera of patients with lung carcinoma are lower than those in medullary thyroid carcinoma, and no clinical syndrome is associated with the secretion of calcitonin or CGRP.
Endothelin

The potent vasoconstrictor peptide endothelin is expressed in hepatocellular carcinoma, breast carcinoma, ovarian carcinoma, and prostate carcinoma. Endothelin receptors are often coexpressed on tumor cells, and endothelin-1 was reported to have paracrine effects on tumor cell growth. Increased serum levels of endothelin-1 and a partially processed form of the peptide big endothelin-1 have been found in hepatocellular carcinoma. Arteriovenous differences, albeit small, have been found across the liver of patients with hepatocellular carcinoma. No clinical manifestations of systemic secretion of endothelin have been reported, but endothelin-1 has been implicated as the factor causing the osteoblastic response to bone metastasis in breast carcinoma, and a similar role was suggested in prostate carcinoma.
Vasoactive Intestinal Peptide

Inappropriate secretion of VIP produces pancreatic cholera, also known as the WDHA syndrome (watery diarrhea, hypokalemia, and achlorhydria) or Verner-Morrison syndrome (see Chapter 33). In addition to pancreatic islet cell tumors, other neuroendocrine tumors, including ganglioneuroma, ganglioneuroblastoma, neuroblastoma, pheochromocytoma, and medullary thyroid carcinoma, can produce the syndrome. These tumors stain for VIP, and removal of the tumor causes return of peripheral VIP levels to normal and reverses the clinical syndrome. Increased VIP levels have also been reported in lung carcinoma and in a neuroendocrine tumor of the kidney. VIP is present in the central and peripheral nervous systems; thus, its production by neuroendocrine tumors may be regarded as eutopic rather than ectopic.
Other Gut Hormones

Somatostatin is frequently detectable in extracts of lung tumor [305] but is secreted by cultured SCLC cells, [306] but elevated serum somatostatin concentrations are uncommon in lung cancer (see Chapter 35). [307] Only one case of the somatostatinoma syndrome has been attributed to SCLC. [308] The glucagonoma syndrome occurred in a patient with a renal neuroendocrine tumor [309] and in a patient with a large cell lung carcinoma. [310] The glucagonoma syndrome occurred in a patient with a renal neuroendocrine tumor and in a patient with a large cell lung carcinoma.

GRP is often found in lung carcinomas, cultured SCLC cells, and other tumors but elevated serum levels are uncommon. [311] Pro-GRP may be a better tumor marker. [312] The peptide is a mitogen for SCLC cells, [313] and neutralizing studies with antibodies and antagonists suggested that it has an autocrine role as a growth factor. GRP was also reported to affect the motility of tumor cells. A variant form of the GRP receptor is expressed in human lung carcinoma cell lines. GRP is expressed in neuroendocrine cells of bronchial mucosa, particularly at branch points, and appears to have a developmental role in the regulation of branching morphogenesis of airways. Pancreatic polypeptide is occasionally detectable in the sera of patients with carcinoid tumors.

Pancreatic polypeptide is occasionally detectable in the sera of patients with carcinoid tumors.
References


1830

1830

1830


References


236. Daughaday WH, Trivedi B, Baxter RC. Abnormal serum IGF-II transport in nonislet cell tumor hypoglycemia results from abnormalities of both IGF binding protein-3 and acid label subunit and leads to elevation of serum free IGF-II. Endocrinology 1995; 133:262733.


Chapter 41 - Carcinoid Tumors, Carcinoid Syndrome, and Related Disorders

Kjell öberg

The first clinical and histopathologic description of carcinoid tumor was made by Otto Lubarsch in 1888. He was impressed by the multicentric origin of carcinoid tumors of the gastrointestinal tract, their lack of gland formation, and their lack of similarity with the usual adenocarcinoma of the alimentary system.

The term Karzinoide was introduced in 1907 by the pathologist Oberndorffer as a descriptive name for what he considered to be a "benign" type of neoplasm of the ileum, which could nevertheless behave like a carcinoma. It was subsequently generally accepted that the carcinoid tumor was a very-slow-growing and benign neoplasm with no potential for invasiveness and no tendency to give rise to metastases. This myth of benignity has survived to the present, even though in 1949 Pearson and Fitzgerald described a large series of metastasizing carcinoid tumors.

Carcinoid tumors have subsequently been reported in a wide range of organs, but they most commonly involve the lungs and gastrointestinal tract. Carcinoid tumors of the thymus, ovaries, testes, heart, and middle ear have also been described. The clinically well-known carcinoid syndrome was described by Thorson and associates in 1954; 1 year earlier, Lembeck had extracted serotonin from a carcinoid tumor.
PHYLOGENESIS AND EMBRYOLOGY

Carcinoid tumors are derived from neuroendocrine cells, and Gosset and Masson in 1914 were the first to point out the neuroendocrine properties of carcinoid tumors. Masson later described the remarkable affinity for silver salts displayed by intracytoplasmic granules in tumor cells and noted that carcinoid tumors originate from enterochromaffin cells, the so-called Kulchitsky cells in the crypts of Lieberkühn in the intestinal epithelium. Furthermore, he suggested that the tumors were of endocrine origin (Fig. 41-1).

The mammalian gastrointestinal tract and pancreas contain a large number of endocrine cell types, which initially were thought to originate from the neuroectoderm. This observation gave rise to the APUD concept (amine precursor uptake and decarboxylation) because of the ability of these cells to take up and decarboxylate amino acid precursors of biogenic amines such as serotonin and catecholamines. It was later revised by others who postulated that these endocrine cells might also be derived from mesoderm and endoderm. The neuronal phenotype is clearly seen when culturing carcinoid tumor cells in vitro. The enterochromaffin cells, from which many carcinoid tumors derive, have the property of producing and secreting amines (such as serotonin) and polypeptides (such as neurokinin-A and substance P).

Carcinoid tumors may also originate from other neuroendocrine cells, such as the enterochromaffin-like (ECL) cells of the gut and endocrine cells in the bronchi. The tumors derived from these cells are able to produce a wide range of hormones, such as gastrin, gastrin-releasing peptide (GRP), calcitonin, pancreatic polypeptide, adrenocorticotropic hormone (ACTH), corticotropin-releasing hormone (CRH), and growth hormonereleasing hormone (GHRH) as well as somatostatin, glucagon, and calcitonin gene related peptide (CGRP). A common secretory product from all types of carcinoid tumors is the glycoprotein chromogranin-A (CgA) the most important general tumor marker in these patients (see later).
MOLECULAR GENETICS

Despite advances in the diagnosis, localization, and treatment of carcinoid tumors, no etiologic factor associated with the development of these tumors has been identified. Little is known about molecular genetic changes underlying tumorigenesis. Sporadic foregut carcinoids as well as the familial-type multiple endocrine neoplasia type 1 (MEN-1) frequently display allelic losses at chromosome 11q13, and somatic MEN-1 gene mutations has been reported in one third of sporadic foregut tumors. In contrast with foregut carcinoids, molecular and cytogenetic data for midgut carcinoids are quite limited, and these tumors are not included in MEN-1 syndrome. Deletions of chromosomes 18q and 18p have been reported in 38% and 33%, respectively, of gastrointestinal carcinoids.

In one recent publication, deletions on chromosome 18 were found in 88% of midgut carcinoid tumors but Smad 4/DPC4 locus was not deleted. In addition to the consistent finding of deletions on chromosome 18, multiple deletions on other chromosomes (4, 5, 7, 9, 14, 20) were noticed in single tumors. The region telomeric to Smad 4/DPC 4/DCC loci must be further explored for possible losses of a tumor suppressor gene in this area.

| TABLE 41-1 -- Classification of Carcinoid Tumors |
|---------------------|---------------------|---------------------|
| **Histopathology**  | **Foregut**          | **Midgut**          | **Hindgut**         |
| Argyrophilic        | Argentaffin-positive | Argyrophilic        |
| CgA-positive        | CgA-positive         | SVP-2-positive      |
| NSE-positive        | NSE-positive         | (CgA-positive NSE-positive) |
| Synaptophysin-positive | Synaptophysin-positive | (Synaptophysin-positive) |
| **Molecular Genetics** | Chromosome 11q13 deletion | Chromosome 18q, 18p deletion | Unknown |
| **Secretory Products** | CgA, 5-HT, 5-HTP, histamine, ACTH, GHRH, CGRP, somatostatin, AVP, glucagon, gastrin, NKA, substance-P, neurotensin, GRP | CgA, 5-HT, NKA, substance-P, prostaglandins E₁ and F₂ , bradykinin | PP, PYY, somatostatin |
| **Carcinoid Syndrome** | Present (30%) | Present (70%) | Absent |

ACTH, adrenocorticotropic hormone; AVP, arginine vasopressin; CgA, chromogranin-A, CGRP, calcitonin gene-related peptide; GHRH, growth hormone-releasing hormone; GRP, gastrin-releasing peptide; 5-HT, 5-hydroxytryptamine; 5-HTP, 5-hydroxytryptophan; NKA, neurokinin; NSE, neuron-specific enolase; PP, pancreatic peptide; PYY, peptide YY; SVP2, synaptic vesicle protein 2.
CLASSIFICATION

In 1963, Williams and Sandler reported a relationship between the embryonic origin of carcinoid tumors and the histologic, biochemical, and, to some extent, clinical features of the tumors. Three distinct groups were formed:

1. Foregut carcinoids (i.e., intrathoracic, gastric, and duodenal carcinoids).
2. Midgut carcinoids (carcinoids of the small intestine, appendix, and proximal colon).
3. Hindgut carcinoids (carcinoid tumors of the distal colon and rectum).

Although this original classification has been useful in the clinical assessment of patients with carcinoid tumors, it has demonstrated significant shortcomings. As a result, many investigators have adopted a new classification system that takes into account not only the site of origin but also variations in the histopathologic characteristics of carcinoid tumors. In this revised system, typical tumors are classified as well-differentiated neuroendocrine tumors with their characteristic growth pattern. These tumors are usually slow-growing, with low proliferation capacity (proliferation index <2%). They are usually confined to the mucosa and submucosa and are less than 1 to 2 cm in diameter. Tumors with increased nuclear atypia and high proliferation index (>10%) have been termed atypical or anaplastic carcinoids and have been subclassified to well-differentiated and poorly differentiated neuroendocrine carcinomas, depending on the growth pattern.

The incidence of carcinoid tumors is similar in Western countries and is estimated to be 2.8 to 21 per 1 million people. Because many carcinoid tumors are indolent, the true incidence may be higher. In particular, appendiceal carcinoids have not been included in many studies, but the high incidence of 21 per 1 million was found in an autopsy study when appendiceal carcinoids were included. The incidence of patients with a carcinoid syndrome is about 0.5 per 100,000. Data from the United States, based on results from the End Results Group and the Third National Cancer Survey, 1950 to 1969 and 1969 to 1971, respectively, found that the appendix was the most common site of carcinoid tumors, followed by the rectum, ileum, lungs, and bronchi.

An analysis done in the Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute between 1973 and 1991 reported an increase in the proportion of pulmonary and gastric carcinoids and a decrease in the proportion of appendiceal carcinoids.
BIOCHEMISTRY

The production of hormones appears to be a highly organized function of carcinoid cells. In 1953, Lembeck isolated serotonin from a carcinoid tumor; since then, the carcinoid syndrome has been related to serotonin overproduction. The biosynthesis of serotonin and its metabolic degradation are outlined in Figure 41.3.

Carcinoid tumors of the midgut and foregut region with metastatic disease secrete serotonin and show elevated urinary excretion of 5-hydroxyindoleacetic acid (5-HIAA) in 75% and 30%, respectively. Carcinoid tumors arising from the foregut, however, frequently have low levels of L-amino-acid decarboxylase, which converts 5-hydroxytryptophan (5-HTP) to serotonin. Thus, these tumors secrete primarily 5-HTP.

For many years, it was believed that the entire carcinoid syndrome could be explained by the secretion of these biologically active amines. However, further studies have indicated that serotonin is mainly involved in the pathogenesis of diarrhea and that other biologically active substances play a more important part in the carcinoid flush and bronchoconstriction.

Oates and associates proposed that kallikrein, an enzyme found in carcinoid tumors, is released in association with flush and stimulates plasma kininogen to liberate lysyl-bradykinin and bradykinin. These are biologically active substances that cause vasodilation, hypotension, tachycardia, and edema. Furthermore, prostaglandins (E1, E2, F1α, F2α) may also play a role in the carcinoid syndrome. Gastric carcinoids as well as lung carcinoids have been found to contain and secrete histamine, which might be responsible for the characteristic bright red flush seen in these patients. Metabolites of histamine are frequently present in high concentration in the urine from these patients. Dopamine and norepinephrine have also been found in carcinoid tumors.

The occurrence of substance-P in carcinoid tumors was first demonstrated by Håkansson and co-workers in 1977. Substance-P belongs to a family of polypeptides that share the same carboxyl terminus and are called tachykinins (Fig. 41.4). A number of tachykinin-related peptides have been isolated from carcinoid tumors, such as neuropeptide-A, neuropeptide-K, and eledin. During stimulation of flush in patients with midgut carcinoids, multiple forms of tachykinins are released to the circulation (Fig. 41.5).

Many different polypeptides (e.g., insulin, gastrin, somatostatin, S-100 protein, polypeptide YY, pancreatic polypeptide, human chorionic gonadotropin alpha subunit [HCGα], motilin, calcitonin, vasointestinal polypeptide [VIP], and endorphins) have been demonstrated in carcinoid tumors by immunohistochemical staining and sometimes in tumor extracts. Ectopic ACTH or CRH production may be found in foregut carcinoids; in particular, patients with bronchial carcinoids seem susceptible to Cushing’s syndrome. Patients with carcinoid tumors of the foregut type might also present with acneormega due to ectopic secretion of growth hormone-releasing hormone from the tumor. Duodenal carcinoids as part of von Recklinghausen’s disease can secrete somatostatin.

The chromogranin/secretogranin family consists of CgA, CgB (sometimes called secretoglobin I), secretoglobin II (sometimes called CgC), and some other members. CgA was first isolated in 1965 as a water-soluble protein present in chromaffin cells from bovine adrenal medulla. Its immunoreactivity has been found in all parts of the gastrointestinal tract and pancreas and has also been isolated from all endocrine glands.

<table>
<thead>
<tr>
<th>TABLE 41-2 — Clinicopathologic Classification of Intestinal Endocrine Tumors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Well-differentiated endocrine tumor</td>
</tr>
<tr>
<td>Serotonin-producing tumor</td>
</tr>
<tr>
<td>Enteroglucagon-producing tumor</td>
</tr>
<tr>
<td>Uncertain behavior: functioning or nonfunctioning, confined to mucosasubmucosa, or nonangioinvasive</td>
</tr>
<tr>
<td>&gt;1 or 2 cm in diameter</td>
</tr>
<tr>
<td>Serotonin-producing tumor</td>
</tr>
<tr>
<td>Enteroglucagon-producing tumor</td>
</tr>
<tr>
<td>Low-grade malignant endocrine carcinoma</td>
</tr>
<tr>
<td>Serotonin-producing carcinoma or without carcinoid syndrome</td>
</tr>
<tr>
<td>Poorly differentiated endocrine carcinoma</td>
</tr>
<tr>
<td>High-grade malignant: small to intermediate cell carcinoma</td>
</tr>
</tbody>
</table>


CgA is an acidic glycoprotein of 439 amino acids with a molecular weight of 48 kD. It can be spliced into smaller fragments at dibasic cleavage sites, generating multiple bioactive fragments such as vasostatins, chromostatin, and pancreastatin. Both amines and peptides are co-released (Fig. 41.6).

Amines and hormones are stored intracellularly in two types of vesicles: (1) large dense core vesicles and (2) small synaptic-like vesicles. These vesicles are released on stimulation. Large dense-core vesicles contain the hormones and one or more members of the chromogranin/secretoglobin family of proteins. Both amines and peptides are co-released (Fig. 41.7).

The physiologic function of CgA is not fully elucidated. Its ubiquitous presence in neuroendocrine tissues and its co-secretion with peptide hormones and amines indicate a storage role of the peptide within the secretory granule. It also acts as a prohormone that can generate bioactive smaller fragments. CgA is an important tissue and serum marker for different types of carcinoid tumors, including those of the foregut, midgut, and hindgut (see Table 41-1 and later discussion).
CLINICAL PRESENTATION

The clinical presentation of carcinoid tumors depends on localization, hormone production, and extent of the disease. Usually, a lung carcinoid is diagnosed incidentally on routine pulmonary radiography, whereas a midgut carcinoid may be identified as a bowel obstruction or as a cause of abdominal discomfort or pain. Rectal carcinoids may cause bleeding or obstruction. However, lung carcinoids may also present clinically with Cushing's syndrome, due to secretion of CRH or ACTH, or with the carcinoid syndrome, due to production of serotonin, 5-HTP, or histamine. A midgut carcinoid often presents with the carcinoid syndrome, due to production of serotonin and tachykinins.

The clinical manifestations at referral depend on the type of referral center. At our institution, which cares for patients with malignant tumors, 74% of the patients present with the carcinoid syndrome, 13% with abdominal pain, 12% with carcinoid heart disease, and 2% with bronchial constriction. When unbiased material is analyzed, bowel obstruction is the most frequent problem leading to the diagnosis of ileal carcinoid tumor. The second most frequent symptom is abdominal pain. Flushing and diarrhea, which are components of the carcinoid syndrome, make up only the third most frequent presentations. Because many patients have vague symptoms, however, diagnosis of the tumor may be delayed by approximately 2 to 3 years.

The syndrome rarely occurs in patients with rectal carcinoids. This is the classic carcinoid syndrome, but some patients may display only one or two of the features. Other symptoms related to the syndrome are weight loss, sweating, and pellagra-like skin lesions.

In 1954, Thorson and co-workers for the first time described the carcinoid syndrome as having the following features: malignant carcinoid of the small intestine with metastasis to the liver; valvular disease of the right side of the heart (pulmonary stenosis and tricuspidal insufficiency without septal defect); peripheral vasomotor symptoms; bronchial constriction; and an unusual type of cyanosis. One year later, Dr. William Bean gave this colorful description of the carcinoid syndrome:

This witch's brew of unlikely signs and symptoms, intriguing to the most fastidious connoisseur of clinical esoteric, the skin underwent rapid and extreme changes resembling in clinical miniature the febrile phantasmagoria of the aurora borealis.

The syndrome is thus well characterized and includes flushing, diarrhea, right-sided heart failure, and sometimes bronchial constriction and increased urinary levels of 5-HIAA. This is the classic carcinoid syndrome, but some patients may display only one or two of the features. Other symptoms related to the syndrome are weight loss, sweating, and pellagra-like skin lesions.

Development of the carcinoid syndrome is a function of tumor mass, extent and localization of metastases, and localization of the primary tumor. The syndrome is most common in tumors originating in the small intestine and proximal colon; 40% to 60% of patients with these tumors experience the syndrome. The disorders are less frequent in patients with bronchial carcinoids and do not occur in patients with rectal carcinoids. The syndrome rarely occurs in patients with midgut carcinoids and a small tumor burden, such as only regional lymph node metastases. Patients with the full syndrome usually have multiple liver metastases. The association with hepatic metastases is due to inefficient inactivation by the liver of amines and peptides released into the portal circulation. The venous drainage of liver metastases is directly into the systemic circulation and bypasses hepatic inactivation.

Other carcinoid tumors likely to be associated with the carcinoid syndrome in the absence of liver metastases are ovarian carcinoids and bronchial carcinoids, which release mediators directly into the systemic rather than the portal circulation. Retroperitoneal metastases from classic midgut carcinoid also release mediators directly into the circulation and might cause the carcinoid syndrome without any liver metastases.
Flushing

Four types of flushing have been described in the literature.  

The first and most well-known type is the sudden, diffuse, erythematous flush, usually affecting the face, neck, and upper chest (i.e., the normal flushing area) (Fig. 41-8A and B) (see also Color Plate). This type of flush is commonly of short duration, lasting from 1 to 5 minutes, and is related to early-stage midgut carcinoids. Patients usually experience a sensation of warmth during flushing and sometimes heart palpitations. This type of flushing is reported in 20% to 70% of patients with midgut carcinoid at presentation of the disease.  

The second type is the violaceous flush, which affects the same area of the body. It has roughly the same time course or sometimes lasts a little longer. Patients may also have facial telangiectasia. This flush is related to the later stages of midgut carcinoid (Fig. 41-9) (see also Color Plate) and is normally not felt by patients because they have become accustomed to the flushing reaction.  

The third type is prolonged flushing that usually lasts a couple of hours but may last up to several days. This flush sometimes involves the whole body and is associated with profuse lacrimation, swelling of the salivary gland, hypotension, and facial edema (Fig. 41-10) (see also Color Plate). These symptoms are usually associated with malignant bronchial carcinoids.  

The fourth type of flushing is a bright red, patchy flush, seen in patients with chronic atrophic gastritis and ECL-cell hyperplasia, or so-called ECL-oma (derived from ECL cells). This type of flushing is related to an increased release of histamine and histamine metabolites.

Flushes may be spontaneous or may be precipitated by (1) stress (physical and mental); (2) infection; (3) alcohol; (4) certain foods (spicy); or (5) drugs, such as by injections of cate-cholamines, calcium, or pentagastrin (see later). The pathophysiology of flushing in the carcinoid syndrome is not yet elucidated. It was previously thought to be totally related to excess production of serotonin or serotonin metabolites. However, several patients with high levels of plasma serotonin did not have any flushing, nor did a serotonin antagonist (e.g., methysergide, cyproheptadine, or ketanserin) have any effect on the flushing.  

In a study from our own group in which we measured the release of tachykinins, neuropeptide-K, and substance-P during flushing provoked by pentagastrin or alcohol, a clear correlation was found between the onset and intensity of the flushing reaction and the release of tachykinins (see Fig. 41-5). Furthermore, when the release of tachykinins was blocked by pretimulatory administration of octreotide, little or no flushing was observed in the same patient (Fig. 41-11). Other mediators of the flushing reaction may be kallikrein and bradykinins, which are released during provoked flushing. Histamine may be a mediator of the flushes seen both in lung carcinoids and in gastric carcinoids (ECL-omas). Tachykinins, bradykinins, and histamines are well-known vasodilators, and somatostatin analogues may alleviate flushing by reducing circulating levels of these agents (see later). Furschgott and Zawadski have suggested that flushing is caused by an indirect vasodilatation mediated by endothelium-derived relaxing factor (EDRF) or by nitric oxide released by 5-HTP during platelet activation.  

The facial flushing associated with carcinoid tumors should be distinguished from idiopathic flushing and menopausal "hot flashes." Patients with idiopathic flushes usually have a long history of flushing starting rather early in life and sometimes with a family history without occurrence of a tumor. Menopausal hot flashes usually involve the whole body and are accompanied by intense sweating. Postmenopausal women in whom a true carcinoid syndrome is developing can tell the difference between the two types of flushes.
Diarrhea

Diarrhea occurs in 30% to 80% of patients with the carcinoid syndrome. Its pathophysiology is poorly understood but is probably multifactorial. The diarrhea is frequently accompanied by abdominal cramping, and endocrine, paracrine, and mechanical factors contribute to this condition. A variety of tumor products, including serotonin, tachykinins, histamines, kallikrein, and prostaglandins, can stimulate peristalsis, electromechanical activity, and tone in the intestine.

Secretory diarrhea may occur with fluid and electrolyte imbalance. Malabsorption may result from intestinal resections, lymphangiectasia, secondary to mesenteric fibrosis, bacterial overgrowth, and secondary to a tumor partially obstructing the small bowel or rapid intestinal transit. Increased secretion by the small bowel, malabsorption, or accelerated transit may overwhelm the normal storage and absorptive capacity of proximal colon and result in diarrhea, which may be aggravated if the reabsorbed function of the colon is impaired.

In a study of patients with elevated serotonin levels and the carcinoid syndrome, transit time in the small bowel and colon was significantly decreased in comparison with that of normal subjects. The volume of the ascending colon was significantly smaller than in normal subjects, and the postprandial colonic tone was markedly increased. This indicates that in patients in whom the carcinoid syndrome is associated with diarrhea, major alterations in gut motor function occur that affect both the small intestine and the colon. Many patients with carcinoid tumors have undergone wide resection of the small intestine, and they may be affected by the symptoms of short-bowel syndrome.

Serotonin is thought to be responsible for the diarrhea in the carcinoid syndrome by its effects on gut motility and intestinal electrolyte and fluid secretion. Serotonin receptor antagonists, such as ondansetron and ketanserin, relieve the diarrhea to a certain degree.
Carcinoid Heart Disease

A unique endocrine effect of carcinoid tumors is the development of plaque-like thickenings of the endocardium of the heart, valve leaflets, atria, and ventricles in 10% to 20% of the patients. This fibrotic involvement causes stenosis and regurgitation of the blood flow. Findings of new collagen beneath the endothelium of the endocardium is almost pathognomonic for carcinoid heart disease. The incidence of these lesions depends on the diagnostic methodology. Echocardiography can demonstrate early lesions in about 70% of patients with the carcinoid syndrome, whereas routine clinical examinations detect them in only 30% to 40%. These figures have significantly dropped to 10% to 15%, probably because of earlier diagnosis and the use of biologic antitumor treatments such as somatostatin analogues and -interferons. Both of these agents control the hormonal release and excess that might be involved in the fibrotic process. In a study performed 15 years ago, 40% of patients with carcinoid tumors died of cardiac complications related to the carcinoid disease. More recent data reveal that this complication is a rare event, and patients usually die of the effects of a progressive tumor.

The precise mechanism behind the fibrosis in the right side of the heart has not been solved at the moment, but it occurs mainly in patients with liver metastases who usually also have the carcinoid syndrome. Substances inducing fibrosis are thought to be released directly into the right side of the heart and are then neutralized or degraded through the lung circulation because few patients present with similar lesions on the left side. However, patients with lung carcinoids occasionally display the same fibrotic changes on the left side. Histologically, the plaque-like thickenings in the endocardium consist of myofibroblasts and fibroblasts embedded in a stroma that is rich in mucopolysaccharides and collagen.

We have previously shown that the transforming growth factor (TGF-) family of growth factors is up-regulated in carcinoid fibrous plaques on the right side of the heart. The TGF- family of growth factors is known to stimulate matrix formation and collagen deposition. The substances that induce TGF- locally in the heart are not known, but serotonin, tachykinins, and insulin-like growth factor I (IGF-I) may be mediators.

A correlation has been found between circulating levels of serotonin and tachykinins and the degree and frequency of carcinoid heart lesion. The weight-reducing drugs fenfluramine and dexfenfluramine appear to interfere with normal serotonin metabolism and have been associated with valvular lesions identical to those seen in carcinoid heart disease. However, treatment resulting in decreased urinary 5-HIAA excretion does not result in regression of cardiac lesions.

Another possible mediator might be IGF-I, which is released from carcinoid tumor cells. Treatment with somatostatin analogues, which down-regulate circulating IGF-I, has been able to prevent further development of carcinoid heart disease in two patients (data to be published).
Bronchial Constriction

A true asthma episode is a rare event in patients with the carcinoid syndrome. The causative agents of bronchial constriction are not known, but both tachykinins and bradykinins have been suggested as mediators. These agents can constrict smooth muscles in the respiratory tract and may also cause local edema in the airways.
Other Manifestations of the Carcinoid Syndrome

Fibrotic complications other than heart lesions may be found in patients with carcinoid tumors. These include (1) intra-abdominal and retroperitoneal fibrosis, (2) occlusion of the mesenteric arteries and veins, (3) Peyronie's disease of the penis, and (4) carcinoid arthropathy. Intra-abdominal fibrosis can lead to intestinal adhesions and bowel obstruction and is a more common cause of bowel obstruction than is the primary carcinoid tumor itself. Retroperitoneal fibrosis can result in urethral obstruction that impairs kidney function, which sometimes requires treatment with urethral stents. Narrowing and occlusion of arteries and veins by fibrosis are potentially life-threatening. Ischemic loops of the bowel may have to be removed, and this procedure ultimately causes short-bowel syndrome. Other rare features of the syndrome are pellagra-like skin lesions with hyperkeratosis and pigmentation, myopathy, and sexual dysfunction.
Carcinoid Crisis

Carcinoid crisis has become a rare event since the introduction of treatment with somatostatin analogues. It might occur spontaneously or during induction of anesthesia, embolization procedures, chemotherapy, or infection. Carcinoid crisis is a clinical condition characterized by severe flushing, diarrhea, hypotension, hyperthermia, and tachycardia. Without treatment, patients might die during the crisis.

Intravenous (IV) and/or subcutaneous somatostatin analogues are given before, during, and after surgery to prevent the development of carcinoid crisis. Patients with metastatic lung carcinoids are particularly difficult to treat during crisis. IV infusions of octreotide at doses of 50 to 100 µg/hour, supplemented with histamine H₁-receptor and H₂-receptor blockers and IV sodium chloride, are recommended.
Other Clinical Manifestations of Carcinoid Tumors

Ectopic secretion of CRH and ACTH from pulmonary carcinoid tumors and thymic carcinoids accounts for 1% of all cases of Cushing's syndrome. Acromegaly due to ectopic secretion of GHRH has also been reported in foregut carcinoids. Gastric carcinoid tumors make up less than 1% of gastric neoplasms. They can be separated into three distinct groups on the basis of both clinical and histologic characteristics and originate from gastric ECL cells:

1. Those associated with chronic atrophic gastritis type A (80%) (type I).
2. Those associated with Zollinger-Ellison syndrome as part of MEN-1 syndrome (6%) (type II).
3. Sporadic gastric carcinoids (type III), which occur without hypergastrinemia and pursue a more malignant behavior, with 50% to 60% developing metastases.

About 80% of gastric carcinoids are associated with chronic atrophic gastritis type A, and more than 50% of patients with these carcinoids also have pernicious anemia. These tumors are more common in women than in men and are usually identified endoscopically during diagnostic evaluation for anemia or abdominal pain. They are often multifocal and localized in the gastric fundus area, and they are derived from ECL cells. Patients have hypochlorhydria and hypergastrinemia. Gastrin hypersecretion has been postulated to result in hyperplasia of the ECL cells, which might later develop into carcinoid tumors. Hyperplasia of ECL cells has been noticed in patients with long-standing proton-pump inhibitor therapy.
DIAGNOSIS

The diagnosis of a suspected carcinoid tumor must take into consideration tumor biology, histopathology, biochemistry, and localization. The diagnosis of a carcinoid may be suspected from clinical symptoms suggestive of the carcinoid syndrome or from the presence of other clinical symptoms, or it can be made in relatively asymptomatic patients from the histopathology at surgery or after liver biopsy for unknown hepatic lesions.

In one study involving 154 consecutive patients with gastrointestinal carcinoids found at surgery, 60% were asymptomatic. In patients with symptomatic tumors, the time from onset of symptoms until diagnosis is frequently delayed, varying from 1 to 2 years. The current tumor biology program includes growth factors (platelet-derived growth factor, epidermal growth factor, IGF-I, TGF-), and proliferation factors (measurements of the nuclear antigen Ki-67) as a proliferation index. Such index correlates with tumor aggressiveness and survival. Adhesion molecules such as CD-44, particularly exon-V6 and exon-V9, have been related to improved survival. Determination of the expression of angiogenic factors basic fibroblast growth factor (b-FGF) and vascular endothelial growth factor (VEGF) should also be included in a tumor biology program. Somatostatin analogues are cornerstones in the treatment of the carcinoid syndrome; therefore, determination of the different subtypes of somatostatin receptors (sst-1 to sst-5) with specific antibodies is warranted.

The histopathologic diagnosis of carcinoids is based on immunohistochemistry using antibodies against CgA, synaptophysin, and neuron-specific enolase. These immunohistochemical stainings have replaced the old silver stainings by Grimelius and Sevier-Munger. The argentaffin staining by Masson to demonstrate content of serotonin has also been replaced by immunocytochemistry with serotonin antibodies. These neuroendocrine markers can be supplemented by specific immunocytochemistry to different hormones such as substance-P, gastrin, and ACTH.

Biochemical Diagnosis

In patients with flushing and other manifestations of the carcinoid syndrome, the diagnosis can be established by measuring the urinary excretion of 5-HIAA because levels are invariably elevated under these circumstances. Patients with carcinoid tumors usually have urinary 5-HIAA levels of 100 to 3000 µmol/24 hours (15 to 60 mg/24 hours) (reference range <50 µmol/24 hours [10 mg/24 hours]). Assays for urinary 5-HIAA include high-pressure liquid chromatography (HPLC) with electrochemical detection and colorimetric and fluorescence methods. Various foods and drugs can interfere with the measurement of urinary 5-HIAA, and patients should avoid these agents during the 24-hour sampling. Normally, two 24-hour urine collections are recommended. In a study of patients with malignant midgut carcinoid tumors, 60% to 73% presented with increased urinary 5-HIAA levels, with a specificity of almost 100%.

Today measurement of urinary 5-HIAA for diagnosis of carcinoid tumor is the predominant biochemical analytic procedure. However, urinary and platelet measurement of serotonin itself may give additional information. In some studies, platelet serotonin levels were more sensitive than urinary 5-HIAA and urinary serotonin levels and were not affected by the patient’s diet, as are 5-HIAA levels.

In a comparative study of 44 consecutive patients with carcinoid, the platelet serotonin, urinary 5-HIAA, and urinary serotonin levels were measured. In foregut carcinoids the sensitivities were 50%, 29%, and 55%, respectively. For midgut carcinoids, the sensitivities were 100%, 92%, and 82%, respectively, and for hindgut carcinoids, 20%, 0%, and 60%, respectively.

Elevations of 5-HIAA can occur in malabsorption states and a number of other conditions. Foregut carcinoids tend to produce an atypical carcinoid syndrome with increased plasma 5-HTP, but not serotonin, because they lack the appropriate decarboxylase. That results in normal urinary 5-HIAA. However,

| TABLE 41-3 -- Factors That Interfere with Determination of Urinary 5-HIAA |
|--------------------------|-----------------------------------------------|
| **Foods**                |                                               |
| Avocado                  |                                               |
| Banana                   |                                               |
| Chocolate                |                                               |
| Coffee                   |                                               |
| Eggplant                 |                                               |
| Pecan                    |                                               |
| Pineapple                |                                               |
| Plum                     |                                               |
| Tea                      |                                               |
| Walnuts                  |                                               |
| **Drugs**                |                                               |
| Acetaminophen            |                                               |
| Acelanilid               |                                               |
| Caffeine                 |                                               |
| Fluourouracil            |                                               |
| Guafenesin               |                                               |
| l-Dopa                   |                                               |
| Melphalan                |                                               |
| Mephenesin               |                                               |
| Methamphetamine         |                                               |
| Methocarbamol            |                                               |
| Methysgeride maleate     |                                               |
| Phenmetrazine            |                                               |
| Reserpine                |                                               |
| Salicylates              |                                               |
Factors That Cause False-Negative Results

Drugs

Corticotropin

β-Chlorophenylalanine

Chlorpromazine

Isopropylamine

Isoniazid

Methenamine mandelate

Methyldopa

Monoamine oxidase inhibitors

Phenothiazine

Promethazine

Figure 41-12 Plasma levels of chromogranin-A (CgA), CgB, and CgC in patients with various neuroendocrine tumors. EPTs, endocrine pancreatic tumors; MEN-1, multiple endocrine neoplasia type 1.

some of the 5-HTP is decarboxylated in the intestine and other tissues, and many of these patients have slightly elevated U-5-HT or 5-HIAA levels.

Attempts have been made to identify more specific and sensitive serum markers for carcinoid that may allow earlier diagnosis. One such marker is CgA. It has been shown previously that CgA and CgB are more abundant than CgC in human neuroendocrine tissues. In 44 patients with carcinoid tumors, CgA was increased in 99%, CgB in 88%, and CgC in only 6%. In a study of 75 patients with midgut carcinoids and the carcinoid syndrome, CgA was elevated in 87% of carcinoid patients. Furthermore, a correlation between levels of plasma chromogranin and extent of disease was found (P < .0001). In the same study, urinary 5-HIAA was elevated in 76% of midgut carcinoids, and there were no correlations with tumor size or extent of disease.

CgA is a more sensitive marker than urinary 5-HIAA in detection of carcinoid tumors, but because CgA is released and secreted from various types of neuroendocrine tumors, the specificity is lower. Therefore, in a work-up of patients with the carcinoid syndrome, one should combine the determination of plasma CgA with urinary 5-HIAA or serotonin. Plasma neuron-specific enolase shows a lower sensitivity and specificity than does plasma CgA. Serum hCG- has been reported to be increased in 60% of patients with foregut carcinoid tumors and 50% of hindgut carcinoids but in only 11% of those patients with midgut carcinoids and the carcinoid syndrome. Plasma neuropeptide-K levels have been reported to be elevated in 46% of patients with midgut carcinoids, whereas only 9% of patients with foregut carcinoids displayed elevated levels. Plasma substance-P has a sensitivity of 32% and a specificity of 85%. Pancreatic polypeptide levels are also elevated in about one third of patients with midgut carcinoids and in as many with foregut carcinoids.

During therapy with somatostatin analogues, neither plasma CgA nor urinary 5-HIAA is a reliable marker of tumor size because somatostatin inhibits the synthesis and release of the hormones without changes in tumor size.
Localization Procedures

Numerous imaging techniques, including endoscopy, barium enema, chest radiography, ultrasonography, computed tomography, magnetic resonance imaging (MRI), and angiography, have been used to determine the location of the primary tumor as well as the metastases in patients with carcinoid tumors. In more recent years, somatostatin-receptor scintigraphy (SRS) and iodinated meta-iodobenzylguanidine (131I-MIBG) scanning have been used to localize and stage the disease. Bronchial carcinoids are usually detected by chest radiography, CT, or, occasionally, by bronchoscopy. The primary midgut tumor is usually small and difficult to localize with traditional diagnostic methods such as barium enema, CT scan, or MRI. Some of these tumors can be localized by angiography or SRS. Liver metastases are usually detected by CT or MRI. At present, CT or MRI and SRS are the primary diagnostic modalities for tumor staging (Fig. 41-13 A and B) (see also Color Plate).

A more sensitive method is positron emission tomography (PET) using 11C-5-HTP, the precursor of serotonin synthesis (Fig. 41-14) (see also Color Plate). This isotope is accumulated in carcinoid tumors, and with the recent development of PET cameras, tumors as small as 0.3 cm in diameter can be detected. During treatment a close relationship has been found among changes in the PET scan, transport rate constant, and urinary 5-HIAA, suggesting that PET scanning may be useful in monitoring the results of therapy. PET scanning using fluorodeoxyglucose 18 (18FDG) is not useful in detecting low-proliferating neuroendocrine tumors but can be beneficial in identifying poorly differentiated anaplastic tumors.

Carcinoid tumors contain high-affinity receptors for somatostatin in 80% to 100% of cases. The receptors are present in both the primary tumor and metastases. Five subtypes of somatostatin receptors have been cloned (sst-1 to sst-5), and somatostatin receptor type 2 is the predominant subtype expressed in carcinoid tumors.

The most commonly available somatostatin analogue, octreotide, binds with high affinity to sst-2 and with lower affinity to sst-3 and sst-5. SRS with 111In-DTPA-Phe-octreotide has been reported with a sensitivity of 80% to 90% in patients with carcinoids. Many studies have demonstrated that SRS has greater sensitivity for localizing carcinoids compared with conventional imaging studies. False-positive scans can be encountered in patients with granulomas (e.g., sarcoidosis, tuberculosis), activated lymphocytes (lymphomas, chronic infection), thyroid diseases (goiter, thyroiditis), endocrine pancreatic tumors, and other endocrine tumors. Because of its high sensitivity and ability to image, whole-body SRS should be the initial imaging procedure to localize and establish the stage of the disease. Bone metastases, which are common in carcinoid tumors, are efficiently picked up by SRS, which is as sensitive as traditional bone scanning with technetium.

A diagnostic algorithm is outlined in Figure 41-15.
TREATMENT

Treatment of carcinoid tumors with the carcinoid syndrome requires a multimodal approach, including symptomatic control as well as tumor reduction. Most patients with the carcinoid syndrome have metastatic disease. The therapeutic goals are to ameliorate and improve clinical symptoms, abrogate the tumor growth, improve quality of life, and if possible, prolong overall survival.

Symptomatic control of the carcinoid syndrome includes lifestyle changes, dietary supplementations, and specific medical treatment that reduces the clinical symptoms related to the different components of the carcinoid syndrome. Avoiding stress, both psychological and physical, as well as substances such as alcohol, spicy food, and medication that precipitate a flushing reaction might be sufficient in early cases. Production of serotonin by the tumor consumes tryptophan. Normally, about 1% of the tryptophan is used for the synthesis of serotonin, which can result in tryptophan and niacin deficiency. Therefore, supplemental niacin to prevent the development of pellagra has been recommended over the years. Because many patients have undergone resection of the terminal ileum, which may result in vitamin B₁₂ and folic acid deficiency, vitamin B₁₂ supplementation is needed in those patients.

Heart failure due to carcinoid heart disease may require diuretics or angiotensin-converting enzyme (ACE) inhibitors. A small number of patients need bronchodilators such as salbutamol, which interacts with -adrenergic receptors and does not induce flushing. The diarrhea seen in the carcinoid syndrome might be controlled by loperamide or diphenoxylate. If patients still have the carcinoid syndrome, they receive somatostatin analogue treatment, which has replaced most of the earlier types of serotonin and serotonin receptor inhibitors. Serotonin inhibitors (e.g., parachlorophenylalanine and -methyldopa), which inhibit serotonin synthesis, and serotonin receptor antagonists (e.g., ciproheptadine, methysergide, and ketanserin) are not used routinely clinically. These earlier treatments had limited efficacy in terms of inhibiting flushing and diarrhea and were accompanied by significant side effects. A combination of histamine H₁ and H₂-receptor antagonists is effective in the carcinoid syndrome that is caused by foregut carcinoids due to concomitant secretion of histamine and serotonin. Prednisolone in doses of 15 to 30 mg/day gives occasional relief in some cases with severe flushing and diarrhea.

Somatostatin Analogues

Although natural somatostatin-14 reduces symptoms in patients with the carcinoid syndrome, its use is limited by its short half-life (2.5 minutes). During the last two decades, synthetic somatostatin analogues (octapeptides) have been developed for clinical use. Octreotide is the most commonly available drug; other analogues are lanreotide and vapreotide.

The somatostatin analogues used in clinical practice (octreotide, lanreotide) both bind to receptors sst-1 and sst-5 and, with lower affinity, to sst-3. They exert their cellular action through interaction with specific cell and transmembrane receptors belonging to the superfamily of G protein-coupled membrane receptors. They inhibit adenylate cyclase activity, activate phosphotyrosine phosphatases (PTPs), and modulate mitogen-activated protein kinases (MAPKs). Receptor subtypes 2 and 5 modulate K⁺ and Ca²⁺ fluxes in the cell. Activation of all these pathways results in inhibition of known growth factor production and release as well as antiproliferative effects.

Somatostatin receptor subtype 3 is known to mediate PTP-dependent apoptosis accompanied by activation of p53 and Bax. Four of the five somatostatin receptor subtypes (sst-2 to sst-5) undergo rapid internalization after ligand binding, which has been explored by tumor-targeted radioactive somatostatin analogue therapy. An antiproliferative effect has been reported, probably through a combination of receptor subtype 2 and 5 activities, which inhibits MAPK and K⁺ and Ca²⁺ fluxes leading to cell cycle arrest; the precise antitumor mechanism, however, is not known.

It is now known that different subtypes of somatostatin receptors form heterodimers (sst-1 and sst-5) and heterodimers with dopamine receptor D2R. This cross-talk...
modulates the intracellular signal and gives a “fine tuning” of the mediated effects. \[148\]

All five subtypes of somatostatin receptors are expressed in carcinoid tumors; they are expressed in various combinations, although some tumors express all five subtypes.\[148\] \[149\] The receptors are expressed not only on tumor cells but also in periluminal veins. \[148\] Antiangiogenesis might be another antitumor mechanism of somatostatin analogues. \[152\]

Subcutaneous administration of octreotide and lanreotide every 8 to 12 hours can control the clinical symptoms in about 60% to 70% of patients with the carcinoid syndrome; these agents are considered the drugs of choice.\[153\] Octreotide and lanreotide decrease serotonin and urinary 5-HIAA levels as well as plasma tachykinin and CgA levels. The recommended dose for octreotide is 100 µg two or three times a day, a standard treatment for controlling clinical symptoms. \[154\] However, some patients may require higher doses, up to a total of 3000 µg/day, to control the clinical symptoms and tumor growth, particularly during long-term therapy.

Tachyphylaxis (reduced sensitivity) to somatostatin analogues may develop during long-term therapy. \[158\] Long-acting, slow-release formulations of octreotide and lanreotide have been developed, and doses of 20 to 30 mg given once a month (octreotide) or every 2 weeks (lanreotide) control clinical symptoms and hormone levels in 50% to 60% of patients with the carcinoid syndrome. \[158\] The long-acting formulations of somatostatin analogues have clearly improved the quality of life of patients by reducing the number of injections and provide more stable control of clinical symptoms. \[159\]

High-dose therapy with lanreotide (12 mg/day) and octreotide (3 mg/day) has generated an increased number of significant tumor reductions (12% versus 5% for the standard dose).\[164\] \[165\] Induction of apoptosis has been reported during high-dose therapy, \[166\] possibly mediated through activation of receptor subtype 3.

For patients at risk for carcinoid crisis, somatostatin analogue therapy is the treatment of choice. Carcinoid crisis is a life-threatening complication of the carcinoid syndrome and may occur spontaneously or may be associated with stress and anesthesia, chemotherapy, and infections (see earlier). Patients usually experience severe flushing, diarrhea, abdominal pain, and hypotension. Continuous infusion with somatostatin analogues, 50 to 100 µg/hour, is recommended and usually alters the life-threatening condition. It is also recommended that patients be given subcutaneous somatostatin analogues before surgery or before other stressful situations.

Side effects of somatostatin analogue therapy have been of a low-grade variety, occurring in 20% to 40% of patients. These include pain at the injection site, gas formation, diarrhea, and abdominal cramping. Significant long-term side effects include gallstone formation, “sludge” in the gallbladder, steatorrhea, deterioration of glucose tolerance, and hypocalcemia.\[168\] \[169\] The incidence of gallstones in patients treated over the long term has varied from 5% to 70%, and the incidence of symptomatic gallstones requiring surgical treatment is less than 10%. \[168\]
Interferons

Interferon alone or in combination with somatostatin analogues effective in the treatment of the carcinoid syndrome. Symptomatic and biochemical control may be obtained in 40% to 50% of the patients with the recommended doses of 3 to 5 million units of recombinant interferon alfa-2a or alfa-2b three to five times per week subcutaneously. Significant tumor reduction is reported in 10% to 20% of the patients. Interferon exerts a direct effect on the tumor cells by blocking cell division in the G1/S-phase, by inhibiting protein and hormone synthesis, and by reducing angiogenesis through inhibition of angiogenic factors b-FGF and VEGF. It has also an indirect effect through stimulation of the immune system, particularly T cells and natural killer cells. Response to interferon can be predicted by analyzing induction of 2’5’-oligoadenylate synthetase or P68 (PKR) protein kinase, enzymes involved in cell cycle regulation and protein synthesis.

Treatment with interferon induces an intratumoral fibrosis that is not picked up by regular CT scanning or ultrasonography; therefore, tumor size may remain unchanged. The side effects of interferons are more pronounced than with somatostatin analogues and include chronic fatigue syndrome, anemia, leukopenia, and thrombocytopenia as well as the development of autoimmune reactions in 10% to 15% of the patients. Most of the side effects are dose-dependent and can be managed by individualizing the dose.

Patients with the carcinoid syndrome who have not responded to octreotide or interferon alone may be given a combination of both agents. Such combinations have generated symptomatic control in 70% of patients and stabilization of tumor growth in 40% to 50% of patients. The combination also offers better tolerance of interferons when somatostatin analogues are added. Moreover, somatostatin analogue treatment is hampered by development of tachyphylaxis with time, which means less sensitivity to the somatostatin analogue, necessitating escalating doses and, finally, withdrawal of the compound for several months, when interferon therapy can continue. Conversely, interferon can be withdrawn and somatostatin analogue can be continued if severe side effects to interferon (mainly chronic fatigue syndrome or mental depression) develop.
Chemotherapy

Most agree that patients with classic midgut carcinoids and the carcinoid syndrome, in which tumors show low proliferation capacity, should not receive chemotherapy. The results in various studies have been disappointing, with response rates not more than 5% to 10%, short-lasting, and accompanied by considerable side effects. The combination of streptozotocin and 5-fluorouracil, which has demonstrated antitumor effect in endocrine pancreatic tumors, has not shown similar effects in classic midgut carcinoids. In foregut carcinoids, which usually present a more malignant behavior, however, cytotoxic treatment may be attempted. Such combinations include streptozotocin plus 5-fluorouracil, doxorubicin, cisplatin plus etoposide, and dacarbazine plus 5-fluorouracil. All of these cytotoxic treatments can be combined with a somatostatin analogue.
Surgery

Because most patients with the carcinoid syndrome already have malignancy at clinical presentation, surgical cure is seldom obtained. Resection of local disease or regional nodular metastatic disease can cure some patients; however, even if radical surgery cannot be performed, debulking procedures and bypass should always be considered and can be performed at any time during the course of treatment. In recent years, a more proactive attitude among surgeons has emerged, and wider resections and debulking procedures are being performed today than 10 years ago. In contrast to other metastatic tumors to the liver, in which liver transplantation has generally given poor results, an interest in liver transplantation is increasing for patients with metastatic carcinoids. In a review of 103 patients with malignant neuroendocrine tumors, including both carcinoids and pancreatic endocrine tumors, 5- and 2-year survival rates were 16% and 47%, respectively; however, recurrence-free survival was less than 24%. Liver transplantation might be considered in younger patients (<50 years of age) with a life-threatening uncontrolled carcinoid syndrome during medical therapy or tumor-targeted radioactive treatment without known metastatic spread outside the liver.

Another means of tumor reduction is hepatic artery embolization, which not only improves the carcinoid syndrome in about 50% of the patients but also reduces the tumor size in as many. The therapeutic effect may last for 9 to 12 months, and the procedure can be repeated. Chemoembolization, simultaneous embolization with surgical gel (Gelfoam), and chemotherapy (doxorubicin, mitomycin C, cisplatin, 5-fluorouracil), or interferon has resulted in symptomatic improvement in a significant number of patients with the carcinoid syndrome. However, hepatic artery occlusion or embolization may result in serious side effects (nausea, vomiting, liver pain, fever) and major complications (hepatorenal syndrome, sepsis, gallbladder perforation, and intestinal necrosis). Complications are seen in 5% to 7% of patients.

Other cytoreductive treatments may include cryotherapy and radiofrequency ablation. However, this procedure is limited to patients with smaller tumor burden, tumors less than 3 cm in diameter, and a limited number of metastases.
Irradiation

External irradiation has demonstrated limited efficacy and is used mainly to palliate symptoms related to bone and brain metastases. MIBG is taken up by carcinoids and is concentrated. The possibility of radiolabeled MIBG therapy has been evaluated in a limited number of patients. The response rate has been reported to be about 30% with 125I-MIBG or 131I-MIBG.

Somatostatin analogue-based tumor-targeted radioactive treatment has been applied through the last few years using 111In-DTPA-octreotide. Symptomatic improvement is reported in about 40% of the patients and tumor stabilization in about 30%. Indium 111 (111In) is a weak irradiator (Auger electrons) and seems to be replaced by yttrium 90 (90Y) (β emitters).

Studies with 90Y-DOTA-octreotide have been reported with promising results. It is an attractive mode of treatment because the radioactive ligand, after binding to the receptor, is internalized and transported to the cell nucleus, causing DNA damage. Because tumor cells usually have higher-density somatostatin receptors (sst-2 and sst-5) than do surrounding normal tissues, the treatment might be better tolerated.
PROGNOSIS

Clinically, the carcinoid syndrome is generally a manifestation of advanced disease. Carcinoids from various sites differ not only in the percentage developing the carcinoid syndrome but also in their aggressiveness. Survival rates for patients with various carcinoids depend on the site and the extent of the tumor. In patients with only localized disease, the 5-year survival rate is about 65%, not essentially higher than that for patients with regional metastases. In patients with distant metastases, the 5-year survival rate is reduced to 36%.

One of the main determinants of survival in carcinoid patients is the presence of metastases. Female gender and a younger age are associated with a better prognosis. Other factors that correlate with impaired survival are high CgA level at diagnosis and high proliferation index (Ki-67). During the 1990s, there was a reduced incidence of death from carcinoid heart disease, possibly a result of earlier diagnosis, active surgery, and the introduction of somatostatin analogues and interferons. In an earlier study performed by our group, 30% of the patients died of carcinoid heart complications. In a more recent study, this rate was reduced to less than 10%. Clinically significant carcinoid heart disease is now a rare event. Five percent to 10% of patients with carcinoids are at an increased risk for development of synchronous adenocarcinoma of the large intestine. The occurrence of a second malignancy is associated with a worse prognosis.
OTHER FLUSHING DISORDERS

Medullary Thyroid Carcinoma and VIPoma

Other neuroendocrine tumors, such as medullary thyroid carcinoma (MTC) and VIP-producing tumors (ganglioneuroma, endocrine pancreatic tumors), may present with flushing syndromes. Patients may also have diarrhea, particularly those with VIP-producing tumors, which are accompanied by a severe secretory diarrhea. In patients with MTC, flushing and diarrhea are infrequent symptoms and are seen mainly in patients with high circulating levels of calcitonin and CGRP.

The mechanism behind the flushing and diarrhea is unknown, but it has been postulated to be mediated through prostaglandins stimulated by calcitonin. The frequency of flushing and diarrhea is usually less than 5% in patients with advanced metastatic MTC. Treatment is directed against tumor growth and may consist of surgical resection, embolization of liver metastases, and cytotoxic treatment (doxorubicin-based combination therapies). Somatostatin analogue therapy may alleviate the diarrhea.

VIPoma or WDHA (watery diarrhea, hypokalemia, and achlorhydria) syndrome (Verner-Morrison syndrome) is associated with severe secretory diarrhea (up to 15 L/day), and some patients also display a continuous whole-body violaceous flushing and hypotension. The syndrome also includes achlorhydria, hypokalemia, and metabolic acidosis and is related to overproduction of VIP and a related peptidyl peptide histidine methionine. These patients have tumors in the pancreas, lung, or sympathetic ganglia.

The diagnosis is confirmed by measuring plasma VIP, usually exceeding 70 pmol/L. Treatment is directed against the tumor and hormone excess. Administration of somatostatin analogues by either subcutaneous or intravenous infusion in the worst cases can control clinical symptoms. Cytotoxic treatment with streptozotocin-based combinations, 5-fluorouracil, or doxorubicin is recommended for malignant cases.
Mastocytosis and Related Disorders

Mastocytosis as well as other systemic mast cell activations is clinically related to flushing disorders. Most patients with mastocytosis have an indolent course, but some forms of mastocytosis are aggressive. Symptoms are attributed primarily to paroxysmal mast cell activation. [219][220]

Most patients with mastocytosis have evidence of cutaneous involvement, the most common being multiple, small pigmented lesions that produce urticaria on stroking with a blunt object (Darier's sign); this condition is called urticaria pigmentosa. [221] Another form of cutaneous mastocytosis is a more telangiectatic form called telangiectasia macularis eruptiva perstans. Hepatomegaly and splenomegaly can be due to infiltration of mast cells, and hepatic fibrosis is also common. [222][223]

Bone involvement can be manifested by either osteoporosis or osteosclerosis. [224] Systemic mastocytosis can also involve the gastrointestinal tract with mucosal nodules in the ileum, stomach, and large bowel. [225]

Hematologic abnormalities are nonspecific with marked mast cell infiltration of the bone, anemia, leukocytosis, eosinophilia, sometimes lymphadenopathy, and eosinophilia. [226] In a subgroup of patients, the mastocytosis is secondary to primary hematologic disorders, usually myeloproliferative or myelodysplastic disease. [227][228]

Mast cell leukemia has been reported in rare cases. [229] Mast cell leukemia has been reported in rare cases. [229] The diagnosis is made by measurement of histamine and histamine metabolites in the urine. [230][231] Quantification of histamine metabolites (N-methylhistamine and methylimidazole-acetic acid) appears to be more sensitive for overproduction of histamine in patients with mastocytosis. [231] Measurement of endogenous production of prostaglandin D$_2$ can be assessed by quantifying the major urinary metabolite (9-hydroxy-11,15-dioxo-2,3,18,19-tetranorprost-5-ene-1,20-dioic acid). [232] However, these measurements can be done only in specialized laboratories. Measurement of the tryptase release is easier to perform, and increased quantities of this granula-associated enzyme tryptase can be detected by immunoassay. [233]

Clinical signs of systemic mastocytosis include flushing, tachycardia, hypotension, and sometimes nausea, vomiting, and diarrhea. This syndrome resembles the carcinoid syndrome. Histamine is a potent vasodilator and is released from mast cells. Other mediators of the syndrome are the release of prostaglandin D$_2$, tryptase, and heparin. [234] Prostaglandin D$_2$ is a more potent mediator than is histamine.

The diagnosis is made by measurement of histamine and histamine metabolites in the urine. [235][236] Treatment depends on the severity of the disease. As in the treatment of allergic anaphylaxis, epinephrine is effective in reversing the hypotension associated with mast cell mediator release; thus, these patients should have constant access to epinephrine in the form of subcutaneous injection or inhalation. Chronic therapy to prevent acute attacks includes antihistamine therapy combined with inhibition of prostaglandin biosynthesis. Blockade of both histamine H$_1$ and H$_2$ receptors is required to prevent the vasodilator effect of histamine. [237]

Nonsteroidal anti-inflammatory drugs (NSAIDs) inhibit the cyclooxygenase enzyme that catalyzes the formation of prostaglandins. Aspirin has been used, but some patients cannot tolerate it because of side effects in the gut and allergic reactions. [238] In patients resistant to both antihistamine and NSAIDs, interferon has been attempted with a reduction in mast cell numbers as well as excretion of mast cell mediators. Treatment with interferon is still considered experimental. [239]
References


Robiolo PA, Rigolin VH, Wilson JS, et al. Carcinoid heart disease: correlation of high serotonin levels with valvular abnormalities deleted by cardiac catheterization and echocardiography.


<table>
<thead>
<tr>
<th>Page</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>114.</td>
<td>USA 17:27.</td>
</tr>
</tbody>
</table>


References


1876


APPENDIX: Reference Values

The reference values are meant to be used only with this text because values and ranges vary among laboratories. In preparing the book and the reference values, the editors have taken into account the fact that the system of international units (SI, Système International d'Unités) is used in most medical and scientific journals and in clinical laboratories in many countries. However, clinical laboratories in some countries, including the United States, report results in conventional units. Therefore, in the book and in the Reference Values section, we use both systems. In the text, values in SI units appear first, and conventional units appear in parentheses after the SI units.

For some assays that measure mixtures of products in serum (growth hormone, luteinizing hormone), both the SI and the conventional units are given as units of weight rather than as molarity. Most conversions from one system to the other can be made as follows:

Conversion of mEq/L to mmol/L is made by dividing mEq/L by the valence of the molecule. For the convenience of the reader, factors for converting conventional to SI units are included (see Young DS. Implementation of SI units for clinical laboratory data. Ann Intern Med 106:114128, 1987). For the Units of Radiation, 1 bequerel (Bq) = 2.7 × 10^{-11} curies (Ci), 37 mBq = 1 mCi, 1 gray (Gy) = 100 rad, 1 sievert (Sv) = 100 rem.

<table>
<thead>
<tr>
<th>Laboratory Parameter</th>
<th>SI</th>
<th>Conventional (C)</th>
<th>Conversion Factor (CF) CF × C = SI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetoacetate, plasma</td>
<td>&lt;100 µmol/L</td>
<td>&lt;1 mg/dL</td>
<td>97.95</td>
</tr>
<tr>
<td>Adrenal Steroids, plasma</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aldosterone, supine, saline suppression</td>
<td>&lt;240 pmol/L</td>
<td>&lt;8.5 ng/dL</td>
<td>27.74</td>
</tr>
<tr>
<td>Aldosterone, upright, normal diet</td>
<td>140560 pmol/L</td>
<td>520 ng/dL</td>
<td>27.74</td>
</tr>
<tr>
<td>Cortisol</td>
<td>8 AM</td>
<td>140690 nmol/L</td>
<td>525 µg/dL</td>
</tr>
<tr>
<td></td>
<td>4 PM</td>
<td>80330 nmol/L</td>
<td>312 µg/dL</td>
</tr>
<tr>
<td></td>
<td>Overnight dexamethasone suppression</td>
<td>&lt;140 nmol/L</td>
<td>&lt;5 µg/dL</td>
</tr>
<tr>
<td>Dehydroepiandrosterone (DHEA)</td>
<td>731 nmol/L</td>
<td>29 µg/dL</td>
<td>3.467</td>
</tr>
<tr>
<td>Dehydroepiandrosterone Sulfate (DHEAS)</td>
<td>1.36.8 µmol/L</td>
<td>5002500 ng/ml</td>
<td>0.002714</td>
</tr>
<tr>
<td>11-Deoxycortisol</td>
<td>&lt;30 nmol/L</td>
<td>&lt;1 µg/dL</td>
<td>28.86</td>
</tr>
<tr>
<td>17-Hydroxyprogesterone</td>
<td>Women, follicular phase</td>
<td>0.63 nmol/L</td>
<td>0.21 µg/L</td>
</tr>
<tr>
<td></td>
<td>Women, luteal phase</td>
<td>1.510.6 nmol/L</td>
<td>0.53.5 µg/L</td>
</tr>
<tr>
<td></td>
<td>Men</td>
<td>1.68 nmol/L</td>
<td>0.63 µg/L</td>
</tr>
<tr>
<td>Adrenal Steroids, urine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aldosterone</td>
<td>1453 nmol/d</td>
<td>519 µg/d</td>
<td>2.774</td>
</tr>
<tr>
<td>Cortisol, free</td>
<td>55276 nmol/d</td>
<td>20100 µg/d</td>
<td>2.759</td>
</tr>
<tr>
<td>17-Hydroxycorticosteroids</td>
<td>5.427.6 µmol/d</td>
<td>210 mg/d</td>
<td>2.759</td>
</tr>
<tr>
<td>17-Ketosteroids</td>
<td>Men</td>
<td>2588 µmol/d</td>
<td>725 mg/d</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>1453 µmol/d</td>
<td>416 mg/d</td>
</tr>
<tr>
<td>Ammonia, as NH₃, plasma</td>
<td>647 µmol/L</td>
<td>1080 µg/dL</td>
<td>0.5872</td>
</tr>
<tr>
<td>Angiotensin II, plasma</td>
<td>1060 ng/L</td>
<td>1060 pg/mL</td>
<td></td>
</tr>
<tr>
<td>Arginine Vasopressin (AVP), plasma</td>
<td>Random fluid intake</td>
<td>0.022.8 pmol/L</td>
<td>13 pg/ml</td>
</tr>
<tr>
<td></td>
<td>Dehydration 1824 h</td>
<td>5.513 pmol/L</td>
<td>414 pg/ml</td>
</tr>
<tr>
<td>Calciferols (as Cholecalciferol, Vitamin D₃) plasma</td>
<td>1,25-Dihydroxycholecalciferol [1,25(OH)₂D]</td>
<td>36144 pmol/L</td>
<td>1560 pg/mL</td>
</tr>
<tr>
<td></td>
<td>25-Hydroxycholecalciferol (25-OH-D)</td>
<td>20100 nmol/L</td>
<td>840 ng/mL</td>
</tr>
<tr>
<td>Calcitonin, plasma</td>
<td>Normal</td>
<td>&lt;19 ng/L</td>
<td>&lt;19 pg/mL</td>
</tr>
<tr>
<td></td>
<td>Medullary cancer</td>
<td>&gt;100 ng/L</td>
<td>&gt;100 pg/mL</td>
</tr>
<tr>
<td>Calcium</td>
<td>Ionized serum</td>
<td>11.4 mmol/L</td>
<td>45.6 mg/dL</td>
</tr>
<tr>
<td></td>
<td>Total serum</td>
<td>2.22.6 mmol/L</td>
<td>910.5 mg/dL</td>
</tr>
<tr>
<td>Catecholamines, urine</td>
<td>Free Catecholamines</td>
<td>&lt;580 nmol/d</td>
<td>&lt;100 µg/d</td>
</tr>
<tr>
<td></td>
<td>Epinephrine</td>
<td>&lt;275 nmol/d</td>
<td>&lt;50 µg/d</td>
</tr>
<tr>
<td></td>
<td>Metanephrines</td>
<td>&lt;7 µmol/d</td>
<td>&lt;1.3 ng/d</td>
</tr>
<tr>
<td></td>
<td>Norepinephrine</td>
<td>89473 nmol/d</td>
<td>1589 µg/d</td>
</tr>
<tr>
<td></td>
<td>Vanillylmandelic Acid (VMA)</td>
<td>&lt;40 µmol/d</td>
<td>&lt;8 mg/d</td>
</tr>
<tr>
<td>Chloride, serum</td>
<td></td>
<td>98106 µmol/L</td>
<td>98106 mEq/L</td>
</tr>
<tr>
<td>Cholesterol, total plasma</td>
<td>Desirable</td>
<td>&lt;5.20 mmol/L</td>
<td>&lt;200 mg/dL</td>
</tr>
<tr>
<td></td>
<td>Borderline</td>
<td>5.206.18 mmol/L</td>
<td>200239 mg/dL</td>
</tr>
<tr>
<td><strong>Undesirable</strong></td>
<td><strong>Desirable</strong></td>
<td><strong>Borderline</strong></td>
<td><strong>Undesirable</strong></td>
</tr>
<tr>
<td>----------------</td>
<td>-------------</td>
<td>---------------</td>
<td>---------------</td>
</tr>
<tr>
<td>6.21 mmol/L</td>
<td>&gt;4.05 mmol/L</td>
<td>≤4.11 mmol/L</td>
<td>≤1.55 mmol/L</td>
</tr>
<tr>
<td>240 mg/dL</td>
<td>&gt;130 mg/dL</td>
<td>≤130 mg/dL</td>
<td>≤69 mg/dL</td>
</tr>
<tr>
<td>0.02586</td>
<td>0.02586</td>
<td>0.02586</td>
<td>0.02586</td>
</tr>
</tbody>
</table>

| **Cholesterol, High-Density Lipoprotein (HDL Cholesterol), plasma** |
|------------------|------------------|
| **Desirable**    | **Borderline**   |
| 3.364.11 mmol/L | ≤3.91 mmol/L     |
| 130159 mg/dL    | ≤159 mg/dL       |
| 0.02586         | 0.02586          |

| **Cholesterol, Low-Density Lipoprotein (LDL Cholesterol), plasma** |
|------------------|------------------|
| **Desirable**    | **Borderline**   |
| 4.14 mmol/L      | ≥3.36 mmol/L     |
| 160 mg/dL        | ≥130 mg/dL       |
| 0.02586          | 0.02586          |

<table>
<thead>
<tr>
<th><strong>Corticotropin (ACTH) plasma, 8 AM</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>National Diabetes Data Group</td>
</tr>
<tr>
<td>≥7.8 mmol/L</td>
</tr>
<tr>
<td>≤140 mg/dL</td>
</tr>
<tr>
<td>0.05551</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Glucose, plasma</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overnight fast, normal</strong></td>
</tr>
<tr>
<td>4.26.4 mmol/L</td>
</tr>
<tr>
<td>75115 mg/dL</td>
</tr>
<tr>
<td>0.05551</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Glucose Tolerance Test, 2-h postprandial plasma glucose</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Normal</strong></td>
</tr>
<tr>
<td>≤7.8 mmol/L</td>
</tr>
<tr>
<td>≤140 mg/dL</td>
</tr>
<tr>
<td>0.05551</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Gonadal Steroids, plasma</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Androstenedione</strong></td>
</tr>
<tr>
<td>Men</td>
</tr>
<tr>
<td>3.05.0 mmol/L</td>
</tr>
<tr>
<td>0.81.3 ng/mL</td>
</tr>
<tr>
<td>3.492</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Testosterone</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Women</td>
</tr>
<tr>
<td>&lt;3.5 nmol/L</td>
</tr>
<tr>
<td>&lt;1 ng/mL</td>
</tr>
<tr>
<td>3.467</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Gonadotropins, plasma</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Follicle-Stimulating Hormone (FSH)</strong></td>
</tr>
<tr>
<td>Women, basal</td>
</tr>
<tr>
<td>1.49.6 IU/L</td>
</tr>
<tr>
<td>1.49.6 mIU/mL</td>
</tr>
<tr>
<td>Men</td>
</tr>
<tr>
<td>1083 mmol/L</td>
</tr>
<tr>
<td>310 ng/mL</td>
</tr>
<tr>
<td>3.467</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Insulin, plasma</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Fasting</td>
</tr>
<tr>
<td>35145 pmol/L</td>
</tr>
<tr>
<td>520 uU/mL</td>
</tr>
<tr>
<td>7.175</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Structure</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Human Chorionic Gonadotropin Subunit (-hCG), plasma</strong></td>
</tr>
<tr>
<td>Free</td>
</tr>
<tr>
<td>&lt;5 IU/L</td>
</tr>
<tr>
<td>96.05</td>
</tr>
<tr>
<td>Test</td>
</tr>
<tr>
<td>-------------------------------------</td>
</tr>
<tr>
<td><strong>Insulin C Peptide</strong>, plasma</td>
</tr>
<tr>
<td><strong>Insulin-Like Growth Factor I (IGF-I, Somatomedin-C)</strong></td>
</tr>
<tr>
<td><strong>Lactate</strong>, plasma</td>
</tr>
<tr>
<td><strong>Magnesium</strong>, serum</td>
</tr>
<tr>
<td><strong>Osmolality</strong>, plasma</td>
</tr>
<tr>
<td><strong>Oxytocin</strong>, plasma</td>
</tr>
<tr>
<td><strong>Phosphorus</strong>, inorganic, serum</td>
</tr>
<tr>
<td><strong>Pyruvate</strong>, plasma</td>
</tr>
<tr>
<td><strong>Renin Activity</strong>, plasma, normal-sodium intake</td>
</tr>
<tr>
<td><strong>Sodium</strong>, serum</td>
</tr>
<tr>
<td><strong>Thyroid Function Tests</strong></td>
</tr>
<tr>
<td>Free thyroxine estimate</td>
</tr>
<tr>
<td>Radioactive iodine uptake, 24 h</td>
</tr>
<tr>
<td>Resin T₃ uptake, serum</td>
</tr>
<tr>
<td>Reverse triiodothyronine (rT₃), serum</td>
</tr>
<tr>
<td>Thyrotrpin (TSH), serum</td>
</tr>
<tr>
<td>Triiodothyronine (T₃), serum</td>
</tr>
<tr>
<td>Triglycerides, plasma</td>
</tr>
<tr>
<td><strong>Vitamin D</strong>, see Calciferols</td>
</tr>
</tbody>
</table>
Figure 1-1 Comparison of determinants of endocrine (A) and paracrine (B) signaling.
Figure 1-3 Diseases caused by mutations in G-protein-coupled receptors. All are human conditions with the exception of the final two entries, which refer to the mouse. (AD = autosomal dominant; AR = autosomal recessive inheritance.) Loss of function refers to inactivating mutations of the receptor, and gain of function to activating mutations. Abbreviations for G-protein-coupled receptors: ACTH = adrenocorticotropic hormone; LH = luteinizing hormone; TSH = thyroid-stimulating hormone; PTH-PTHrP = parathyroid hormone and parathyroid hormone-related peptide; MSH = melanocyte-stimulating hormone; GHRH = growth hormone-releasing hormone; FSH = follicle-stimulating hormone. (Reproduced from Mayo K. In Conn PM, Melmed S (eds), Endocrinology: Basic and Clinical Principles. Totowa, NJ, Humana Press, 1997, page 27.)
**Figure 1-4** Peripheral feedback mechanism and a million-fold amplifying cascade of hormonal signals. Environmental signals are transmitted to the central nervous system, which innervates the hypothalamus, which responds by secreting nanogram amounts of a specific hormone. Releasing hormones are transported down a closed portal system, pass the blood-brain barrier at either end through fenestrations, and bind to specific anterior pituitary cell membrane receptors to elicit secretion of micrograms of specific anterior pituitary hormones. These enter the venous circulation through fenestrated local capillaries, bind to specific target gland receptors, trigger release of micrograms to milligrams of daily hormone amounts, and elicit responses by binding to receptors in distal target tissues. Peripheral hormone receptors enable widespread cell signaling by a single initiating environmental signal, thus facilitating intimate homeostatic association with the external environment. Arrows with a black dot at their origin indicate a secretory process. *(Reproduced from Normal AW, Litwack G. Hormones, 2nd edn. New York, Academic Press, 1997, p 14.)*
Figure 1-5 Model for regulation of anterior pituitary hormone secretion by three tiers of control. Hypothalamic hormones impinge directly on their respective target cells. Intrapituitary cytokines and growth factors regulate tropic cell function by paracrine (and autocrine) control. Peripheral hormones exert negative feedback inhibition of respective pituitary trophic hormone synthesis and secretion. (Reproduced with permission from Ray D, Melmed S. Pituitary cytokine and growth factor expression and action. Endocrin Rev 1997; 18:206228.)
Figure 2-1 Simple pedigree of the propositi (arrow) and first-order relatives should be the standard family history in a new patient workup. If the patient has children, their health status should be included as well.
Figure 3-1 Different modes of utilization of polypeptide hormones in expression of their biologic actions. The peptide hormones are expressed in at least four ways in fulfilling their functions as cellular messenger molecules: (1) endocrine mode, for purposes of communication among organs (e.g., pituitary-thyroid axis); (2) paracrine mode, for communication among adjacent cells, often located within endocrine organs; (3) neuroendocrine mode, for synthesis and release of peptides from specialized peptidergic neurons for action on distant organs through the blood stream (e.g., neuroendocrine peptides of hypothalamus); and (4) neurotransmitter mode, for action of peptides in concert with classical amino acid-derived amnergic transmitters in the neuronal communication network. Identical polypeptides are often utilized in the nervous system both as neuroendocrine hormones and as neurotransmitters. In some instances, the same gene product is used in all four modes of expression.
Figure 3-2 Steps in the cellular synthesis of polypeptide hormones. Steps that take place within the nucleus include transcription of genetic information into a messenger ribonucleic acid (mRNA) precursor (pre-mRNA) followed by post-transcriptional processing, which includes RNA cleavage, excision of introns, and rejoicing of exons, resulting in formation of mRNA. Ends of mRNA are modified by addition of methylguanosine caps at the 5' end and addition of poly(A) tracts at the 3' ends. The cytoplasmic mRNA is assembled with ribosomes. Amino acids, carried by aminoacylated transfer RNAs (tRNAs), are then polymerized into a polypeptide chain. The final step in protein synthesis is that of post-translational processing. These processes take place both during growth of the nascent polypeptide chain (cotranslational) and after release of the completed chain (post-translational), and they include proteolytic cleavages of polypeptide chain (conversion of pre-prohormones or prohormones to hormones), derivatizations of amino acids (e.g., glycosylation, phosphorylation), and cross-linking and assembly of the polypeptide chain into its conformed structure. The diagram depicts post-translational synthesis and processing of a typical secreted polypeptide, which requires vectorial, or unidirectional, transport of the polypeptide chain across the membrane bilayer of the endoplasmic reticulum, thus resulting in sequestration of the polypeptide in the cisternae of the endoplasmic reticulum, a first step in the export of proteins destined for secretion from the cell (see Fig. 3-6). Most translational processing occurs within the cell as depicted (presecretory) and in some instances outside the cell, when further proteolytic cleavages or modifications of the protein may take place (postsecretory). CHO, carbohydrate.
Figure 3-3 Diagrammatic depiction of two configurations of precursors of polypeptide hormones. Diagrams represent polypeptide backbones of protein sequences encoded in mRNA. One form of precursor consists of the NH₂-terminal signal, or presequence, followed by the apoprotein portion of the polypeptide that needs no further proteolytic processing for activity. A second form of precursor is a pre-prohormone that consists of the NH₂-terminal signal sequence followed by a polyprotein, or prohormone, sequence consisting of two or more peptide domains linked together that are subsequently liberated by cleavages during post-translational processing of the prohormone. The reason for synthesis of polypeptide hormones in the form of precursors is only partly understood. Clearly, NH₂-terminal signal sequences function in the early stages of transport of polypeptide into the secretory pathway. Prohormones, or polyproteins, often serve to provide a source of multiple bioactive peptides. However, many prohormones contain peptide sequences that are removed by cleavage and have no known biologic activity, and they are referred to as cryptic peptides. Other peptides may serve as spacer sequences between two bioactive peptides (e.g., the C peptide of proinsulin). In instances in which a bioactive peptide is located at the COOH terminus of the prohormone, the NH₂-terminal prohormone sequence may simply facilitate cotranslational translocation of polypeptide in endoplasmic reticulum.
Figure 3-4 Diagrammatic illustration of primary structures of several prohormones. The *darkly shaded areas* of prohormones denote regions of sequence that constitute known biologically active peptides after their post-translational cleavage from prohormones. Sequences indicated by hatching denote regions of precursor that alter the biologic specificity of that region of precursor. For example, the precursor contains the sequence of -melanocyte-stimulating hormone (-MSH), but when the latter is covalently attached to the clip peptide, it constitutes adrenocorticotropic hormone (corticotropin, ACTH). Somatostatin-28 (SS-28) is an NH$_2$-terminally extended form of somatostatin-14 (SS-14) that has higher potency than somatostatin-14 on certain receptors. The neurophysin sequence linked to the COOH terminus of vasopressin (ADH) functions as a carrier protein for hormone during its transport down the axon of neurons in which it is synthesized. Precursor proenkephalin represents a polyprotein that contains multiple similar peptides within its sequence, either met-enkephalin (M) or leu-enkephalin (L). Procalcitonin and procalcitonin gene-related product (CGRP) share identical NH$_2$-terminal sequences but differ in their COOH-terminal regions as a result of alternative splicing during the post-transcriptional processing of the RNA precursor. -LPH, -lipotropin; GLP, glucagon-like peptide; IP, intervening peptide.
Figure 3-5 Schematic representation of subcellular organelles involved in transport and secretion of polypeptide hormones or other secreted proteins within a protein-secreting cell. (1) Synthesis of proteins on polyribosomes attached to endoplasmic reticulum (RER) and vectorial discharge of proteins through the membrane into the cisterna. (2) Formation of shuttling vesicles (transition elements) from endoplasmic reticulum followed by their transport to and incorporation by the Golgi complex. (3) Formation of secretory granules in the Golgi complex. (4) Transport of secretory granules to the plasma membrane, fusion with the plasma membrane, and exocytosis resulting in the release of granule contents into the extracellular space. Note that secretion may occur by transport of secretory vesicles and immature granules as well as mature granules. Some granules are taken up and hydrolyzed by lysosomes (crinophagy). Golgi, Golgi complex; RER, rough endoplasmic reticulum; SER, smooth endoplasmic reticulum. (From Habener JF. Hormone biosynthesis and secretion. In Felig P, Baxter JD, Broadus AE, et al. [eds]. Endocrinology and Metabolism. New York, McGraw-Hill, 1981, pp 2959. Copyright © 1981 by McGraw-Hill, Inc. Used by permission of McGraw-Hill Book Company.)
Figure 3-6 Diagram depicting cellular events in initial stages of synthesis of a polypeptide hormone according to the signal hypothesis. In this schema, a signal recognition particle, consisting of a complex of six proteins and an RNA (7S RNA), interacts with the NH₂-terminal signal peptide of the nascent polypeptide chain after approximately 70 amino acids are polymerized, which results in the arrest of further growth of the polypeptide chain. The complex of the signal recognition particle and the polyribosome nascent chain remains in a state of translational arrest until it recognizes and binds to a docking protein, which is a receptor protein located on the cytoplasmic face of the endoplasmic reticular membrane. This interaction of the signal recognition particle complex with docking protein releases the translational block, and protein synthesis resumes. The nascent polypeptide chain is discharged across the membrane bilayer into the cisterna of the endoplasmic reticulum and is released from the signal peptide by cleavage with a signal peptidase located in the cisternal face of the membrane. In this model, the signal peptide is cleaved from the polypeptide chain by signal peptidase before the chain is completed (cotranslational cleavage). The configuration of the polypeptide during transport across the membrane and the forces and mechanisms responsible for its translocation are unknown. The loop, or hairpin, configuration of the chain that is shown is an arbitrary model; other models are equally possible.
Figure 3-7 Regulatory feedback loops of the hypothalamic-pituitary-target organ axis. Being a combination of both stimulatory and inhibitory factors, hormones often act in concert to maintain homeostatic balance in the presence of physiologic or pathophysiologic perturbations. The concerted actions of hormones typically establish closed feedback loops by stimulatory and inhibitory effects coupled to maintain homeostasis.
Figure 3-8 Diagrammatic structure of a "consensus" gene encoding a prototypical polypeptide hormone. Such a gene typically consists of a promoter region and a transcription unit. The transcription unit is the region of deoxyribonucleic acid (DNA) composed of exons and introns that is transcribed into a messenger ribonucleic acid (mRNA) precursor. Transcription begins at the cap site sequence in DNA and extends several hundred bases beyond the poly(A) addition site in the 3' region. During post-transcriptional processing of the RNA precursor, the 5' end of mRNA is capped by addition of methylguanosine residues. The transcript is then cleaved at the poly(A) addition site approximately 20 bases 3' to the AATAAA signal sequence, and the poly(A) tract is added to the 3' end of the RNA. Introns are cleaved from the RNA precursor, and exons are joined together. Dinucleotides GT and AG are invariably found at the 5' and 3' ends of introns. Translation of mRNA invariably starts with the codon ATG for methionine. Translation is terminated when the polyribosome reaches the stop codon TGA, TAA, or TAG. The promoter region of the gene located 5' to the cap site contains numerous short regulatory DNA sequences that are targets for interactions with specific DNA-binding proteins. These sequences consist of the basal constitutive promoter (TATA box), metabolic response elements that modulate transcription (e.g., in response to cAMP, steroid hormone receptors, and thyroid hormone receptors), and tissue-specific enhancers and silencers that permit or prevent transcription of the gene, respectively. The enhancer and silencer elements direct expression of specific subsets of genes to cells of a given phenotype. Whether a gene is or is not expressed in a particular cellular phenotype depends on complex interactions of the various DNA-binding proteins among themselves and, most important, with the TATA box proteins of the basal constitutive promoter.
Figure 3-9 Diagram of the pancreatic glucagon gene and its encoded messenger RNA (mRNA) (complementary DNA). The glucagon gene is an example of a gene in which exons precisely encode separate functional domains. The gene consists of six exons (E1 to E6) and five introns (1A to 1E). The mRNA encoding pre-proglucagon, the protein precursor of glucagon, consists of 10 specific regions: from left to right, a 5'-untranslated sequence (UN-TX, unshaded), a signal sequence (S, stippled), an NH₂-terminal extension sequence (N, hatched), glucagon (Gluc, shaded), a first intervening peptide (IP-I, hatched), a first glucagon-like peptide (GLP-I, shaded), a second intervening peptide (IP-II, hatched), a second glucagon-like peptide (GLP-II, shaded), a dilysyl dipeptide (hatched) after the glucagon-like peptide II sequence, and an untranslated region (UN-TX, unshaded). Exons from left to right encode the 5'-untranslated region, signal sequence, glucagon, glucagon-like peptide I, glucagon-like peptide II, and 3'-untranslated sequence. Letters shown above the mRNA denote amino acids located at positions in pre-proglucagon that are cleaved during cellular processing of precursor. The amino acid methionine (M) marks the initiation of translation of mRNA into pre-proglucagon. H, histidine; K, lysine; Q, glutamine; R, arginine.
Figure 3-10 Diagram of an endocrine cell showing potential control points for regulation of gene expression in hormone production. Specific effector substances bind either to plasma membrane receptors (peptide effectors) or to cytosolic or nuclear receptors (steroids), which leads to initiation of a series of events that couple the effector signal with gene expression. In the illustration shown, peptide effector-receptor complex interactions act initially through activation of adenylate cyclase (AC) coupled with a guanosine triphosphate-binding protein (G). Coupling factors and substances such as glucose, cyclic adenosine monophosphate, and cations activate protein kinases, resulting in a series of phosphorylations of macromolecules. As discussed in the text, specific effectors for various endocrine cells appear to act at one or more of the indicated five levels of gene expression, with the possible exception of post-translational processing of prohormones, for which no definite examples of metabolic regulation have yet been found.
Figure 3-11 Diagram showing three cell-surface receptor-coupled signal transduction pathways involved in the activation of a superfamily of nuclear transcription factors. Peptide hormone molecules (H1, H2, and H3) interact with sensor receptors (R1, R2, and R3) coupled to the diacylglycerol (DAG)-protein kinase C (PKC), the cyclic adenosine monophosphate (cAMP)-protein kinase A (PKA), and the calcium-calmodulin pathways in which small diffusible second messenger molecules are generated (DAG, cAMP, Ca$^{2+}$). The third messengers or effector protein kinases are generated and phosphorylate transcription factors such as members of the CREB/ATF and jun/AP-1 families of DNA-binding proteins to modulate DNA-binding affinities or transcriptional activation, or both. The various proteins bind as dimers determined by a poorly understood code that is not promiscuous in as much as only certain homodimer or heterodimer combinations are permissible. AP-1, activator protein 1; ATF, activating transcription factor; CaMK, calcium/calmodulin-dependent protein kinase; CREB, cAMP response element-binding protein.
Figure 3-12 Schema indicating levels in expression of genetic information at which diversification of information encoded in a gene may take place. The three major levels of genetic diversification are (1) gene duplication, a process that occurs in terms of evolutionary time; (2) variation in the processing of ribonucleic acid (RNA) precursors, which results in formation of two or more messenger RNAs (mRNAs) by way of alternative pathways of splicing of transcript (see Fig. 3-13 and Fig. 3-14); and (3) use of alternative patterns in processing of protein biosynthetic precursors (polypeptides, or prohormones). These three levels in gene expression provide a means for diversification of gene expression at levels of deoxyribonucleic acid (DNA), RNA, or protein. One or a combination of these processes leads to formation of the final biologically active peptide or hormone. In the diagram, loops depicted in transcripts denote introns; in diagrammatic structures of proteins, the stippled, shaded, and unshaded areas denote exons. See text for details.
Figure 3-13 Utilization of alternative promoters in the expression of genes as a means to generate biologic diversification of gene expression. The use of alternative promoters allows a gene to be expressed in a variety of unique contexts that alter the properties of the messenger ribonucleic acid (mRNA) that is expressed. Such alternative promoter usage may render the mRNA more or less stable, affect translational efficiencies, or switch the translation of one protein isoform to another. The use of alternative promoters in genes characteristically occurs during development, or after development is completed, to designate tissue-specific patterns of expression of the gene. Exons are shown as boxes whose protein-coding regions are shaded. Introns are designated by horizontal lines. Dashed lines indicate introns that are spliced out. (Adapted from Ayoubi TAY, Van De Ven WJM. Regulation of gene expression by alternative promoters. FASEB J 1996; 10:453460.)
Figure 3-14  Alternative exon splicing provides a means to generate biologic diversification of gene expression. Mechanisms of exon skipping or switching and intron slippage are frequently utilized in the alternative processing of premessenger ribonucleic acids (mRNAs) to provide unique mRNAs and encoded proteins during development and in a tissue-specific pattern of expression in the fully differentiated tissues or organs. Exons are shown as boxes with protein-coding regions shaded to designate origin of protein isoforms. Introns are depicted as horizontal lines. Dashed lines denote spliced-out introns.
Figure 3-15 Alternative translational initiation sites are used to change the coding sequences of messenger ribonucleic acids to encode different protein isoforms. The two mechanisms illustrated involve loose scanning and reinitiation of translation. See text.
Figure 4-1 Signal transduction by hormones and other ligands that act via nuclear receptors. HRE, hormone response element; mRNA, messenger ribonucleic acid.
Figure 4-2 Domain structure of nuclear receptors.
Figure 4-3 Structural basis of nuclear receptor ligand binding and cofactor recruitment.
Figure 4-4 Structural basis of nuclear receptor (NR) DNA binding specificity. Ribbon diagrams of receptor DNA-binding domains (DBDs) are shown. A, Steroid hormone receptor binding as homodimer to inverted repeat (arrows) of AGAACA half-site. B, RXR-NR heterodimer binding to direct repeat of AGGTCA. The position of the P-box, the region of the DBD that makes direct contact with DNA, is shown. N, number of base pairs between the two half-sites; RXR, retinoid X receptor.
Figure 4-5 Coactivators and corepressors in transcriptional regulation by nuclear receptors. CBP, calcium-binding protein; DRIP, D receptor-interacting protein; HRE, hormone response element; HAT, histone acetyltransferase; HDAC, histone deacetylase; N-CoR, nuclear receptor corepressor; NR, nuclear receptor; PCAF, p300/CBP-associated factor; SMRT, silencing mediator of retinoid and thyroid receptors; TRAP, thyroid hormone receptor-associated protein.
Figure 4-6 Repression and activation functions augmenting the dynamic range of transcriptional regulation by nuclear receptors. HRE, hormone response element.
Figure 5-1 Receptor tyrosine kinases. This diagram illustrates 3 of the 16 families of receptor tyrosine kinases. All receptor tyrosine kinases possess an extracellular domain containing the ligand-binding site, a single transmembrane domain, and an intracellular domain containing the tyrosine kinase domain. Several structural motifs (i.e., cysteine-rich domain, immunoglobulin-like domain, tyrosine kinase domain) in these receptor tyrosine kinases are indicated on the right side of the figure. Cys, cysteine; EGF, epidermal growth factor; Ig, immunoglobulin; PDGF, platelet-derived growth factor.
Figure 5-2 Ligand-induced dimerization of receptors. Two molecular mechanisms of ligand-induced receptor dimerization are illustrated. In the case of the platelet-derived growth factor, the ligand is dimeric and therefore contains two receptor binding sites. In the case of growth hormone, a single ligand molecule contains two binding sites so that it can bind simultaneously to two receptor molecules.
Figure 5-3: Phosphorylation of tyrosine residues in the activation loop leads to activation of the insulin receptor tyrosine kinase. A hypothetical mechanism for ligand-stimulated activation of the insulin receptor tyrosine kinase is illustrated. The model is based on the three-dimensional structure of the isolated insulin receptor tyrosine kinase as determined by x-ray crystallography. In the inactive insulin receptor kinase (left), Tyr1162 blocks the active site so that substrates cannot bind. In contrast, when the tyrosine residues in the activation loop (including Tyr1162) become phosphorylated (right), Tyr1162 moves out of the way, and there is a conformational change that allows binding of adenosine triphosphate (ATP) and protein substrate so that the kinase reaction can proceed.
Figure 5-4 Simplified model of signaling pathways downstream from the insulin receptor. Insulin binds to the insulin receptor, thereby activating the receptor tyrosine kinase to phosphorylate tyrosine residues on insulin receptor substrates (IRSs) including IRS-1 and IRS-2. Consequently, phosphorylated tyrosine residues in IRS molecules bind to Src homology 2 (SH2) domains in molecules such as growth factor receptor-binding protein 2 (Grb-2) and the p85 regulatory subunit of phosphatidylinositol (PI) 3-kinase. These SH2 domain-containing proteins initiate two distinct branches of the signaling pathway. Activation of PI 3-kinase leads to activation of phosphoinositide-dependent kinases (PDKs) 1 and 2, which activates multiple protein kinases including Akt/protein kinase B, atypical protein kinase C (PKC) isoforms, and serum/glucocorticoid-activated protein kinases (Sgk). Grb-2 interacts with m-SOS, a guanine nucleotide exchange factor that activates Ras. Activation of Ras triggers a cascade of protein kinases leading to the activation of mitogen-activated protein (MAP) kinase.
Figure 5-5 Mutations leading to constitutive activation of Ret. "Wild-type" Ret has intramolecular disulfide bonds formed by two cysteine residues in the same receptor molecule (left). When one of the two cysteine residues is mutated, the unpaired cysteine residue is available to form an intermolecular disulfide bond with a cysteine residue on another receptor molecule. This leads to receptor dimerization (right), which in turn leads to constitutive activation of the receptor tyrosine kinase. This type of mutation has been identified in patients with multiple endocrine neoplasia type 2.
Figure 5-6 Cytokine receptors are composed of multiple subunits and bind to one or more members of the Janus kinase (JAK) family of tyrosine kinases. A, Growth hormone (GH), like prolactin and leptin, binds to receptor homodimers and activates JAK2. B, Interferon (IFN) homodimers bind to their ligand-binding R1 subunits. The R2 subunits are then recruited, leading to activation of JAK1, which binds to R1 subunit, and JAK2, which binds to R2 subunit. Both subunits and both JAKs are necessary for responses to IFN. C, Interleukin-2 (IL-2) binds to receptors composed of three subunits: a c subunit shared with receptors for ILs 4, 7, 9, and 15; an IL-2R subunit shared with the IL-15 receptor; and a noncytokine receptor subunit, IL-2R subunit. IL-2 activates both JAK3, bound to the c subunit, and JAK1, bound to IL2-R. Extracellular regions of homology are indicated by the black lines and patterns. Intracellular regions of homology are indicated by the white boxes.
Cytokines activate signal transducers and activators of transcription (STATs). STAT proteins are latent cytoplasmic transcription factors. STATs bind, through their Src homology 2 (SH2) domains, to one or more phosphorylated tyrosines in activated receptor-JAK complexes. Once bound, they themselves are tyrosyl phosphorylated, presumably by the receptor-associated JAKs. STATs then dissociate from the receptor-JAK complexes, homodimerize or heterodimerize with other STAT proteins, move to the nucleus, and bind to gamma-activated sequence-like elements (GLES) in the promoters of cytokine-responsive genes. (Adapted from figure by J. Herrington, with permission.)
**Figure 5-8** The G protein-coupled receptor (GPCR) superfamily: diversity in ligand binding and structure. Each panel depicts various members of the GPCR superfamily in cartoon form. The seven membrane-spanning helices are shown as cylinders with the extracellular amino terminus and three extracellular loops above and the intracellular carboxyl terminus and three intracellular loops below. The superfamily can be divided into three subfamilies on the basis of amino acid sequence conservation within the transmembrane helices. Family 1 includes (A) the opsins, in which light (jagged arrow) causes isomerization of retinal covalently bound within the pocket created by the transmembrane helices (bar); (B) monoamine receptors, in which agonists (arrow) bind noncovalently within the pocket created by the transmembrane helices (bar); (C) receptors for peptides such as vasopressin, in which agonist binding (arrow) may involve parts of the extracellular amino terminus and loops as well as the transmembrane helices (bar); and (D) glycoprotein hormone receptors, in which agonists (oval) bind to the large extracellular amino terminus, thereby activating the receptor through as yet undefined interactions with the extracellular loops or transmembrane helices (arrow). Family 2 includes receptors for peptide hormones such as parathyroid hormone (PTH) and secretin. Agonists (arrow) may bind to residues in the extracellular amino terminus and loops as well as transmembrane helices (bar). Family 3 includes the extracellular Ca$^{2+}$ sensing receptor and metabotropic glutamate receptors. Agonists (sphere) bind in a cleft of the Venus flytrap-like domain in the large extracellular amino terminus, thereby activating the receptor through as yet undefined interactions with the extracellular loops or transmembrane helices (arrow).
**Figure 5-9** The G protein guanosine triphosphatase (GTPase) and G protein-coupled receptor (GPCR) desensitization-resensitization cycle. In each panel, the stippled region denotes the plasma membrane with extracellular above and intracellular below. In the basal state, the G protein is a heterotrimer with guanosine diphosphate (GDP) tightly bound to the subunit. The agonist-activated GPCR catalyzes release of GDP, which permits guanosine triphosphate (GTP) to bind. The GTP-bound subunit dissociates from the dimer. Arrows from subunit to effector and from dimer to effector indicate regulation of effector activity by the respective subunits. Arrow from effector to subunit indicates regulation of its GTPase activity by effector interaction. Under physiologic conditions, effector regulation by G protein subunits is transient and is terminated by the GTPase activity of the subunit. The latter converts bound GTP to GDP, thus returning the subunit to its inactivated state with high affinity for the dimer, which reassociates to form the heterotrimer in the basal state. In the basal state, the receptor kinase and arrestin are shown as cytosolic proteins. Dissociation of the GTP-bound subunit from the dimer permits the latter to facilitate binding of receptor kinase to the plasma membrane (arrow from dimer to receptor kinase). Plasma membrane binding permits the receptor kinase to phosphorylate the agonist-bound GPCR (depicted here as occurring on the carboxyl-terminal tail of the GPCR, but sites on intracellular loops are also possible). GPCR phosphorylation in turn facilitates arrestin binding to GPCR, resulting in desensitization. Endocytic trafficking of arrestin-bound GPCR and recycling to the plasma membrane during resensitization are not depicted here.
Figure 6-1 Seven-logarithm range of normal concentrations for the plasma concentrations of endocrine tests. DHEA, dehydroepiandrosterone; FSH, follicle-stimulating hormone; FT₄, free thyroxine; FT₃, free triiodothyronine; LH, luteinizing hormone; T₃, triiodothyronine; T₄, thyroxine; TSH, thyrotropin.
Figure 6-2 A, Principles of competitive binding assays. B, Typical dose-response curve.
Figure 6-3 Comparison of an immunologic technique for measuring hormone concentration versus a receptor technique for measuring hormone activity. ATP, adenosine triphosphate; cAMP, cyclic adenosine monophosphate.
Figure 6-4 A, Principles of immunometric assays. Ag, antigen; ATB1, capture antiserum; ATB2, signal antiserum. B, Typical dose-response curve. NSB, nonspecific binding.
Figure 6-5 Immunometric "high-dose hook effect." The response signal reaches a maximum and then decreases when the antigen concentration exceeds the limit of the assay.
Figure 6-6 Assay interferences caused by heterophile antibodies, which result in either false high or false low results. Ag, antigen; ATB1, capture antiserum; ATB2, signal antiserum.

False High

Capture ATB1
Binding ATB
Signal ATB

False Low

Capture ATB1
Binding ATB
Ag correctly bound
Figure 6-7 Mass spectrum illustrating the concurrent measurement of 10 cortisol-related compounds with one assay. cps, counts per second; LC, Supelcosil LC-18 column; MS/MS, tandem mass spectrometry. (Courtesy of Dr. R. Singh, Mayo Clinic.)
Figure 6-8 Flow diagram illustrating appropriate preanalytic conditions for measuring plasma catecholamines. EDTA, edetate.
Figure 6-9 Effect of analytic bias, or shift, on the number of patients with elevated levels of thyrotropin (TSH).
**Figure 6-10** Nonproportional dilutions. Discordant values are produced when samples do not dilute linearly. (NI = undiluted [neat].)
Figure 7-1 Three types of hypothalamic neurosecretory cells. Left, A magnocellular neuron that secretes arginine vasopressin or oxytocin (AVP, OXY). The cell body, which is located in the supraoptic or paraventricular hypothalamic nucleus (SON, PVH), projects its neuronal process into the neural lobe, and neurohormone is released from nerve endings. Center, Similar peptidergic neurons are located in the medial basal hypothalamus in nuclear groups including the PVH and arcuate nucleus of the hypothalamus (Arc). The neuropeptides in this case are released into the specialized blood supply to the pituitary to regulate its secretion. Similar in plan are neurosecretory neurons that terminate in relation to another neuron (right). These projection neurons are found in sites including the PVH, Arc, and lateral hypothalamic area (LHA) that project to autonomic preganglionic neurons in the brain stem and spinal cord. Such substances act as neurotransmitters or neuromodulators. ACTH, corticotropin; CART, cocaine and amphetamine-regulated transcript; CRH, corticotropin-releasing hormone; FSH, follicle-stimulating hormone; GH, growth hormone; GHRH, growth hormone-releasing hormone; GnRH, gonadotropin-releasing hormone; LH, luteinizing hormone; MCH, melanin-concentrating hormone; ORX, orexin-hypocretin; POMC, pro-opiomelanocortin; TRH, thyrotropin-releasing hormone; TSH, thyrotropin.
Neurobiologic features of the peptidergic neuron. Neurosecretory neurons can be regarded as having secretory functions that are in many ways analogous to those of glandular cells. A secretory product, which is formed on the endoplasmic reticulum under the direction of messenger ribonucleic acid, is packaged in granules and transported along the axon by axoplasmic flow to reach nerve terminals, where the granules are released. Virtually all neurons carry out similar functions. Some secrete neurotransmitters, such as acetylcholine or norepinephrine; others, such as motor nerves, secrete acetylcholine and myotropic factors. In all neurons there is a constant orthodox (forward) flow of cytoplasm and formed elements such as mitochondria. Retrograde flow also takes place to bring substances that enter nerve endings back to the body of the cell. In typical neurotransmitter neurons, the neurotransmitters are synthesized by enzymes and are packaged into secretory granules. These granules are transported in a manner similar to that of neuropeptide-containing granules. In many neurons cosecretion of one or two peptides may occur in association with secretion of a classic neurotransmitter. (From Reichlin S. Summarizing comments. In Gotto AM Jr, Peck EJ Jr, Boyd AE III, et al [eds]. Brain Peptides: A New Endocrinology. New York, Elsevier/North-Holland, 1979, pp 379-403.)
Figure 7-3 A. Human hypothalamic-pituitary unit showing the relationship to the sella turcica, brain membranes, and optic chiasm. B. Midsagittal nuclear magnetic resonance scan of the brain of a normal woman, which corresponds to the diagram in A. Note the location of the pituitary stalk, the intense signal from the posterior pituitary, and the anatomic relationship to the optic commissure and the optic nerve. (See also Fig. 7-4A.) (Courtesy of Dr. Samuel Wolpert.)
Figure 7-4a A. Midsagittal view of the human brain showing the hypothalamus and neighboring structures.
Figure 7-4b. Base of the human brain, showing the hypothalamus and neighboring structures. On gross inspection, several landmarks outline the hypothalamus. It is bounded anteriorly by the optic chiasm, laterally by the sulci formed with the temporal lobes, and posteriorly by the mammillary bodies (in which the mammillary nuclei are located). Dorsally, the hypothalamus is delineated from the thalamus by the hypothalamic sulcus. The smooth, rounded base of the hypothalamus is the tuber cinereum; the pituitary stalk descends from its central region, which is termed the median eminence. The median eminence stands out from the rest of the tuber cinereum because of its dense vascularity, which is formed by the primary plexus of the hypophyseal portal system. The long portal veins run along the ventral surface of the pituitary stalk. (From Nauta WJ, Haymaker W. Hypothalamic nuclei and fiber connections. In Haymaker W, Anderson E, Nauta WJ [eds]. The Hypothalamus. Springfield, Ill, Charles C Thomas, 1969, pp 136209.)
Figure 7-7 The paraventricular nucleus of the hypothalamus (PVH). A series of photomicrographs illustrate distinct subdivisions of the PVH and the presence of multiple neuropeptide mediators. Magnocellular neurons project to the posterior lobe of the pituitary gland to release arginine vasopressin or oxytocin (AVP, OT). Parvicellular hypophyseotropic neurons project to the median eminence and release releasing factors such as corticotropin-releasing hormone (CRH). In addition to neurons that make neuropeptides within the nucleus, the PVH receives a dense innervation from neurotransmitters and neuropeptide neurons, including neuropeptide Y (NPY). Note the subdivisions of the PVH including the dorsal (dp), ventral (vp), and medial (mp) parvicellular and posterior magnocellular (pm) divisions. IR, immunoreactivity; 3v, third ventricle.
Figure 7-8 The median eminence is the functional connection of the hypothalamus and the pituitary gland. A and B, Distribution of corticotropin-releasing hormone and thyrotropin-releasing hormone (CRH-IR and TRH-IR) immunoreactivity in the external layer of the median eminence (ME ext) of the rat. CRH and TRH cell bodies reside in the medial division of the paraventricular hypothalamic nucleus. C, Arginine vasopressin (AVP) immunoreactivity in nerve endings in the internal layer of the median eminence (ME int). Arc, arcuate nucleus; 3v, third ventricle.
Figure 7-9 The tuberoinfundibular system is revealed by retrograde transport of cholera toxin subunit B (CtB). The location of hypothalamic cell bodies of neurons projecting to the median eminence (ME) and the posterior pituitary can be identified by microinjecting a small volume of retrograde tracer (CtB) into the median eminence of the rat (see Wiegand,112 Lechan113). A, Retrogradely labeled cells can be seen in the paraventricular and supraoptic nuclei of the hypothalamus (PVH, SON). B, Magnocellular neurons are observed in the SON. C, Labeled neurons are found in the posterior magnocellular group (pm) as well as the medial parvicellular subdivision (mp). The labeled cells in the PVH include those that contain corticotropin-releasing hormone and thyrotropin-releasing hormone. D, Retrogradely labeled cells are also found in the arcuate nucleus of the hypothalamus (Arc). These include neurons that release growth hormone-releasing hormone and dopamine. 3v, third ventricle; ot, optic tract.
Figure 7-10 Median sagittal section through the human brain to show the circumventricular organs (black). AP, area postrema; ME, median eminence; NH, neurohypophysis; OVLT, organum vasculosum of the lamina terminalis; PI, pineal body; SFO, subfornical organ; SCO, subcommissural organ; CP, choroid plexus. (From Weindl A. Neuroendocrine aspects of circumventricular organs. In Ganong WF, Martini L [eds]. Frontiers in Neuroendocrinology, vol 3. New York, Oxford University Press, 1973, pp 332.)
Figure 7-11 Biosynthesis of melatonin from tryptophan in the pineal gland. Step 1 is catalyzed by tryptophan hydroxylase, step 2 by aromatic-
N
Figure 7-15 Prolactin (PRL) and thyrotropin (TSH) secretory responses to intravenous injection of 800 µg of thyrotropin-releasing hormone (TRH) in humans. This figure shows that TRH induces discharge of both PRL and thyrotropin, that the effect in females is greater than that in males (presumably owing to estrogen sensitization of the pituitary), and that thyrotoxicosis inhibits the response of both PRL and thyrotropin to TRH. An inhibitory effect on the TRH response is noted at the upper limit of the normal range of thyroid hormone levels and is a sensitive test of minor degrees of thyroid hormone excess. Although TRH is a potent prolactin-releasing factor (PRF), there is evidence that there is another PRF physiologically connected to PRL regulation. (Replotted from data of Bowers C, Friesen HG, Hwang P, et al. Prolactin and thyrotropin release in man by synthetic pyroglutamylhistidyl-prolinamide. Biochem Biophys Res Commun 1971; 45:1033-1041.)
Figure 7-16 Regulation of the hypothalamic-pituitary-thyroid axis. AGRP, agouti-related protein; CART, cocaine and amphetamine-regulated transcript; CRH, corticotropin-releasing hormone; NPY, neuropeptide Y; POMC, proopiomelanocortin; T₃, triiodothyronine; T₄, thyroxine; TRH, thyrotropin-releasing hormone; TSH, thyrotropin; OB-R, leptin receptor.
Figure 7-17 Relationship between plasma thyrotropin levels and thyroid hormone as determined by plasma protein-bound iodine (PBI) measurements in humans and rats. These curves illustrate, in the human (A) and the rat (B), that plasma thyrotropin levels are a curvilinear function of plasma thyroid hormone level. Human studies were carried out by giving myxedematous patients successive increments of thyrotropin $T_4$ at approximately 10-day intervals. Each point represents simultaneous measurements of plasma PBI and plasma thyrotropin at various times in the six patients studied. The rat studies were performed by treating thyroidectomized animals with various doses of $T_4$ for 2 weeks before assay of plasma thyrotropin and plasma PBI. These curves illustrate that the secretion of thyrotropin is regulated over the entire range of thyroid hormone levels. At the normal set point for $T_4$, the small changes above and below the control level are followed by appropriate increases or decreases in plasma thyrotropin. TSH, thyrotropin; $T_4$, thyroxine. (A from Reichlin S, Utiger RD. Regulation of the pituitary thyroid axis in man: relationship of TSH concentration to concentration of free and total thyroxine in plasma. J Clin Endocrinol Metab 1967; 27:251255, copyright by The Endocrine Society. B from Reichlin S, Martin JB, Boshans RL, et al. Measurement of TSH in plasma and pituitary of the rat by a radioimmunoassay utilizing bovine TSH: effect of thyroidectomy or thyroxine administration on plasma TSH levels. Endocrinology 1970; 87:10221031, copyright by The Endocrine Society.)
Figure 7-19 Structure of human corticotropin-releasing hormone (CRH) gene and protein. The sequence coding for CRH occurs at the terminus of the prohormone. Cleavage sites and the terminal Gly position are shown. PAM, peptidylglycine alpha-amidating monooxygenase; PC1/PC2, prohormone convertases 1 and 2; ERE, estrogen regulating element; GRE, glucocorticoid regulating element; CRE, cyclic AMP-responsive element; UTR, untranslated. (Redrawn from data of Shibahara S, Morimoto Y, Furutani Y, et al. Isolation and sequence analysis of the human corticotropin-releasing factor precursor gene. EMBO J 1983; 2:775-779.)
Figure 7-20 Sequence comparison of members of the corticotropin-releasing hormone (CRH) peptide family. SPP, stresscopin related peptide; SCP, stresscopin.
Figure 7-22 Changes in plasma levels of corticotropin and serum levels of cortisol after intravenous injection of corticotropin-releasing hormone in a group of six normal men. The initial prompt response in corticotropin is followed by a somewhat delayed secondary change in cortisol. To convert corticotropin values to picomoles per liter, multiply by 0.2202. To convert cortisol values to millimoles per liter, multiply by 27.59. ACTH, adrenocorticotropic hormone. (From Grossman A, Kruseman ACN, Perry L, et al. New hypothalamic hormone, corticotropin-releasing factor, specifically stimulates the release of adrenocorticotropic hormone and cortisol in man. Lancet 1982; 1:921922.)
Figure 7-24 Regulation of neurons of the paraventricular nucleus (PVH) by diverse stressors. ADX, adrenalectomy; dp, dorsal PVH; IL-1, interleukin-1; mp, magnocellular PVH; NGFI-B, nerve growth factor I-B; pm, medial PVH. (Reprinted from Sawchenzko PE, et al. The paraventricular nucleus of the hypothalamus and the functional neuroanatomy of visceromotor responses to stress. Prog Brain Res 1996; 107:208. With permission from Elsevier Science.)
Figure 7-25 Neuronal inputs to neurons of the paraventricular nucleus. AVP, arginine vasopressin; BST, bed nucleus of the stria terminalis; CG, central gray; IGL, intergeniculate leaf; LDT, laterodorsal tegmental nucleus; MePO, medial preoptic nucleus; NTS, nucleus of the tractus solitarius; OT, oxytocin; OVLT, organum vasculosum of the lamina terminalis; PB, parabrachial nucleus; PIN, posterior intralaminar nucleus; PP, peripeduncular nucleus; PPN, pedunculopontine nucleus; SFO, subfornical organ. (Reprinted from Sawchenzo PE, et al. The paraventricular nucleus of the hypothalamus and the functional neuroanatomy of visceromotor responses to stress. Prog Brain Res 1996; 107:204. With permission from Elsevier Science.)
Figure 7-26 Regulation of the hypothalamic-pituitary-adrenal axis. ACTH, adrenocorticotropic hormone; AVP, arginine vasopressin; BST, bed nucleus of the stria terminalis; CNS, central nervous system; CRH, corticotropin-releasing hormone; CRIF, corticotropin release inhibiting factor; GABA, aminobutyric acid; 5-HT, 5-hydroxytryptamine; IL-1, interleukin-1; MeA, medial amygdala; MePO, medial preoptic; NPY, neuropeptide Y; NTS, nucleus of the tractus solitarius; OVLT, organum vasculosum of the lamina terminalis; POMC, pro-opiomelanocortin.
Figure 7-27 Diagram illustrating the genomic organization, messenger ribonucleic acid structure, and post-translational processing of the human growth hormone-releasing hormone (GHRH) prohormone. Few details are known about the transcriptional regulation of the GHRH gene except that distinct promoter sequences and alternative 5' exons are utilized by hypothalamic neurons and extrahypothalamic tissues. All of the amino acid residues required for bioactive GHRH peptides are encoded by exon 3. An amino-terminal exopeptidase that cleaves the Tyr-Ala dipeptide is primarily responsible for the inactivation of GHRH peptides in extracellular compartments. CPE, carboxypeptidase E; PAM, peptidylglycine alpha-amidating monooxygenase; PC1/PC2, prohormone convertases 1 and 2; UTR, untranslated region. (Compiled from data of Mayo K, Cerelli GM, Lebo RV, et al. Gene encoding human growth hormone-releasing factor precursor: structure, sequence, and chromosomal assignment. Proc Natl Acad Sci USA 1985; 82:6367; Frohman LA, Downs TR, Chomczynski P, Frohman MA. Growth hormone-releasing hormone: structure, gene expression and molecular heterogeneity. Acta Paediatr Scand [Suppl] 1990; 367:8186; and González-Crespo S, Boronat A. Expression of the rat growth hormone-releasing hormone gene in placenta is directed by an alternative promoter. Proc Natl Acad Sci USA 1991; 88:87498753.)
Figure 7-28 Response of normal men to growth hormonereleasing hormone (GHRH)(129) (1 µg/kg), ghrelin (1 µg/kg), or GHRH(129) and ghrelin administered by intravenous injection. Note the prompt release of GH, followed by a rather prolonged fall in hormone level in response to both secretagogues. Ghrelin alone was more efficacious than GHRH(129), and there was an additive effect from the two peptides administered simultaneously. (From Arvat E, Macario M, Di Vito L, et al. Endocrine activities of ghrelin, a natural growth hormone secretagogue (GHS), in humans: comparison and interactions with hexarelin, a nonnatural peptidyl GHS, and GH-releasing hormone. J Clin Endocrinol Metab 2001; 86:11691174.)
Figure 7-30 Regulation of the hypothalamic-pituitary-growth hormone (GH) axis. GH secretion by the pituitary is stimulated by GH-releasing hormone (GHRH) and is inhibited by somatostatin (SRIF). Negative feedback control of GH secretion is exerted at the pituitary level by insulin-like growth factor I (IGF-I) and by free fatty acids (FFA). GH itself exerts a short-loop negative feedback by the activation of SRIF neurons in the hypothalamic periventricular nucleus. These SRIF neurons directly synapse on arcuate GHRH neurons and project to the median eminence. Neuropeptide Y (NPY) neurons in the arcuate nucleus also indirectly modulate GH secretion by integrating peripheral GH, leptin, and ghrelin signals and projecting to periventricular SRIF neurons. Ghrelin is secreted from the stomach and is a putative natural ligand for the GH secretagogue receptor that stimulates GH secretion at both the hypothalamic and pituitary levels. On the basis of indirect pharmacologic data, it appears that release of GHRH is stimulated by galanin, -aminobutyric acid (GABA), and 2-adrenergic and dopaminergic stimuli and inhibited by somatostatin. Secretion of somatostatin is inhibited by acetylcholine (muscarinic receptors) and 5-HT (type 1D receptors), and increased by 2-adrenergic stimuli and corticotropin-releasing hormone (CRH). ACh, acetylcholine; CNS, central nervous system; DA, dopamine.
Somatostatin and the somatostatin receptor 2 subtype are involved in the short-loop inhibitory feedback of growth hormone (GH) on arcuate neurons. Activation of neurons in the arcuate nucleus was determined by the quantification of immunoreactive c-Fospositive cells after administration of the growth hormone secretagogue MK-0677 (MK). Preliminary treatment of wild-type mice (SSTR2$^{+/+}$) with either GH or the somatostatin analogue octreotide (Octreo) significantly attenuated the neuronal activation by MK-0677. In contrast, GH and octreotide had no effect on MK-0677 neuronal activation in somatostatin receptor 2-deficient mice (SSTR2$^{-/-}$). (Adapted from Zheng H, Bailey A, Jian M-H, et al. Somatostatin receptor subtype 2 knockout mice are refractory to growth hormone-negative feedback on arcuate neurons. Mol Endocrinol 1997; 11:17091717.)
Figure 7-32 Neural pathways involved in growth hormone (GH) regulation. This diagram illustrates the varied pathways by which impulses from the limbic system and brain stem ultimately impinge on the hypothalamic periventricular and arcuate nuclei to stimulate GH release through the mediation of somatostatin (SRIF) and growth hormone-releasing hormone (GHRH). Psychological stress modulates hypothalamic function indirectly through the bed nucleus of the stria terminalis (BNST) and amygdalar complex (Amyg). Circadian rhythms are entrained in part by projections from the suprachiasmatic nucleus (SCN). Complex reciprocal interactions between sleep stage and GHRH release involve cortex and subcortical nuclei, but the detailed mechanisms are not known. Dopaminergic and histaminergic input are from neurons located in the arcuate and mammillary nuclei, respectively, of the hypothalamus (HYP). Ascending catecholaminergic projections arise in both the nucleus of the tractus solitarius (NTS) and ventral lateral medulla (VLM). Serotonergic (5-HT) afferents are from the raphe nuclei. In addition to these neural pathways, a variety of peripheral hormonal and metabolic signals and cytokines influence GH secretion by actions within the medial basal hypothalamus and pituitary gland.
Figure 7-33 Diagram illustrating the genomic organization, messenger ribonucleic acid structure, and post-translational processing of the human somatostatin prohormone. Transcriptional regulation of the somatostatin gene, including the identification of tissue-specific elements (TSE), upstream elements (UE), and the cyclic adenosine monophosphate (cAMP) response element (CRE) that are binding sites for specific factors, has been studied extensively in pancreatic islet cell lines. It is not known whether all or some of these factors are also involved in the neural-specific expression of somatostatin. SST-28 and SST-14 are cyclic peptides containing a single covalent disulfide bond between a pair of Cys residues. A beta turn containing the tetrapeptide Phe-Trp-Lys-Thr is stabilized by hydrogen bonds to produce the core receptor binding epitope. This minimal structure has been the model for conformationally restrained analogues of somatostatin including octreotide. CPE, carboxypeptidase E; PC1/PC2, prohormone convertases 1 and 2; UTR, untranslated region. (Compiled from data by Shen LP, Rutter WJ. Sequence of the human somatostatin 1 gene. Science 1984; 224:168171; Goudet G, Delhalle S, Biemar F, et al. Functional and cooperative interactions between the homeodomain PDX1, Pbx, and Prep1 factors on the somatostatin promoter. J Biol Chem 1999; 274:40674073; and Milner-White EJ. Predicting the biologically active conformations of short polypeptides. Trends Pharmacol Sci 1989; 10:7074.)
Figure 7-34 The use of $^{111}$In-labeled diethyleneetriaminepentaacetic acid (DTPA)-octreotide (radioactive somatostatin analogue) and external imaging techniques to localize a carcinoid tumor expressing somatostatin receptors. Pictures were taken 24 hours after administration of labeled tracer.  
A, Anterior view of the abdomen showing nodular metastases in an enlarged liver and the primary carcinoid tumor (arrow) in the wall of the jejunum of a patient with severe flushing and diarrhea.  
B, Posterior view of the chest and neck showing a metastasis in a lymph node on the left side of the neck (arrow) and multiple metastases in the ribs and pleura.  
Figure 7-35 Regulation of the hypothalamic-pituitary-prolactin (PRL) axis. The predominant effect of the hypothalamus is inhibitory, an effect mediated principally by dopamine secreted by the tuberohypophyseal dopaminergic neuron system. The dopamine neurons are stimulated by acetylcholine (ACh) and glutamate and inhibited by histamine and opioid peptides. One or more prolactin releasing factors (PRFs) probably mediate acute release of PRL as in suckling and stress. There are several candidate PRFs, including thyrotropin-releasing hormone (TRH), vasoactive intestinal polypeptide (VIP), and oxytocin. PRF neurons are activated by serotonin (5-HT). Estrogen sensitizes the pituitary to release PRL, which feeds back on the pituitary to regulate its own secretion (ultrashort-loop feedback) and also influences gonadotropin secretion by suppressing the release of luteinizing hormone-releasing hormone (LHRH). Short-loop feedback is also mediated indirectly by prolactin receptor regulation of hypothalamic dopamine synthesis, secretion, and turnover. CNS, central nervous system; DA, dopamine; GABA, -aminobutyric acid.
**Figure 7-36** Schematic diagram of the gene for pre-pro-gonadotropin-releasing hormone (GnRH) and the GnRH peptide. Diagrams for the enhancer and promoter regions are specific to the rat gene.
Figure 7-37a Regulation of the hypothalamic-pituitary-gonadal axis.  A, Gonadotropin-releasing hormone (GnRH) neurons in a coronal section of the rat hypothalamus at 4× magnification. The inset is at 20× magnification. (Micrograph provided by Patricia Williamson and Kevin Grove, Oregon National Primate Center.)
Figure 7-37b Schematic diagram of the hypothalamic-pituitary-gonadal axis showing neural systems that regulate GnRH secretion and feedback of gonadal steroid hormones at the level of the hypothalamus and pituitary. CRH, corticotropin-releasing hormone; FSH, follicle-stimulating hormone; GABA, -aminobutyric acid; LH, luteinizing hormone; NPY, neuropeptide Y.
Figure 7-39 The influence of gonadotropin-releasing hormone (GnRH) pulse frequency on luteinizing hormone (LH) and follicle-stimulating hormone (FSH) secretion in a female rhesus monkey with an arcuate nucleus lesion ablating endogenous GnRH support of the pituitary. Decreasing GnRH pulse frequency from 1 pulse/hour to 1 pulse/3 hours leads to a decrease in plasma LH concentrations but an increase in plasma FSH concentrations. (Redrawn from Wildt L, Haulser A, Marshall G, et al. Endocrinology 1981; 109:376385.)
Figure 7-40 Pulsatile luteinizing hormone (LH) secretion in an ovariectomized rhesus monkey (A) and an ovariectomized rhesus monkey treated with estradiol (B). Estradiol causes a rapid and sustained suppression of LH secretion. (Redrawn from Yamaji T, Dierschke DJ, Knobil E. Endocrinology 70: 771777.)
Figure 7-41 Dose and duration requirements for estradiol-induced negative and positive feedback on luteinizing hormone (LH) secretion. Varying amounts of estradiol were implanted into ovariectomized rhesus monkeys. Short-term exposure to estradiol led to negative feedback on LH secretion, 36 hours of exposure led to positive feedback in 6 of 11 monkeys, and 42 hours of exposure led to robust positive feedback resulting in a surge of LH secretion in all monkeys. (*Redrawn from Karsch F, Weick RF, Butler WR, et al. Endocrinology 1973; 92:17401747.*)
Figure 7-42 Diagrammatic representation of changes in plasma levels of estradiol, progesterone, luteinizing hormone (LH) and follicle-stimulating hormone (FSH) and portal levels of gonadotropin-releasing hormone (GnRH) over the human menstrual cycle.
Figure 7-43 Regulation of energy homeostasis by the brain-gut-adipose axis. CCK, cholecystokinin; GLP-1, glucagon-like peptide 1; PYY, peptide YY.
Figure 7-44 Structure of the human leptin gene. cDNA, complementary deoxyribonucleic acid; UTR, untranslated region; rectangle, SP-1 site; triangle, CCAAT/enhancer binding protein (C/EBP); circle, cyclic AMP-responsive element (CRE); diamond, glucocorticoid response element (GRE).
Figure 7-45 Role of the hypothalamus as a sensory organ in the regulation of energy homeostasis. CRH, corticotropin-releasing hormone; DA, dopamine; GHRH, growth hormone-releasing hormone; GnRH, gonadotropin-releasing hormone; NPY, neuropeptide Y; POMC, pro-opiomelanocortin; SRIH, somatotropin release-inhibiting hormone; TRH, thyrotropin-releasing hormone. (Modified with permission from Cone RD, Cowley MA, Butler AA, et al. The arcuate nucleus as a conduit for diverse signals relevant to energy homeostasis. Int J Obes 2001; 25:563-567.)
Figure 7-46 Organization of pro-opiomelanocortin (POMC), the precursor hormone of corticotropin (ACTH on figure), -LPH, and related peptides. The precursor protein contains a leader sequence (signal peptide), followed by a long fragment that includes sequence 5162 corresponding to -MSH. This fragment is cleaved at Lys-Arg bonds to form corticotropin 1B39, which in turn includes the sequences for -MSH (corticotropin 1B13) and corticotropin-like intermediate lobe peptide (CLIP) (corticotropin 1739), and a sequence corresponding to -LPH (191) that includes -LPH 158), and -endorphin (61B91). The -endorphin sequence also includes a sequence corresponding to met-enkephalin. The precursor molecule in the anterior lobe of the pituitary is processed predominantly to corticotropin and -LPH. In the intermediate pituitary lobe (in the rat), corticotropin and -LPH are further processed to -MSH and a -endorphinlike material. In all extrapituitary tissues, post-translational processing of the prohormone resembles that in the intermediate lobe. Hypothalamic processing is similar but not identical to that in the intermediate lobe. In the latter, -endorphin and -MSH are present predominantly in their acetylated forms. -LPH, -lipotropin; -MSH, -melanocyte-stimulating hormone; -LPH, -lipotropin; -endorphin, -melanocyte-stimulating hormone; -LPH, -lipotropin.
Figure 7-48 Average plasma ghrelin, insulin, and leptin concentrations during a 24-hour period in 10 human subjects consuming breakfast (B), lunch (L), and dinner (D) at the times indicated (0800, 1200, and 1730 hours, respectively). (Reprinted with permission from Cummings DE, et al. A preprandial rise in plasma ghrelin levels suggest a role in meal initiation in humans. Diabetes 2001; 50:17141719.)
Figure 7-50 A model of the central nervous system circuitry mediating the activation of the paraventricular hypothalamic nucleus (PVN or PVH) and the hypothalamic-pituitary-adrenal (HPA) axis by immune system stimulation. The immune system probably uses several pathways and sites of entry to communicate with the brain. This model predicts that circumventricular organs (organs devoid of blood-brain barrier; CVOs) and the blood vessels (bv) are crucial target sites of cytokines of systemic origin produced during the acute-phase response, whereas activated regions of the brain stem and deep limbic system might play a determinate role in the integration of information received from the periphery. Among these integrative structures, the PVN is critical in coordinating autonomic and endocrine responses including the activity of the HPA axis. For example, corticotropin-releasing factor (CRF) neurons of the parvicellular PVN expressed c-fos messenger ribonucleic acid, and that transcription of the gene coding CRF is activated essentially in this hypothalamic nucleus indicates the importance and the specificity of this neuroendocrine nucleus in endotoxin-treated animals. The mechanisms and the circuitry controlling the CRF release and the activity of the HPA axis might also be different from those involved in the biosynthetic machinery of CRF during immune challenge. ACTH, adrenocorticotropic hormone; AP, area postrema; ARC, arcuate nucleus; BnST, bed nucleus of the stria terminalis; bv, blood vessels; chp, choroid plexus; CeA, central nucleus of the amygdala; COX-2, cyclooxygenase-2; DMH, dorsomedial nucleus of the hypothalamus; EP, prostaglandin E receptor; IL-1, interleukin-1; IL-1R1, IL-1 type 1 receptor; IL-6, interleukin-6; IB, NF-B inhibitor; LC, locus coeruleus; LDT, laterodorsal tegmental nucleus; LPS, lipopolysaccharide; LRNm, lateral reticular nucleus medial; ME, median eminence; MPOA, medial preoptic area; NF-B, nuclear factor B; NTS, nucleus of the solitary tract; OVLT, organum vasculosum of the lamina terminalis; PGE2, prostaglandin E2; PB, parabrachial nucleus; PP, posterior pituitary; PVN, paraventricular nucleus of the hypothalamus (parvicellular [pc] and magnocellular [mc] divisions); SFO, subfornical organ; SON, supraoptic nucleus; TNF-, tumor necrosis factor; VLM, ventrolateral medulla. (Modified from Rivest S, Lacroix S, Vallieres L, et al. How the blood talks to the brain parenchyma and the paraventricular nucleus of the hypothalamus during systemic inflammatory and infectious stimuli. Proc Soc Exp Biol Med 2000; 223:2238.)
Figure 7-51 Immune stimulation activates key brain regions. A series of photomicrographs demonstrating the distribution of Fos-like immunoreactivity (Fos-IR) in the rat brain 2 hours after intravenous injections of lipopolysaccharide (LPS; 125 µg/kg). LPS administration is a commonly used model of immune stimulation, and Fos-IR is a widely used marker of neuronal activation. LPS activates (induces Fos-IR) in the ventral medial preoptic area and organum vasculosum of the lamina terminalis (VMPO and OVLT; A), in the subfornical organ (SFO; B), in the paraventricular nucleus of the hypothalamus (PVH; C), and in the area postrema and nucleus of the solitary tract in the brain stem (AP, NTS; D). Note that prominent Fos-IR is seen throughout the subdivisions of the PVH including the dorsal (dp), ventral (vp), and medial (mp) parvicellular and posterior magnocellular (pm) divisions. Also note that LPS activates neurons in the circumventricular organs (OVLT, SFO, AP). 3v, third ventricle.
**Figure 7-52** Typical pituitary response to thyrotropin-releasing hormone (TRH) administration in patients with hypothalamic-pituitary disease that has caused hypothyroidism. If there is intrinsic pituitary damage, the response is abnormally low. If there is hypothalamic damage, the response is normal or exaggerated. It must be emphasized that some patients with hypothalamic disease may not respond to TRH and that some patients with pituitary disease may respond to TRH. 

T<sub>4</sub>, thyroxine; TSH, thyrotropin. (From Jackson IMD. Diagnostic tests for the evaluation of pituitary tumors. In Jackson IMD, Reichlin S [eds]. The Pituitary Adenoma. New York, Plenum, 1980, pp 219-238.)
Figure 7-53 Effect of hypothalamic-pituitary disconnection on the growth hormone (GH) secretory responses to GH-releasing hormone (GHRH) (1 µg/kg) and hexarelin (2 µg/kg) administered intravenously to children with GH deficiency. Top, mean responses in a group of 24 prepubertal children with short stature secondary to familial short stature or constitutional growth delay. Children with GH deficiency and an intact vascular pituitary stalk as visualized by dynamic magnetic resonance imaging exhibited a clear, but blunted, GH response to both secretagogues (middle). In contrast, children with pituitary stalk agenesis (both vascular and neural components) had no or a markedly attenuated response to both peptides (bottom). (From Maghnie M, Spica-Russotto V, Cappa M, et al. The growth hormone response to hexarelin in patients with different hypothalamic-pituitary abnormalities. J Clin Endocrinol Metab 1998; 83:3886-3889.)
Figure 7-54 A neuroendocrine syndrome of adrenocorticotropic hormone insufficiency, obesity, and red hair resulting from a null mutation in the pro-opiomelanocortin gene. (Photo kindly provided by Dr. A. Gruters, Berlin.)
Figure 8-5 Model for regulation of anterior pituitary hormone secretion by three tiers of control. Hypothalamic hormones traverse the portal system and impinge directly upon their respective target cells. Intrapituitary cytokines and growth factors regulate tropic cell function by paracrine (and autocrine) control. Peripheral hormones exert negative feedback inhibition of respective pituitary trophic hormone synthesis and secretion. (From Ray D, Melmed S. Pituitary cytokine and growth factor expression and action. Endocr Rev 1997; 18:206228.)
Figure 8-7 Magnetic resonance coronal section of a normal pituitary gland (top). A large pituitary adenoma is seen lifting and distorting the optic chiasm (arrow) and is also invading the sphenoid sinus (middle). A sagittal section of a large macroadenoma with bone invasion and impinging brain structures is shown (bottom).
Figure 8-10 Pathogenesis of Rathke's cysts. Schematic of the embryologic progenitors of sellar and parasellar structures. Rathke's pouch arises from an outpocketing of stomodeum (ectoderm) and gives rise to the adenohypophysis. The pharyngohypophyseal stalk, which connects the stomodeum and Rathke's pouch, is divided by the sphenoid bone as it grows together (arrows), isolating Rathke's pouch and the neurohypophysis within the sella. (From Harrison MJ, Morgello S, Post KD. Epithelial cystic lesions of the sellar and parasellar region: a continuum of ectodermal derivatives? J Neurosurg 1994; 80:1018–1025.)
Figure 8-12 Threshold field test showing bitemporal hemianopsia in a patient with pituitary tumor compressing the optic chiasma (A) and superior bitemporal field cuts (B).
Figure 8-13 Transsphenoidal pituitary surgery.  A, B, Route of the transsphenoidal approach (lateral view) and surgical corridor of the transsphenoidal approach and positioning of the retractor. The extent of removal of bone structures is indicated (gray).  C, Parasellar extensions of pituitary adenomas (coronal sections): intrasellar adenoma (a); displacement of the cavernous sinus (b); focal invasion of the cavernous sinus (c); diffuse invasion of the cavernous sinus by the adenoma (d).  D, Extensions of a pituitary (e). Adenoma (sagittal sections): suprasellar extension; invasion of the sphenoid sinus and of the clivus (f).  (Adapted from Honegger J, Buchfelder M, Fahlbusch R. Surgery for pituitary tumors. In Sheaves R, Jenkins PJ, Wass JAH [eds]. Clinical Endocrine Oncology. Cambridge, Mass, Blackwell Science, 1977, p 179.)
Figure 8-14 Transsphenoidal resection of pituitary adenoma. (Modified from BMI Quarterly 6:5, 1990.)
**Figure 8-15** Endoscope-assisted microsurgery provides a panoramic view of the sphenoid sinus. Using a 30-degree endoscope, a view "around the corner" is possible. Parasellar structures can be visualized and residual tumor detected and resected. (From Fahlbusch R, Buchfelder M, Kreutzer J, Nomikos P. Surgical management of acromegaly. In Wass JAH (ed). Handbook of Acromegaly. Bristol, UK, BioScientifica, 2001, p. 46.)
Figure 8-22 Effect of hyperprolactinemia on suppressing follicle-stimulating hormone (FSH) and luteinizing hormone (LH) secretory patterns leading to hypogonadotropism in a female patient. (Adapted from Tolis G. Prolactin: physiology and pathology. Hosp Pract 1980; 15:8595.)
**Figure 8-23** Prolactin levels in 235 patients with galactorrhea of various causes. Triangles denote patients with acromegaly; open circles or triangles denote patients studied after radiotherapy or surgery. *(From Kleinberg DL, et al. Galactorrhea: a study of 235 cases including 48 with pituitary tumors. N Engl J Med 1977, 296:589-600.)*
Figure 8-24 Prolactin-secreting tumors are more often macroadenomas in men (n = 31) than in women (n = 45). Serum prolactin levels highly correlate with tumor size. (Adapted from Danila DC, Klibanski, A. Prolactin secreting pituitary tumors in men. Endocrinologist 2001; 11:105111.)
Figure 8-25  Top, A prolactin-secreting adenoma removed at surgery with no preoperative dopamine agonist therapy.  Bottom, Prolactin-producing pituitary adenoma removed by surgery from a patient treated with dopamine agonist in the preoperative period. The adenoma cells are small with dark nuclei and a narrow rim of cytoplasm. Accumulation of interstitial connective tissue is apparent. (Hematoxylin-eosin stain; original magnification ×400.)  (Courtesy of Kalman Kovacs.)
**Figure 8-26** Comparison of bromocriptine and cabergoline in suppressing prolactin levels in women with hyperprolactinemia. (From Webster J, et al. A comparison of cabergoline and bromocriptine in the treatment of hyperprolactinemic amenorrhea. N Engl J Med 1994; 331:904909.)
Figure 8-27 Prolactinoma management. After secondary causes of hyperprolactinemia have been excluded, subsequent management decisions are based on clinical imaging and biochemical criteria. MRI, magnetic resonance imaging; PRL, prolactin.
Figure 8-28 Shrinkage of macroadenoma by cabergoline in a woman harboring a macroadenoma (A) at 22 weeks of gestation when prolactin (PRL) was 488 µg/L (B) and further reduction at 3 postpartum weeks (C). (From Liu C, Tyrrell JB. Successful treatment of a large macroprolactinoma with cabergoline during pregnancy. Pituitary 2002; 4:3.)
**Figure 8-31** Schematic depiction of the subunit structure and glycosylation sites of the four glycoprotein hormone heterodimers (α subunit, light gray; β subunit, dark gray). (From http://www.chem.gla.ac.uk/protein/glyco/GPH.html.)
Figure 8-32 Effects of pulsatile or continuous administration of gonadotropin-releasing hormone (GnRH) to ovariectomized monkeys rendered GnRH-deficient by placement of a lesion in the hypothalamus. Gonadotropin secretion was restored by hourly GnRH pulses, reduced during a continuous GnRH infusion, and again increased after reinstatement of pulsatile GnRH administration. FSH, follicle-stimulating hormone. (From Belchetz PE, Plant TM, Nakai Y, et al. Hypophyseal responses to continuous and intermittent delivery of hypothalamic gonadotropin-releasing hormone. Science 1978; 202:631633. Copyright 1978 by the American Association for the Advancement of Science.)
Figure 8-33 Serum luteinizing hormone (LH) levels (open symbols) and follicle-stimulating hormone (FSH) levels (solid symbols) in men as a function of age from three studies. (From Tenover JL. Male hormone replacement therapy including "andropause." Endocrinol Metab Clin North Am 1998; 27:969987; Bhasin. In Melmed S (ed). The Pituitary, 2nd ed. Malden, Mass, Blackwell Scientific, 2002.)
Figure 8-35 Management of nonfunctioning pituitary adenomas. Skilled MRI interpretation is crucial to diagnose non-adenomatous mass (e.g., meningioma, aneurysm, or other sellar lesion). MRI, magnetic resonance imaging.
Figure 8-37 Effect of growth hormone (GH) secretagogues on GH, adrenocorticotrophic hormone (ACTH) and prolactin (PRL) secretion in healthy subjects. Mean (+ SEM) curve responses after administration of ghrelin (1.0 µg/kg), hexarelin (HEX) (1.0 µg/kg), growth hormone-releasing hormone (GHRH) (1.0 µg/kg), or placebo. (Adapted from Arvat E, Maccario M, Di Vito L, et al. Endocrine activities of ghrelin, a natural growth hormone secretagogue (GHS), in humans: comparison and interactions with hexarelin, a nonnatural peptidyl GHS, and GH-releasing hormone. J Clin Endocrinol Metab 2001; 86:11691174.)
Figure 8-38 Growth hormone (GH) axis. Simplified diagram of GH/IGF-I axis involving hypophysiotropic hormones controlling pituitary GH release, circulating GH-binding protein and its GH receptor source, IGF-I and its largely GH-dependent binding proteins, and cellular responsiveness to GH and IGF-I interacting with their specific receptors. FFA, free fatty acids. IGFR, insulin-like growth factor I (IGF-I) receptor. Ghrelin, probably of predominantly gastric origin, also stimulates pituitary GH secretion. (From Rosenbloom A. Growth hormone insensitivity: physiologic and genetic basis, phenotype and treatment, J Pediatr 1999; 135:280289.)
Figure 8-39 Effect of fasting on growth hormone (GH) secretion patterns in a healthy male subject. (From Hartman ML, Veldhuis JD, Johnson ML, et al. Augmented growth hormone (GH) secretory burst frequency and amplitude mediate enhanced GH secretion during a two-day fast in normal men. J Clin Endocrinol Metab 1992; 74:757-765.)
Figure 8-40 Integrated model of the growth hormone-insulin-like growth factor binding protein-IGF (GHIGF-BP-IGF) axis in the growth process. Three mechanisms are proposed: (1) GH stimulates IGF-I production, and circulating IGF-I (endocrine IGF-I) acts at the growth plate; (2) GH regulates hepatic production of IGF-BP3 and acid-labile subunit (ALS): IGF-I binds to IGFBP-3 and thereafter with ALS, forming the 150-kd ternary complex; proteases cleave into fragments that release IGFBP3 into fragments that release IGF-I in the intravascular space and at the growth plate; and (3) GH induces differentiation local IGF-I production, and IGF-I acts through an autocrine and paracrine mechanism to stimulate cell division. (From Spagnoli A, Rosenfeld RG. The mechanism by which growth hormone brings about growth: The relative contributions of growth hormone and insulin-like growth factors. Endocrinol Metab Clin North Am 1996; 25:615631.)
Figure 8-41 Growth hormone (GH) receptors. A. Model of GH activation of JAK2 tyrosine kinase. GH binding to two GH receptors increases the affinity to each receptor for JAK2. The two receptor-associated JAK2 molecules are in close proximity, so that each JAK2 can phosphorylate the activating tyrosine of the other JAK2 molecule (blue arrows), thereby activating it. Activated JAK2 then phosphorylates itself (red arrow) and the cytoplasmic domain of the GH receptor (purple arrows) on tyrosines. These phosphotyrosines within the GH receptor and JAK2 form binding sites for signaling proteins. GH, growth hormone; GHR, growth hormone receptor; JAK2, Janus kinase 2; P, phosphate. B. Regulation of GH receptor-JAK2 signaling. SH2 enhances GH receptor signaling by increasing the activity of JAK2. GH-induced expression of SOCS proteins inhibits further GHR signaling by decreasing the activity of JAK2. Tyrosine phosphatases, such as SHP-2, might also contribute to inhibiting GH receptor signaling by dephosphorylating tyrosines in the GH receptor and/or JAK2. GH, growth hormone; GHR, growth hormone receptor; JAK2, Janus kinase 2; P, phosphatase; SHP-2, src homology 2 domain-containing protein tyrosine phosphatase 2; SOCS, suppressor of cytokine signaling. C. GH receptors signaling pathways. Some of the signaling pathways initiated by GH activation of JAK2 are shown. JAK2 phosphorylates SHC, leading to activation of MAPK (blue arrows). JAK2 also phosphorylates STAT transcription factors. MAPK and STATs are important for GH regulation of gene transcription (purple arrows). JAK2 phosphorylates IRS proteins, which are thought to lead to activation of PI 3'-kinase (PI3 K; red arrows). GH activation of PI 3'-kinase via IRS protein might be important for GH stimulation of glucose transport. GH, growth hormone, GHR, growth hormone receptor; IRS, insulin receptor substrates; JAK2, Janus kinase 2; MAPK, mitogen-activated protein kinase; P, phosphate; PI 3K, phosphatidylinositol 3-kinase, STAT, signal transducers and activators of transcription. (AC, From Herrington J, Carter-Su C. Signaling pathway activated by the growth hormone receptor. Trends Endocrinol Metab 2001; 12:252257.)
**Figure 8-42** Individual growth hormonereleasing hormone (GHRH) plus GHRP-6-mediated GH peaks in control subjects (*black dots*) and GH-deficient adults (*white dots*). GH secretion was a continuum between excessive secretion and abnormally low secretion, although a transition concentration between normality and abnormality may be seen at about the 15 µg/L concentration. Logarithmic representation. (*From Popovic V, Leal A, Micic D, et al. GH-releasing hormone and GH-releasing peptide-6 for diagnostic testing in GH-deficient adults. Lancet 2000; 356:11371142.*)
Figure 8-43 Effect of treatment with human growth hormone (hGH) on body fat in eight studies. (Adapted from Newman CB, Kleinberg DL. Adult growth hormone deficiency. Endocrinologist 1998; 8:178186.)
**Figure 8-44** Abdominal subcutaneous and visceral adipose tissue determined with computed tomography at the level of L45 in one man before (top) and after (bottom) 9 months of rhGH treatment. The scan shows the reduction in both visceral and subcutaneous adipose tissue. (Figures and caption kindly provided by B.A. Bengtsson.)
Figure 8-45  Time course of growth hormone (GH) dose and serum insulin-like growth factor I (IGF-I) concentration in a representative patient (38-year-old female) who was switched from oral to transdermal estrogen therapy during the course of GH replacement.  (From Cook DM, Ludlam WH, Cook MB. Route of estrogen administration helps to determine growth hormone [GH] replacement dose in GH-deficient adults. J Clin Endocrinol Metab 1999; 84:3956-3960.)
Figure 8-46 Management of adult somatotropin deficiency. Patients older than 60 years require lower maintenance doses. Women receiving transdermal estrogen require lower doses than those receiving oral estrogen preparations. GH, growth hormone; IGF-I, insulin-like growth factor I; Rx, treatment.
Figure 8-48 Harvey Cushing's first acromegaly patient. (A) Some years before presentation and (B) at admission. (From Jane JA, Laws ER. History of Acromegaly. In Wass JAH (ed). Handbook of acromegaly. Bristol, UK, BioScientifica, 2001, pp 315.)
Figure 8-49 Clinical features of acromegaly. A to C, Features of acromegaly or gigantism in two identical twins. A 22-year-old man with gigantism related to excess growth hormone is shown to the left of his identical twin. The increased height and prognathism (A) and enlarged hand (B) and foot (C) of the affected twin are apparent. Their clinical features began to diverge at the age of approximately 13 years. D to F, Increased incisor spacing and prognathism (D), macroglossia (E), and a normal tongue (F). (AC, From Gagel R, McCutcheon IE. Images in clinical medicine: pituitary gigantism. N Engl J Med 1999; 324:524; DF, from Turner. Clinical features, investigation and complications of acromegaly. In Wass J [ed]. Handbook of Acromegaly. Bristol, UK, BioScientifica, 2001.)
Figure 8-51 Radiation treatment of acromegaly. Long-term effects of radiation therapy on growth hormone (GH) secretion using a GH nadir after an oral glucose load below 2 µg/L as the cure criterion and the probability of not being cured with time after radiotherapy. The numbers of patients not cured at 5, 10, and 20 years after pituitary irradiation are indicated in parentheses. Each step represents one cure; each cross (+) denotes a patient not cured at the latest follow-up. (From Barrande G, Pittino-Lungo M, Coste J, et al. Hormonal and metabolic effects of radiotherapy in acromegaly: long term results in 128 patients followed in a single center. J Clin Endocrinol Metab 2000; 85:37793785.)
Figure 8-53 Medical management of acromegaly. A, Growth hormone (GH) and insulin-like growth factor I (IGF-I) concentrations with long-term octreotide treatment. Comparison of primary octreotide treatment in 25 previously untreated patients and in 80 patients who had previously undergone surgical resection or irradiation, or both. (From Newman C, Melmed S, George A, et al. Octreotide as primary therapy for acromegaly. J Clin Endocrinol Metab 1998; 83:3034-3040.) B, Pharmacodynamics of octreotide LAR twelve-hour mean serum octreotide and GH concentrations in a representative patient treated with a single 30-mg injection of Sandostatin LAR and observed for 60 days. After injection, drug levels peak at 28 days, and nadir GH levels are sustained for 4 weeks. (Adapted from Lancranjan I, Bruns C, Grass P, et al. Sandostatin LAR: a promising therapeutic tool in the management of acromegalic patients. Metabolism 1996; 45:6771.) C, Mean growth hormone (GH) concentration with octreotide (long-acting release) long-term treatment. Serum GH levels in acromegaly following monthly LAR octreotide injections in 12 patients for 1 year, and 8 patients for 31 months. (From Davies PH, Stewart SE, Lancranjan I, et al. Long-term therapy with long-acting octreotide [Sandostatin-LAR] for the management of acromegaly. Clin Endocrinol (Oxf) 1998; 48:311316.) D, Clinical impact of octreotide in reducing soft tissue swelling. Acromegaly in a patient suffering from obstructive sleep apnea before octreotide. Note the macroglossia, tracheotomy for airway obstruction, and intranasal feeding tube. After 6 months of treatment with octreotide, tongue size was reduced by half. Tracheotomy and nasal tube have been removed and sleep apnea has resolved. (Courtesy of S. Reichlin.)
**Figure 8-54** Action of growth hormone (GH) receptor antagonist. A, Normally, a single molecule of GH binds two GH receptors through sites 1 and 2, and the GH signal transduction pathway is activated. B, Pegvisomant increases binding of GH receptor to site 1 and blocks binding at site 2 to prevent functional GH-receptor dimerization, initiation of GH action, and induction of insulin-like growth factor I (IGF-I) synthesis and secretion. The peripheral effects of excess GH are antagonized at the cellular level, independent of the presence of somatostatin or dopamine receptors on the pituitary tumor. (Adapted from van der Lely AJ, Lamberts S. Medical therapy for acromegaly. In Wass J [ed]. Handbook of Acromegaly. Bristol, UK, BioScientifica, 2001, pp 5154.)
Figure 8-59 Structure of pro-opiomelanocortin (POMC) gene. Exon 1 encodes the RNA leader sequence, exon 2 encodes the initiator methionine (ATG), the signal peptide and several N-terminal residues of the precursor peptide, the remainder of which is encoded by exon 3. Corticotroph expression is determined by the upstream "pituitary" promoter (longer white arrowhead), whereas peripheral expression of the short POMC mRNA is determined by the "downstream" promoter (shorter white arrowhead). Translation of these shorter transcripts initiates from the initiator methionines (ATG) indicated in exon 3. The precursor peptide coding region is shaded light gray and the ACTH coding region is black. (From Clark AJL, Swords FM. Molecular pathology of corticotroph function. In Rappaport R, Amselem S [eds]. Hypothalamic-Pituitary Development: Genetic and Clinical Aspects. Basel, Karger, 2001, pp 140161.)
Figure 8-60 Processing and cleavage of pro-opiomelanocortin (POMC). The mature POMC precursor peptide is sequentially cleaved by PC-1 in the anterior pituitary corticotroph. In the neurointermediate lobe and other cell types, cleavage by PC2 allows release of -MSH and/or -endorphin. Carboxypeptidase H (not shown) removes residual basic amino acids at cleavage sites. -LPH, -lipotropin; JP, joining peptide; -LPH, -lipotropin, CLIP, corticotropin-like intermediate lobe peptide. (From Clark AJL, Swords FM. Molecular pathology of corticotroph function. In Rappaport R, Amselem S [eds]. Hypothalamic-Pituitary Development: Genetic and Clinical Aspects. Basel, Karger, 2001, pp 140161.)
Figure 8-63 Management of thyroid-stimulating hormone (TSH)-secreting pituitary tumors. PTU, propylthiouracil; T₃, triiodothyronine; T₄, thyroxine; TRH, thyrotropin-releasing hormone.
Figure 8-64 Life-table analysis indicating probabilities of initially normal hypothalamic-pituitary-target gland axes remaining normal after radiotherapy (3750 to 4250 cGy). Growth hormone (GH) secretion is the most sensitive of the anterior pituitary hormones to the effects of external radiotherapy, and TSH secretion is the most resistant. In two thirds of patients, gonadotropin deficiency develops before adrenocorticotropic hormone (ACTH) deficiency. The reverse occurs in the remaining third. (From Littley MD, Shalet SM, Beardwell CG, et al. Hypopituitarism following external radiotherapy for pituitary tumors in adults. Q J Med 1989; 70:145160.)
Figure 8-65 Incidence of growth hormone (GH) deficiency in children receiving 27 to 32 Gy or 35 Gy of cranial irradiation for a brain tumor in relation to time from irradiation (dxt). This illustrates that the speed at which individual pituitary hormone deficits develop is dose-dependent; the higher the radiation dose, the earlier GH deficiency occurs. (Courtesy of the Department of Medical Illustrations, Wilkington Hospital, Manchester, England. From Shalet S. Pituitary failure. In DeGroot LJ, Jameson JL [eds]. Endocrinology, 4th ed. Philadelphia, WB Saunders, 2001.)
Figure 9-1 Magnicellular neurons and axon tracts in the hypothalamus. A, Coronal section of the hypothalamus in which neurophysin antibodies are used in an immunoperoxidase technique to demonstrate the magnicellular neurons and axons. The paired paraventricular nuclei (PVN) are at the top in the walls of the third ventricle (V). The axons course laterally and ventrally to the supraoptic nucleus (SON) at the lateral extremes of the optic chiasm (OC) and then inferiorly to the median eminence. (Adapted from Zimmerman EA, Anatomy of vasopressin in producing cells. In Czernichow P, Robinson AG [eds]. Diabetes Insipidus in Man. Basel, Karger, 1985, pp 121. Modified to add right SON by A. G. Robinson, University of California, Los Angeles.) B, In the median eminence, the axons from the two sides coalesce to form the supraopticohypophyseal tract that descends through the stalk to the axon terminals in the posterior pituitary. (Adapted from Robinson AG, Zimmerman EA. Cerebrospinal fluid and ependymal neurophysin. J Clin Invest 1973; 52:12601267. Modified to remove tissue tear by A. G. Robinson, University of California, Los Angeles.)
Figure 9-2 Comparison of the chemical structures of arginine vasopressin, oxytocin, and desmopressin. The differences are illustrated by the shaded areas. Oxytocin differs from vasopressin in position 3 (Ile for Phe) and position 8 (Leu for Arg). Desmopressin differs from arginine vasopressin in that the terminal cystine is deaminated and the arginine in position 8 is a d rather than an l isomer. (© 2003, UCLA, AG Robinson.)
Figure 9-3 Vasopressin synthesis in a magnicellular neuron.  A. The vasopressin gene is located on the short arm of chromosome 20. In the nucleus, the gene is transcribed to heteronuclear ribonucleic acid (RNA).  B. The introns are then excised, and the three exons are spliced to form mature RNA, which consists of exon A, exon B, and exon C. The mature RNA exits the nucleus to the cytoplasm. It is targeted to the endoplasmic reticulum, where it is attached to ribosomes.  C and D. There is translation of the three exons into pre-provasopressin. Exon A is translated to the 19-amino-acid signal peptide (SP), the nonapeptide arginine vasopressin (AVP), and the amino-terminal portion of the 93- to 95-amino-acid neurophysin (NP). Exon B encodes the highly conserved middle region of neurophysin. Exon C encodes the variable carboxyl terminus of neurophysin and a 39-amino-acid glycopeptide (GP). The pre-provasopressin is transferred across the endoplasmic reticulum, the glycopeptide is glycosylated, and the signal peptide is cleaved (D).  E. Provasopressin enters the Golgi apparatus, where the entire provasopressin complex is packaged into neurosecretory granules. The neurosecretory granules attach to microtubules and are transported along the microtubules to the posterior pituitary, where the neurosecretory granules are stored.  F. During transport, enzymes in the acidic granules cleave the prohormone to vasopressin (which is amidated), to neurophysin, and to the glycopeptide. Neurophysins form dimers and, subsequently, tetramers with one vasopressin attached to each neurophysin. There is an auxiliary fifth binding site for vasopressin, which spans the four neurophysin molecules of the tetramer.  G. When there is an action potential signaling release, a neurosecretory granule fuses with the axon membrane and the vasopressin, neurophysin, and glycopeptide are secreted into the extracellular space and, hence, into plasma, where they circulate independently of each other.  (© 2003, UCLA, AG Robinson.)
**Figure 9-4** Comparison in humans of the release of vasopressin in response to percentage changes of osmolality (increase) and pressure or volume (decrease). Note: To increase plasma vasopressin, the change in osmolality is much more sensitive, responding to as little as a 1% increase in osmolality, whereas volume and pressure require greater than a 10% to 15% change to stimulate release of vasopressin. (Redrawn from Robertson GL, Berl T. Water metabolism. In Brenner BM, Rector FC Jr [eds]. The Kidney, vol 1, 3rd ed. Philadelphia, WB Saunders, 1986, p 385.)
Figure 9-5 Relationship of plasma osmolality (mOsm/kg of H₂O) to plasma vasopressin (pg/mL) to urine osmolality (mOsm/kg of H₂O) to urine volume (L/day). AVP, arginine vasopressin; P, plasma; U, urine. A, Small changes in osmolality induce changes in vasopressin from less than 0.5 to 5 to 6 pg/mL. B, These small changes in plasma vasopressin induce changes in urine osmolality through the full range from maximally dilute to maximally concentrated urine. Plasma vasopressin can rise to higher levels than 6 pg/mL, as illustrated in A, but this does not translate into increased urine osmolality, which has a maximum determined by the osmolality of an inner medulla of the kidney. C, The relationship of volume to urine osmolality is logarithmic, assuming a constant osmolar load and the urine volume that would excrete that osmolar load at the urine osmolality indicated. The interrelationship between the three graphs is illustrated by the shaded area that represents the normal range. Urine volume changes relatively little with small changes in the other parameters until there is nearly complete absence of vasopressin, and the urine volume then increases dramatically. (Calculated from formulas in Robertson GL, Shelton RL, Athar S. The osmoregulation of vasopressin. Kidney Int 1976; 10:2537.) (© 2003, UCLA, AG Robinson.)
Figure 9-6 Comparison of the normal response of plasma vasopressin to plasma osmolality (A) with that in midgestation (B) and late gestation (C). During midgestation (B), there is a reset of the osmostat to a lower level, shifting the curve to the left. In late gestation, the osmotic threshold remains the same but there is a change in the slope and less vasopressin is secreted. (Adapted from Davison JM, Shields EA, Phillips PR, Lindheimer MD. Serial evaluation of vasopressin release and thirst in human pregnancy: role of human chorionic gonadotrophin in the osmoregulatory changes of gestation. J Clin Invest 1988; 81:798806.)
Figure 9-7 Typical triphasic response of urine volume after section of the pituitary stalk induced by surgery or head trauma. The first phase of diabetes insipidus occurs immediately after operation and continues to day 6. The second phase of antidiuresis occurs from day 7 and continues to day 12. The third stage is the recurrence of diabetes insipidus on day 13. (Durations vary; see text for discussion.)
Figure 9-9 Plasma arginine vasopressin (AVP) levels in patients with syndrome of inappropriate antidiuretic hormone secretion as a function of plasma osmolality. Each point depicts one patient at a single point in time. The shaded area represents AVP levels in normal subjects over physiologic ranges of plasma osmolality. The lowest measurable plasma AVP level using this radioimmunoassay was 0.5 pg/mL. (From Robertson GL, Aycinena P, Zerbe RL. Neurogenic disorders of osmoregulation. Am J Med 1982; 72:339-353.)
Figure 9-10 Schematic summary of different patterns of arginine vasopressin (AVP) secretion in patients with syndrome of inappropriate antidiuretic hormone secretion. Each line (a to d) represents the relation between plasma AVP and plasma osmolality of individual patients in whom osmolality was increased by infusion of hypertonic NaCl. The shaded area represents plasma AVP levels in normal subjects over physiologic ranges of plasma osmolality. (From Robertson GL. Thirst and vasopressin function in normal and disordered states of water balance. J Lab Clin Med 1983; 101:351371.)
**Figure 10-1** Structure of thyroid hormone and related compounds. The thyronine nucleus, the precursor iodinated amino acids, and the secreted hormones, thyroxine (T₄) and triiodothyronine (T₃). Iodinated thyronines are formed by the oxidative coupling of the precursor iodotyrosines monoiodotyrosine (MIT) and diiodotyrosine (DIT) in the thyroglobulin molecule.
Figure 10-2 Schematic illustration of a thyroid follicular cell showing the key aspects of thyroid iodine transport and thyroid hormone synthesis. AMP, adenosine monophosphate; cAMP, cyclic AMP; DIT, diiodotyrosine; MIT, monoiodotyrosine; NIS, sodium iodide symporter; T₃, triiodothyronine; T₄, thyroxine; Tg, thyroglobulin; TPO, thyroid peroxidase; TSH, thyrotropin; TSHR, thyrotropin receptor. (Modified from Spitzweg C, Heufelder AE, Morris JC. Thyroid iodine transport. Thyroid 2000; 10:321330.)
Figure 10-3 Pathways for thyroid hormone activation and inactivation catalyzed by human iodothyronine selenodeiodinases. Numbers refer to the iodine positions in the iodothyronine nucleus. The iodothyronine deiodinases are abbreviated D1, D2, and D3 for types 1, 2, and 3 deiodinases, respectively. Arrows refer to monodeiodination of the outer or inner ring of the iodothyronine nucleus, termed 5' or 5 by convention. The parentheses around D1 emphasize that D3, not D1, is probably the major enzyme catalyzing inner ring deiodination of T_4 and T_3.
Figure 10-5  Schematic diagram of thyroxine (T₄) 5'-deiodination reaction as catalyzed by the type 1 iodothyronine deiodinase (D1). The reaction assumes the formation of a selenyl-iodide intermediate, which requires a cystolic cofactor that is likely to be an -SH compound, such as reduced glutathione (GSH). Heavy metals with a single positive charge, such as gold (Au⁺), inhibit deiodination by interaction with the negatively charged selenium atom. Propylthiouracil (PTU) is thought to form a relatively stable Se-S complex, thereby blocking regeneration of the active enzyme. (Modified from Leonard JL, Visser TJ. Biochemistry of deiodination. In Hennemann G [ed]. Thyroid Hormone Metabolism. New York, Marcel Dekker, 1986, pp 189-229.)
Figure 10-6  Schematic diagram of the origin of the specifically bound nuclear triiodothyronine (T₃) in various rat tissues. Data are derived from studies in which double-isotope labeling techniques were used to estimate the sources of specifically bound nuclear T₃. In tissues having a receptor saturation significantly greater than 50%, the additional T₃ is provided by D₂-catalyzed T₄ to T₃ conversion. T₃ in rat plasma is derived from thyroid secretion (40%) with the remainder from D₁- and D₂-catalyzed T₂ to T₃ conversion. BAT, brown adipose tissue; PIT, pituitary gland.
Figure 10-7 Comparison of the chemical structure of thyroxine (T₄) with the structures of two agents that block the deiodination of the iodothyronines. The inhibition of T₄ to triiodothyronine (T₃) conversion, which occurs in patients receiving amiodarone, may be due to the drug itself or to a metabolic product. Iopanoic acid and related iodoanilines are competitive inhibitors of all three iodothyronine deiodinases.
Figure 10-8 Schematic diagram of thyroid hormone activation and inactivation in a cell expressing D2 and D3, such as an astroglial cell or a neuron. The triiodothyronine (T₃) that enters the cell can either be deiodinated to 3,3′ T₂ (diiodothyronine) or can enter the nucleus and bind to the thyroid hormone receptor. An additional source of T₃ is that generated by outer ring deiodination of thyroxine (T₄) within the cell. The interaction of T₃ with the thyroid hormone receptor (TR) bound as a heterodimer with a retinoid X receptor (RXR) to the thyroid hormone response element (TRE), causes either an increase or a decrease in the transcription of that gene. This leads to parallel changes in the concentrations of critical proteins, thus producing the thyroid hormone response of a given cell. mRNA, messenger RNA; rT₃, reverse T₃. (See Chapter 4 for specific details.)
Figure 10-9 The functional domains in the thyroid hormone receptor are similar in both α and isoforms of the thyroid hormone receptors. These isoforms differ primarily in their NH₂-terminal domains. AF, activating factor; DBD, DNA-binding domain; T₃, triiodothyronine.
Figure 10-10 Role of thyroxine and triiodothyronine (T\(_4\) and T\(_3\)) in the feedback regulation of thyrotropin-releasing hormone (TRH) and thyrotropin (TSH) secretion. Secreted T\(_4\) must be converted to T\(_3\) to produce its effects. This conversion may take place in tissues such as the liver (L) and kidney (K) and thyroid (T) catalyzed by D1 or D2 in the human thyroid gland (T), skeletal muscle (SM), and, possibly, cardiac muscle (CM). SRIH, somatostatin.
Figure 10-11 The log/linear relationship between thyrotropin (TSH) (vertical axis) and free thyroxine (FT₄) concentrations. Typical free T₄ concentrations in hypothyroid, euthyroid, and hyperthyroid patients are shown. (Modified from Spencer CA, LoPresti JS, Patel A, et al. Applications of a new chemiluminometric thyrotropin assay to subnormal measurement. J Clin Endocrinol Metab 1990; 70:453460.)
Figure 10-12 Effects of acute depletion of dietary iodine on serum triiodothyronine (T₃), thyroxine (T₄), and thyrotropin (TSH) in rats. Animals received a low-iodine diet (LID) without or with supplementation of potassium iodide (KI) in drinking water. (From Riesco G, Taurog A, Larsen PR, Krulich L. Acute and chronic responses to iodine deficiency in rats. Endocrinology 1977; 100:303313. © The Endocrine Society.)
**Figure 10-13** Newborn infant with iodide-induced goiter due to Lugol's solution treatment of the mother during the third trimester. This illustration shows the danger of chronic excess iodide administration during gestation.
Figure 10-14 Changes in various critical components of the thyroid-pituitary axis during pregnancy. Note the early increase in free thyroxine (T₄), probably due to thyroidal stimulation by human chorionic gonadotropin (hCG), which causes a reciprocal modest suppression of serum thyrotropin (TSH) during the late first trimester. TBG, thyroxine-binding globulin. (From Burrow GN, Fisher DA, Larsen PR. Mechanisms of disease: maternal and fetal thyroid function. N Engl J Med 1994; 331:10721078. © 1994, Massachusetts Medical Society.)
Figure 10-15 With the development of modern thyrotropin (TSH) assays with greater sensitivity, it is possible to distinguish between normal thyrotropin concentrations and the suppressed values of hyperthyroidism. Each generation of assays represents a 10-fold improvement in functional sensitivity, that is, 20% interassay coefficient of variation (CV) value. Black bars denote the 95% confidence limits of measurement at different thyrotropin concentrations. (From Nicoloff JT, Spencer CA. Clinical review 12: the use and misuse of the sensitive thyrotropin assays. J Clin Endocrinol Metab 1990; 71:553558.)
Figure 10-16 Pattern of changes in total serum thyroxine (T4) concentrations and thyroid hormone-binding ratio (THBR) in euthyroid patients with alterations in circulating concentrations of thyroxine-binding globulin (TBG). To convert T4 from nmol/L to µg/dL (total) or pmol/L (free), divide by 12.87.
Figure 10-17 Pattern of changes in total serum thyroxine (T₄) concentration and thyroid hormone-binding ratio (THBR) in patients with hyperthyroidism or hypothyroidism with normal serum thyroxine-binding globulin (TBG) concentration.
Figure 10-18 Examination of the thyroid gland. A. Sagittal section demonstrates relations of the isthmus of the normal thyroid gland. The superior border is inferior to the cricoid cartilage. The inferior thyroid border is essentially at the level of the superior surface of the manubrium. The inferior portions of the lateral lobes (not shown) extend more inferiorly than the isthmus. B. The cricoid cartilage is regarded as an important landmark. Especially when the thyroid gland is thought to be essentially normal or subnormal in size, the cricoid should be located. This is easily accomplished. The index fingers are then inserted so that their superior portion rests against the inferior portion of the cricoid while the inferior portion of these fingers is over the superior portion of the thyroid. The second and third fingers are rotated over other portions of the gland to evaluate its size, contour, consistency, possible adherence to surrounding structures, and other features. Because there is marked variation among different subjects in the length and thickness of the neck and in the length of the trachea superior to the level of the manubrium, the relative position of the thyroid may vary. In some cases, essentially all of the thyroid gland rests posterior to the sternum. In most instances, however, by having the patient moderately extend the neck (short of tightening the anterior neck muscles) and swallow repeatedly, it is possible to palpate most or all of the gland. Despite marked variations in neck-chest relations, thyroid tissue, when present, is found within 1 cm of the cricoid. By concentrating the palpation meticulously in the area where the thyroid is normally found, with rare exceptions the examiner can outline small as well as enlarged glands.
Figure 11-1  Inhibition of the binding of $^{125}\text{I}$-labeled thyrotropin (TSH) to human thyroid TSH receptors by increasing concentrations of bovine TSH (†) and by increasing concentrations of immunoglobulin G (IgG)-containing TSH receptor antibodies referred to here as TSH-binding inhibitory immunoglobulin (TBII) activity (and . (From Endo K, Kashagi K, Konishi J, et al. Detection and properties of TSH-binding inhibitor immunoglobulins in patients with Graves’ disease and Hashimoto’s thyroiditis. J Clin Endocrinol Metab 1978; 46:734739. © 1978, The Endocrine Society.)
Figure 11-2 Stimulation of adenylate cyclase activity in human thyroid membranes by serum immunoglobulin G in normal control subjects and patients with thyroid disease. (From Bech K, Nishup Madsen SN. Thyroid adenylate cyclase stimulating immunoglobulins in thyroid disease. Clin Endocrinol 1979; 11:4758.)
Figure 11-3 Schematic diagram of thyroid cell stimulation and blockade by antibodies to the thyrotropin-stimulating hormone receptor. Such autoantibodies may act as agonists or antagonists, depending on how they interact within the extracellular domain.
Figure 11-4 Section of thyroid gland of four patients with Graves' disease. A, Untreated. B, After therapy with potassium iodide for 3 weeks. C, After treatment with thiouracil for 5 weeks. D, Three months after three treatments with radioiodine. Note the marked hypertrophy and hyperplasia of the acinar cells and scant amount of colloid in sections A, C, and D. A lymph follicle is present in C. Note the broad bands of scar tissue in D. Section B is almost normal in appearance. Each patient, except the first one, was euthyroid at the time of thyroidectomy.
Figure 11-6  Human thyrotropin receptor (hTSH-R) exon structure. Outline structure of the TSHR (A) in comparison with the porcine luteinizing hormone receptor (pLH-R) and rat follicle-stimulating hormone receptor (rFSH-R) genes and (B) the exon/intron organization of the TSHR gene. A shows the similarity (%) between the hTSHR and the pLH and rFSH receptors, respectively. (From Gross B, Misrahi M, Sar S, et al. Composite structure of the human TSH receptor gene. Biochem Biophys Res Comm 1991; 177:679687.)
Figure 11-7 A current model of the human thyrotropin-stimulating hormone receptor structure. The TSHR has seven transmembrane domains, a large extracellular domain, and a small intracellular domain. The receptor is cleaved, probably after activation, into A (or A) and B (or B) subunits. The subunit is thought to be shed from the cell surface.
Figure 11-8 T cells from a patient with Graves' disease demonstrate a dose-related proliferative response to a thyrotropin-stimulating hormone receptor peptide (aa 181200). Data shown as $^3$H-thymidine uptake at 18 hours.
Figure 11-9 Schematic diagram of a T-cell receptor dimer showing the alpha and beta chains retained by transmembrane regions. V, variable region; C, constant region. (From Davis MM, Chien YH, Gascoigne NR, et al. A murine T cell receptor gene complex. Immunol Rev 1984; 81:235258.)
**Figure 11-10** Possible mechanisms involved in the cause and precipitation of Graves’ disease. MHC, major histocompatibility complex.
Figure 11-11 Photomicrograph of Graves' thyroid tissue stained for HLA class II (DR) antigen expression using the immunoperoxidase technique. Note the brown thyroid epithelial cells indicating the presence of DR antigen. Note also the relative lack of lymphocytic infiltration in this region.
Figure 11-12 CT scans of orbits in two patients with Graves’ orbitopathy. A, Note the obviously grossly swollen medial rectus extraocular muscles in both orbits and the resulting proptosis. B, The patient shows considerable proptosis with only minimal muscle enlargement, suggesting the presence of a large amount of retro-orbital fat. (Courtesy of Dr. Peter Som, New York, NY.)
**Figure 11-13** Section of extraocular muscle from a biopsy taken from a patient with severe Graves’ orbitopathy. Note that within the muscle fibers is a patch of lymphocytic infiltration. (Courtesy of Dr. D. Kendler, University of British Columbia, Vancouver, Canada.)
Figure 11-14 Chronic pretibial myxedema in a patient with Graves’ disease and orbitopathy. The lesions are firm and nonpitting, with a clear edge to feel. (Courtesy of Dr. Andrew Werner, New York, N.Y.)
Figure 11-15 Characteristic signs of Graves’ orbitopathy (A) subsequently corrected by orbital decompression surgery (B). Note the thyroid stare, the asymmetry, the proptosis, and the periorbital edema prior to correction. (Courtesy of Dr. Jack Rootman, University of British Colombia, Vancouver, Canada.)
Figure 11-16 Graves' orbitopathy. A, Palpebral edema. This patient's eyes protruded anteriorly 1 cm more than normal, but there is no "popeye" appearance, owing to edema of the surrounding structures. B, Marked widening of palpebral fissures and slight palpebral swelling. C, Unequal degrees of ophthalmopathy. D, Unilateral lid retraction. E, Palpebral swelling, presumably because of fat pads and edema, and paralysis of the right external rectus muscle. F, Marked conjunctival injection and chemosis, together with ophthalmoplegia. G, Failure to close lid on the right because of marked exophthalmos, corneal scarring, and panophthalmitis; the eye had to be enucleated.
Figure 11-17 Ophthalmoplegia in Graves' disease. Other than slight conjunctival injection, the only ocular abnormality was paralysis of upward gaze on the right in this woman with severe Graves' disease.
Figure 11-18 Rare thyroid acropachy in a patient with Graves' disease. The hypermetabolic state leads to axial bone destruction, presumably secondary to enhanced osteoclast activity. Acropachy is not to be confused with clubbing, which is usually painless. (Courtesy of Dr. Andrew Werner, New York, N.Y.)
**Figure 11-19** Effects of antithyroid agents on the serum levels of triiodothyronine (T\(_3\)) and thyroxine (T\(_4\)) in patients with Graves' disease. The **left panels** show the effects of potassium iodide (SSKI, 5 drops every 8 hours). A rapid reduction in T\(_3\) concentration occurs in all patients over the first 5 days of therapy. Methimazole (MMI) at the indicated doses has a variable effect on serum T\(_3\) concentrations. In one patient the serum T\(_3\) level falls rapidly over the first 3 days, whereas in the other two individuals, despite an even larger dosage, there is no change. Serum T\(_4\) concentration does not change significantly over this time interval. The **right panels** show that the administration of high-dose propylthiouracil (PTU) causes a marked decrease in serum T\(_3\) concentrations to one third to one half of initial levels. This decrease is due to the PTU-induced inhibition of type 1 iodothyronine 5'-deiodinase. *(Data from Abuid J, Larsen PR. Triiodothyronine and thyroxine in thyrotoxicosis: acute response to therapy with antithyroid agents. J Clin Invest 1974; 39:263-268.)*
Figure 11-20 Influence of carbimazole and placebo treatment on thyroid peroxidase antibody levels (measured as antimicrosomal antibodies) in patients with Hashimoto's thyroiditis. These data illustrate the immunosuppressive effect of carbimazole in patients who remained euthyroid throughout observation. SEM, standard error of the mean. (From McGregor AM, Ibbertson HK, Smith BR, Hall R. Carbimazole and autoantibody synthesis in Hashimoto's thyroiditis. Br Med J 1980; 281:968-970.)
Figure 11-21 Incidence of postradioiodine hypothyroidism in relation to the duration of follow-up. The total number of patients followed for each of the indicated time periods is shown in parentheses. (From Dunn JT, Chapman EM. Rising incidence of hypothyroidism after radioactive iodine therapy in thyrotoxicosis. Reprinted by permission of The New England Journal of Medicine, 1964; 271:10371042.)
Figure 11-22 Probability of the development of worsening of orbitopathy in patients with Graves’ disease. The serum triiodothyronine (T₃) levels are shown before treatment, and the type of therapy is shown as a variable. (From Tallestedt L, Lundell G, Torring O, et al. Occurrence of ophthalmopathy after treatment for Graves’ disease. N Engl J Med 1992; 326:17331738. Copyright © 1992, Massachusetts Medical Society. All rights reserved.)
**Figure 11-24** $^{131}$I scanning of a hyperfunctioning hot nodule corresponding to physical examination with a faint outline of the remaining suppressed gland. In this unusual case, Graves’ disease developed a few months later after an oral contrast agent load. *(From Soule J, Mayfield R. Graves' disease after $^{131}$I therapy for toxic nodule. Thyroid 2001; 11:9192.)*
Figure 11-25 Subacute thyroiditis. Intrafollicular giant cell surrounding a central core of colloid. (From Meachim G, Young MH. De Quervain’s subacute granulomatous thyroiditis: histological identification and incidence. J Clin Pathol 1963; 16:189199.)
Figure 11-26 Thyroid function in a patient in the course of de Quervain's (subacute) thyroiditis. During the thyrotoxic phase (days 10 to 20) the serum thyroglobulin (TG) concentration was elevated, the free thyroxine index (FTI) was high, and thyrotropin (TSH) was suppressed. The erythrocyte sedimentation rate was 86 mm/hour, and the thyroidal radioactive iodine uptake was 2%. The Tg level and FTI declined in parallel. During the phase of hypothyroidism (days 30 to 63), when the FTI was below normal, the serum Tg level transiently increased in parallel with the increase in serum TSH. All parameters of thyroid function were normal by day 150, 5 months after the onset of symptoms. (From DeGroot LJ, Larsen PR, Hennemann G [eds]: Acute and subacute thyroiditis. In The Thyroid and Its Diseases, 6th ed. New York, Churchill Livingstone, 1996, p 705.)
**Figure 11-27** Low-power (A) and high-power (B) magnification of a thyroid gland biopsy during the hypothyroid phase of "silent thyroiditis." Note the extensive lymphocytic infiltration and patchy distribution of poorly preserved follicles. (From Woolf PD. Transient painless thyroiditis with hyperthyroidism: a variant of lymphocytic thyroiditis? Endocr Rev 1980; 1:411420. © 1980, The Endocrine Society.)
Figure 11-28 Schematic of the typical course in chronic thyroiditis with transient thyrotoxicosis. The duration of each phase may vary, and some patients do not experience a discernible hyperthyroid or hypothyroid phase. (From Woolf PD. Transient painless thyroiditis with hyperthyroidism: a variant of lymphocytic thyroiditis? Endocr Rev 1980; 1:411-420. © 1980, The Endocrine Society.)
Figure 11-29 Changes in free thyroxine in various types of postpartum thyroid disease. Numbers indicate the number of months post partum. (Adapted from Davies TF [ed]. Autoimmune Endocrine Disease. New York, John Wiley & Sons, 1963, p 255.)
Figure 11-30 The structure of amiodarone illustrating the characteristics of thyroid hormone. The molecule is 37% iodine by weight.
Figure 12-1 A and B, Typical appearance of patients with moderately severe primary hypothyroidism or myxedema. Note dry skin, sallow complexion, with the absence of scleral pigmentation differentiating the carotenemia from jaundice. Both individuals demonstrate periorbital myxedema. The patient in B illustrates the loss of the lateral aspect of the eyebrow, sometimes termed Queen Anne's sign. That finding is not unusual in the age group that is commonly affected by severe hypothyroidism and should not be considered to be a specific sign of the condition.
Figure 12-2 A and B. Chest roentgenograms in a patient with myxedema heart disease. The patient had signs of severe congestive heart failure and was given thyroid hormone alone. Within 4 months, the heart had returned to normal size and there was no evidence of underlying heart disease.
Figure 12-3 The consequences of untreated congenital hypothyroidism are demonstrated in this 17-year-old girl. Her condition had been diagnosed at birth but, through a series of misunderstandings, was not treated with thyroid hormone. Note her size, the poorly developed nasal bridge, the wide-set eyes, and the ears, which are larger than are appropriate for head size. Her tongue is enlarged, and her extremities are inappropriately short in relation to her trunk. (Courtesy of Dr. Ronald B. Stein.)
Figure 12-4 X-ray films of the skull and hand of the 17-year-old patient illustrated in Figure 12-3. A, Skull film showing that the posterior and anterior fontanelles are open and that the sutures are not fused. The deciduous and permanent teeth are present. B, Radiograph of the wrist and hand showing the delayed appearance of the epiphyseal centers of the bones of the hand and the absence of the distal radial epiphysis. The estimated bone age is 9 months. (Courtesy of Dr. Ronald B. Stein.)
Figure 12-5 Frequency of hypothyroid symptoms and signs (percentage) in 50 patients with overt hypothyroidism and in 80 euthyroid controls. Two symptoms (pulse rate and cold intolerance, marked by asterisks) showed positive and negative predictive values of less than 70% and were thus excluded from the new score. (From Zulewski H, Müller B, Exer P, et al. Estimation of tissue hypothyroidism by a new clinical score: evaluation of patients with various grades of hypothyroidism and controls. J Clin Endocrinol Metab 1997; 82:771776.)
Figure 12-6 Strategy for the laboratory evaluation of patients with suspected hypothyroidism. The principal differential diagnosis is between primary and central hypothyroidism (see Chapter 8). The serum thyrotropin (TSH) concentration is the critical laboratory determination that, in general, allows recognition of the cause of the disease. An exception is the individual with a recent history of thyrotoxicosis (and suppressed TSH) in whom a low free thyroxine (T₄) level may be associated with a reduced TSH level for several months after relief of the thyrotoxicosis. In patients with primary hypothyroidism, the absence of thyroid peroxidase (TPO) antibodies raises a possible diagnosis of transient hypothyroidism following an undiagnosed episode of subacute or postviral thyroiditis. In such patients, a trial of a reduced levothyroxine dosage after 4 months may reveal recovery of thyroid function thus avoiding permanent levothyroxine replacement. MRI, magnetic resonance imaging.
Figure 12-7 Possible involvement of Fas-FasL in the apoptosis of Hashimoto's thyroiditis. In Graves' disease, the thyroid follicles thrive because the thyroid cells do not express many functional Fas molecules (shown in red) and, therefore, the thyroid cells are resistant to the FasL (Fas ligand) on both the thyroid cells themselves and on the Th2 T cells (shown in blue). Further apoptotic resistance may be driven by Th2 cytokines, such as interleukin-4 (IL-4) and IL-10. However, the Th2 cells express Fas and may themselves be deleted by the FasL constitutively expressed on the thyroid cells. The result is thyroid cell survival with T-cell destruction. In contrast, in Hashimoto's disease, the thyroid cell expresses many functional Fas molecules, perhaps induced by gamma interferon from the Th1 type T cells associated with this disease. The expression of thyroid cell Fas may lead to self (homophilic) apoptosis via thyroid cell FasL, or apoptosis may result from attack by the Fas1 armed Th1 cells. The result is thyroid follicle destruction and T-cell proliferation.
Figure 12-8 A T-cell clone (G) from a patient with Hashimoto's disease was able to specifically lyse autologous thyroid cells. Data are shown as percent cytotoxicity from a radioactive chromium release assay. Note that clone C was able to lyse all cell targets and was, therefore, not thyroid cellspecific. (From McKenzie WA, Davies TF. An intrathyroid T-cell clone specifically cytotoxic for human thyroid cells. Immunology 1987; 61:101103).
Figure 12-9 Hashimoto's disease. A. Note the exaggeration of the normal lobular pattern. B. Interfollicular infiltration by lymphocytes and plasma cells. C. Granular, oxyphilic changes in the cytoplasm of the follicular epithelium (Askanazy cells). (From Woolner LB, McConahey WM, Beahrs OH. Struma lymphomatosa [Hashimoto’s thyroiditis] and related thyroidal disorders. J Clin Endocrinol Metab 1959; 19:5363. © 1959, The Endocrine Society.)
**Figure 13-1** Transverse composite sonogram (A) and corresponding anatomic map (B) of the normal thyroid gland. C, common carotid artery; CVII, seventh cervical vertebra; LC, longus colli muscle; SM, strap muscles; SCM, sternocleidomastoid muscle; T, thyroid; TR, trachea. (From Rifkin MD, Charboneau JW, Laing FC. Special course: ultrasound 1991. In Reading CC [ed]. Syllabus: Thyroid, Parathyroid, and Cervical Lymph Nodes. Oak Brook, Ill, Radiological Society of North America, 1991, pp 363-377.)
**Figure 13-2** Cervical map, derived from sonographic images, helps to communicate anatomic relationships of pathology to clinicians and serves as a reference for follow-up examinations. SMG, submandibular gland. (From James EM, Charboneau JW, Hay ID. The thyroid. In Rumack CM, Wilson SR, Charboneau JW [eds]. Diagnostic Ultrasound, vol 1. St. Louis, Mosby Year Book, 1991, pp 507528.)
Figure 13-3 Outer and cut surfaces of a nontoxic nodular goiter of 15 years' duration. Note variations in size and structure of the nodules; there are thick areas of fibrous tissue, flecks of calcium, scattered areas of thyroid tissue, cysts, and small hemorrhages.
**Figure 13-5** Sonographically guided thyroid nodule fine-needle aspiration. Transverse sonogram of the right thyroid lobe (A, left panel) shows a 1.5-cm solid thyroid nodule (arrows) containing a central cystic component. C, common carotid artery; J, jugular vein. Palpation-guided aspiration biopsy obtained nondiagnostic fluid only. B, right panel, shows sonographically guided needle aspiration biopsy (curved arrow) of the solid portion of the nodule, which proved that this was a benign adenomatous nodule. (From Rifkin MD, Charboneau JW, Laing FC. Special course: ultrasound 1991. In Reading CC [ed]. Syllabus: Thyroid, Parathyroid, and Cervical Lymph Nodes. Oak Brook, Ill, Radiological Society of North America, 1991, pp 363377.)
Figure 13-6 Management of nodular goiter based on fine-needle aspiration (FNA) biopsy as the first diagnostic test. Subsequent management is based on cytologic results. Percentages in parentheses indicate satisfactory or unsatisfactory biopsy results. (From Gharib H. Fine needle aspiration biopsy of thyroid nodules: advantages, limitations, and effect. Mayo Clin Proc 1994; 69:4449.)
Figure 13-7 Genetic events in thyroid tumorigenesis. Activating point mutations of the RAS genes are found with a similar frequency in follicular adenomas and follicular carcinomas and are considered an early event in follicular tumorigenesis. The PPAR-PAX8 rearrangement was found only in follicular carcinomas. Rearrangements of transmembrane receptors with tyrosine kinase activity (RET, TRK genes) are found only in papillary thyroid carcinomas. Inactivating point mutations of the P53 gene are found only in poorly differentiated and anaplastic thyroid carcinomas. Activation of the cyclic adenosine monophosphate pathway, by point mutation of the thyrotropin receptor (TSH-R) or the subunit of the G protein genes, leads to the appearance of hyperfunctioning thyroid nodules. Gs stimulatory guanyl nucleotide protein.
Figure 13-8 Distribution of pathologic-Tumor-Node-Metastases (pTNM) stages in 2284 patients with papillary thyroid carcinoma (upper left), 218 patients with medullary thyroid cancer (lower left), 141 patients with follicular thyroid cancer (upper right), and 125 patients with Hurthle cell cancer (lower right) undergoing primary surgical treatment at the Mayo Clinic from 1940 to 1997.
Figure 13-9 Cause-specific survival according to pathologic-Tumor-Node-Metastases (pTNM) stage in a cohort of 2284 patients with papillary thyroid carcinoma treated at the Mayo Clinic from 1940 to 1997. The numbers in parentheses represent the percentages of patients in each pTNM stage grouping.
**Figure 13-10** Development of neck nodal metastases, local recurrences, and distant metastases in the first 20 years after definitive surgery for papillary thyroid cancer (PTC) or medullary thyroid cancer (MTC) performed at the Mayo Clinic from 1940 to 1997. Based on 2150 consecutive PTC (left) and 194 MTC (right) patients who had complete surgical resection (i.e., had no gross residual disease) and were without distant metastases on initial examination. Postop, postoperative.
**Figure 13-11** Development of neck nodal metastases (NM), local recurrences (LR), and distant metastases (DM) in the first 20 years after definitive surgery for follicular thyroid cancer (FTC) or Hürthle cell cancer (HCC) performed at the Mayo Clinic from 1940 to 1997. Based on 110 consecutive FTC patients (left) and 115 HCC patients (right) who had complete surgical resection and were without distant metastases on initial examination.
Figure 13-12 Cumulative cause-specific mortality rates for patients with differentiated thyroid carcinoma in the first 25 years after treatment with initial surgery performed at the Mayo Clinic from 1940 to 1997. Based on 2768 consecutively treated patients (2284 with papillary thyroid carcinoma [PTC], 141 with follicular thyroid cancer [FTC], 125 with Hurthle cell cancer [HCC], and 218 with medullary thyroid cancer [MTC]).
Figure 13-13 Survival to death from all causes and to death from thyroid cancer (cause-specific mortality) in 2284 consecutive patients with papillary thyroid carcinoma undergoing initial management at the Mayo Clinic from 1940 to 1997. Also plotted is the expected survival (all causes) of persons of the same age and sex and with the same date of treatment but living under mortality conditions of the northwest central United States.
Figure 13-14 Lack of influence of nodal metastases at initial operation on cumulative mortality from papillary thyroid carcinoma in 1941 patients with pT1-3 intrathyroidal tumors (completely confined to the thyroid gland) and 209 pT4 patients with extrathyroidal (locally invasive) tumors. All patients had initial surgical treatment at the Mayo Clinic from 1940 to 1997. DM, distant metastases.
Figure 13-15 Cumulative mortality from papillary thyroid carcinoma in patients at either minimal risk or higher risk of cancer-related death as defined by International Union Against Cancer (UICC) pathologic-Tumor-Node-Metastases (pTNM) stages (upper left), AGES scores (upper right), AMES risk groups (lower left), and MACIS scores (lower right). The minimal risk group constitutes 81% of the 2284 patients when defined by pTNM stages I and II, 86% as defined by AGES scores less than 4, 88% as defined by AMES low-risk, and 83% when defined by a MACIS score less than 6. The cause-specific mortality (CSM) rates at 20 years were 25% for stages III and IV, 36% for AGES scores of 4+, 39% for AMES high-risk, and 32% for patients with MACIS scores of 6+. The CSM ratios between the high-risk and low-risk groups at 20 years were 19 for pTNM, 36 for AGES, 35 for AMES, and 40 for MACIS.
Figure 13-16 Cause-specific survival according to MACIS (metastases, age, completeness of resection, invasion, and size) scores of less than 6, 6 to 6.99, 7 to 7.99, and 8+ in a cohort of 2284 consecutive patients with papillary thyroid carcinoma (PTC) undergoing initial treatment at the Mayo Clinic from 1940 to 1997. The numbers in parentheses represent the numbers and percentages of PTC patients in each of the four risk groups.
Figure 13-17 Postoperative recurrence (any site) in the first 20 years after definitive surgery for differentiated thyroid carcinoma performed at the Mayo Clinic from 1940 to 1997. Based on 2569 consecutive patients (2150 papillary thyroid carcinoma, 110 follicular thyroid carcinoma, 115 Hurthle cell carcinoma, and 194 medullary thyroid carcinoma) who had complete tumor resection and had no distant metastases at presentation. The ages in parentheses represent the median age at diagnosis for each of the four histologic subtypes.
Figure 13-18 Cause-specific survival according to pathologic-Tumor-Node-Metastases (pTNM) stages in a cohort of 141 patients with follicular thyroid carcinoma (left panel) and 125 patients with Hürthle cell carcinoma (right panel) treated at the Mayo Clinic from 1940 to 1997. Numbers in parentheses represent the number of patients in each pTNM stage grouping.
Figure 13-19 Comparison of cause-specific survival in 1472 papillary thyroid carcinoma (PTC) and 250 follicular thyroid carcinoma (FTC) patients treated at the Mayo Clinic from 1940 to 1990. 138 of the PTCs were “pure” papillary in histotype (no follicular elements). 97 of the FTC patients had predominantly oxyphilic tumors. There is a significant difference ($P = .0001$) between the PTC and the FTC survival curves. However, within either the PTC or FTC groups, the two survival curves are insignificantly different. (From Grebe SKG, Hay ID. Follicular thyroid cancer. Endocrinol Metab Clin North Am 1995; 24:761801.)
Figure 13-20 Survival to death from all causes in 141 consecutive patients with follicular thyroid carcinoma (left) and 125 patients with Hürthle cell cancer (right) undergoing initial management at the Mayo Clinic from 1940 to 1997. Also plotted is the expected survival (all causes) of persons of the same age and sex and with the same date of treatment but living under mortality conditions of the northwest central United States.
Figure 13-21 Cumulative cause-specific survival among 100 patients with nonoxyphilic follicular thyroid carcinoma treated at the Mayo Clinic from 1946 to 1970, plotted by high-risk and low-risk categories. High risk means that two or more of the following factors were present: age older than 50 years, marked vascular invasion, and metastatic disease at time of initial diagnosis. (From Brennan MD, Bergstralh EJ, van Heerden JA, et al. Follicular thyroid cancer treated at the Mayo Clinic, 1946 through 1970: initial manifestations, pathologic findings, therapy, and outcome. Mayo Clin Proc 1991; 66:1122.)
**Figure 13-22** Survival differences in low-risk, intermediate-risk, and high-risk groups for 228 consecutive patients with follicular thyroid carcinoma who were seen and treated at the Memorial Sloan-Kettering Cancer Center during a period of 55 years from 1930 to 1985. (From Shaha AR, Loree TR, Shah JP. Prognostic factors and risk group analyses in follicular carcinoma of the thyroid. Surgery 1995; 118:1131-1138.)
Figure 13-23 Cause-specific survival according to pathologic-Tumor-Node-Metastases (pTNM) stage in a cohort of 218 patients with medullary thyroid carcinoma treated at the Mayo Clinic from 1940 to 1997. Numbers in parentheses represent the percentages of patients in each pTNM stage grouping.
Figure 13-24  Total-body scan performed 5 days after administration of 100 mCi (3700 MBq) of radioactive iodine ($^{131}$I). The chest radiograph of this asymptomatic 34-year-old patient, who was being monitored for a papillary thyroid carcinoma, was normal; the only abnormality was an elevated serum thyroglobulin level, at 45 ng/mL, during levothyroxine suppressive treatment. Note the presence of diffuse uptake in the lungs (L) and in the left iliac bone (I). After four treatments with 100 mCi of $^{131}$I, metastatic uptake disappeared and the thyroglobulin level became undetectable during levothyroxine therapy. B, bladder; M, mouth; N, nose; S, stomach.
Figure 13-25 The patient was being monitored for a papillary thyroid carcinoma. The serum thyroglobulin level was 22 ng/mL during levothyroxine suppressive treatment, and local imaging modalities were not interpretable because of three previous extensive neck operations. Left, Total-body scan performed 4 days after administration of 100 mCi (3.7 GBq); there is no visible uptake in the neck. Right, Positron emission tomography scan using $^{18}$F-fluorodeoxyglucose ($^{18}$FDG) demonstrating significant uptake in a paratracheal lymph node (arrow) that measured 12 mm in diameter at surgery.
Figure 13-26 Follow-up of high-risk patients with papillary or follicular thyroid carcinoma after near-total thyroidectomy based on serum thyroglobulin (Tg) measurements and ¹³¹I ablation, total-body scanning. LT₄, levothyroxine; TBS, total-body scan; TSH, thyrotropin. Thyroglobulin values are method specific, and the normal range should be determined in each assay. For the total-body scan, above 0 is positive, with ¹³¹I uptake indicative of neoplastic foci; below 0 is negative.
Figure 14-1 Schematic diagram of the structure of the human adrenal cortex, depicting the outer zona glomerulosa and inner zona fasciculata and zona reticularis.
Figure 14-2 The cyclopentanoperhydrophenanthrene structure of corticosteroid hormones highlighting the structure of some endogenous steroid hormones together with their nomenclature.
Figure 14-3 Adrenal steroidogenesis. After the steroidogenic acute regulatory (SIAR) proteinmediated uptake of cholesterol into mitochondria within adrenocortical cells, aldosterone, cortisol, and adrenal androgens are synthesized through the coordinated action of a series of steroidogenic enzymes in a zone-specific fashion. Androstenedione, DHEA, dehydroepiandrosterone, DOC, deoxycorticosterone.
Figure 14-4  A, Electron shuttle system for the mitochondrial enzymes, CYP11A1 and CYP11B1. Adrenodoxin reductase receives electrons from reduced nicotinamide adenine dinucleotide phosphate (NADPH) and reduces adrenodoxin, which transfers reducing equivalents to the CYP enzyme. The enzyme then transfers electrons, by way of oxygen, to the steroid. Fp, flavoprotein; Fp, reduced form of flavoprotein.  B, Electron shuttle system for the microsomal enzymes, CYP17 and CYP21A2. P450 reductase, a flavoprotein, accepts electrons from NADPH and transfers them to the NADPH-P450 enzyme. The enzyme then transfers electrons, by way of oxygen, to the steroid. A second reducing equivalent may be supplied to CYP17 by NADPH-P450 reductase or cytochrome b5.
Figure 14-5 Synthesis and cleavage of pro-opiomelanocortin (POMC) within the human anterior pituitary gland. Prohormone convertase enzymes sequentially cleave POMC to adrenocorticotropic hormone (ACTH). Shaded areas represent melanocyte-stimulating hormone (MSH) structural units. -LPH, -lipoprotein; N-POC, amino-terminal pro-opiomelanocortin.
Figure 14-6 Normal regulation of adrenal glucocorticoid secretion. Adrenocorticotropic hormone (ACTH) is secreted from the anterior pituitary under the influence of two principal secretagogues, corticotropin-releasing hormone (CRH) and arginine vasopressin; other factors including cytokines also play a role. CRH secretion is regulated by an inbuilt circadian rhythm and additional stressors operating through the hypothalamus. Secretion of both CRH and ACTH is inhibited by cortisol, highlighting the importance of negative feedback control.
Figure 14-7 Circadian and pulsatile secretion of adrenocorticotropic hormone (ACTH) and cortisol in a normal subject (top two panels) and in a patient with Cushing's disease. In a normal subject, secretion of ACTH and cortisol is highest in early morning and falls to a nadir at midnight. ACTH pulse frequency and pulse amplitude are increased in Cushing's disease, and circadian rhythm secretion is lost.
Figure 14-8 The normal renin-angiotensin-aldosterone regulatory system. Renin, secreted by the kidney, cleaves angiotensin I (A I) from renin substrate (angiotensinogen), an \( \alpha \)-globulin produced by the liver. Angiotensin I is converted into biologically active angiotensin II by angiotensin-converting enzyme (ACE), mainly in the lung. Angiotensin II increases peripheral vascular resistance, and, together with angiotensin III, stimulates aldosterone (ALDO) secretion, which results in sodium retention and increased plasma volume.
Figure 14-9 Schematic structure of the human genes encoding the glucocorticoid receptor (GR) and mineralocorticoid receptor (MR). In both cases splice variants have been described; in the case of the GR, there is evidence that the GR isoform can act as a dominant negative inhibitor of GR action. mRNA, messenger ribonucleic acid.
Figure 14-10 The anti-inflammatory action of glucocorticoids. Cortisol binds to the cytoplasmic glucocorticoid receptor (GR). Conformational changes in the receptor-ligand complex result in dissociation from heat shock proteins (HSPs) 70 and 90 and migration to the nucleus. Binding occurs to specific DNA motifs, glucocorticoid response elements in association with the activator protein-1 (AP-1) comprising c-fos and c-jun. Glucocorticoids mediate their anti-inflammatory effects through several mechanisms: (1) The inhibitory protein IB, which binds and inactivates nuclear factor B (NFB), is induced. (2) The GR-cortisol complex is able to bind NFB and thus prevent initiation of an inflammatory process. (3) Both GR and NFB compete for the limited availability of coactivators that include cyclic adenosine monophosphate response element binding protein (CREB) binding protein and steroid receptor coactivator-1.
**Figure 14-11** Mineralocorticoid hormone action. An epithelial cell is depicted in the distal nephron or distal colon, or both. The much higher concentrations of cortisol are inactivated by the type 2 isozyme of 11-hydroxysteroid dehydrogenase (11-HSD2) to cortisone, permitting the endogenous ligand, aldosterone, to bind to the mineralocorticoid receptor (MR). Relatively few mineralocorticoid target genes have been identified, but these include serum and glucocorticoid-induced kinase (SGK), subunits of the epithelial sodium channel (ENaC), and basolateral Na⁺,K⁺-adenosine triphosphatase.
Figure 14-12 The principal pathways of cortisol metabolism. Interconversion of hormonally active cortisol to inactive cortisone is catalyzed by two isozymes of 11-hydroxysteroid dehydrogenase (11-HSD). 11-HSD1 principally converting cortisone to cortisol and 11-HSD2 the reverse. Cortisol can be hydroxylated at the C6 and C20 positions. A ring reduction is undertaken by 5-reductase or 5-reductase and 3-hydroxysteroid dehydrogenase.
Figure 14-13 The principal sites of action of glucocorticoids in humans highlighting some of the consequences of glucocorticoid excess. CNS, central nervous system; GI, gastrointestinal; FSH, follicle-stimulating hormone; GH, growth hormone; LH, luteinizing hormone; TSH, thyroid-stimulating hormone.
Figure 14-14 Structures of the natural glucocorticoid cortisol, some of the more commonly prescribed synthetic glucocorticoids, and the mineralocorticoid fludrocortisone. Note that triamcinolone is identical to dexamethasone except that a 16-hydroxyl group is substituted for the 16-methyl group. Betamethasone, another widely used glucocorticoid, has a 16-methyl group.
Figure 14-15 Minnie G., Cushing's index patient, at age 23 years. (From Cushing H. The basophil adenomas of the pituitary body and their clinical manifestations [pituitary basophilism]. Bull Johns Hopkins Hosp 1932; 50:137195.)
Figure 14-16 Clinical features of Cushing's syndrome. A. Centripetal and some generalized obesity and dorsal kyphosis in a 30-year-old woman with Cushing's disease. B. Same woman as in A, showing moon facies, plethora, hirsutism, and enlarged supraclavicular fat pads. C. Facial rounding, hirsutism, and acne in a 14-year-old girl with Cushing's disease. D. Central and generalized obesity and moon facies in a 14-year-old boy with Cushing's disease. E and F. Typical centripetal obesity with livid abdominal striae seen in a 41-year-old woman (E) and a 40-year-old man (F) with Cushing's syndrome. G. Striae in a 24-year-old patient with congenital adrenal hyperplasia treated with excessive doses of dexamethasone as “replacement” therapy. H. Typical bruising and thin skin of Cushing's syndrome. In this case, the bruising has occurred without obvious injury.
**Figure 14-18** Patterns of cortisol secretion in three patients with cyclical Cushing’s syndrome. In each case, ratios of early morning urinary cortisol (nmol/L) to creatinine (mmol/L) are plotted against time. Variable periodicity in cortisol hypersecretion is shown. (From Atkinson AB, McCance DR, Kennedy L, et al. Cyclical Cushing’s syndrome first diagnosed after pituitary surgery: a trap for the unwary. Clin Endocrinol 1992; 36:297-299.)
Figure 14-19 Investigation of a patient with suspected Cushing's syndrome. The laboratory diagnosis of Cushing's syndrome and the differential diagnosis of its cause are debatable and differ in any given center depending on many factors, including familiarity, turn-around time of hormone assays, and local expertise in techniques such as inferior petrosal sinus sampling (IPSS). Depicted here is an algorithm in use within many endocrine units based upon the reported sensitivity and specificity of each endocrine test.
Figure 14-20 Plasma adrenocorticotropic hormone (ACTH) concentrations in patients with Cushing's disease and Cushing's syndrome associated with adrenocortical tumors and ectopic ACTH syndrome. To convert values to pmol/L, multiply by 0.2202. (From Besser GM, Edwards CRW. Cushing's syndrome. Clin Endocrinol Metab 1972; 1:451490.)
Figure 14-21 Comparison of the cortisol and adrenocorticotropic hormone (ACTH) responses to an intravenous injection of ovine corticotropin-releasing hormone (1 µg/kg) in normal subjects, patients with Cushing's disease, and patients with ectopic ACTH. (From Chrousos GP, Schulte HM, Oldfield EH, et al. The corticotropin-releasing factor stimulation test: an aid in the evaluation of patients with Cushing's syndrome. N Engl J Med 1984; 310:624-626.)
**Figure 14-22** Anatomy of the venous drainage of the pituitary gland through the inferior petrosal venous sinuses. (From Oldfield EH, Chrousos GP, Schulte HM, et al. Preoperative lateralization of ACTH-secreting pituitary microadenomas by bilateral and simultaneous inferior petrosal sinus sampling. Reprinted by permission of The New England Journal of Medicine 1985; 312:100103.)
Figure 14-23 A. Magnetic resonance imaging (MRI) scan of pituitary demonstrating the typical appearance of a pituitary microadenoma. A hypodense lesion is seen in the right side of the gland with deviation of the pituitary stalk away from the lesion. After a biochemical diagnosis of Cushing's disease, this patient was cured following transsphenoidal hypophysectomy. B. MRI scan of the pituitary gland demonstrating a large macroadenoma in a patient with Cushing's disease. In contrast to smaller tumors, these tumors are invariably invasive and recur after surgery.
Figure 14-24  A, Adrenal computed tomographic (CT) scan demonstrating bilateral adrenal hyperplasia in a patient with Cushing’s disease.  B, CT scan of a typical solitary left adrenal adenoma causing Cushing's syndrome.  C, Cushing’s syndrome caused by massive macronodular hyperplasia. Adrenal glands are replaced by multiple nodules (arrows). Combined weight of adrenal glands was over 100 g.  D, Cushing’s syndrome caused by surgically proven primary pigmented nodular adrenal disease in a 21-year-old patient. Notice the multiple small nodules with the relatively atrophic internodular adrenocortical tissue involving the medial limb of the right adrenal gland (arrow). (C and D from Findling JW, Doppman JL. Biochemical and radiologic diagnosis of Cushing’s syndrome. Endocrinol Metab Clin North Am 1994; 23:511537.)
Figure 14-25 Computed tomographic scan of a patient with rapidly progressing Cushing's syndrome caused by an adrenal carcinoma. An irregular right adrenal mass is shown (A) with a large liver metastasis (B).
Figure 14-27 A young woman with Cushing's disease, photographed initially beside her identical twin sister (A). In this case, treatment with bilateral adrenalectomy was undertaken. Several years later, the patient presented with Nelson's syndrome and a right third cranial nerve palsy (B and C) related to cavernous sinus infiltration from a locally invasive corticotropinoma (D). Hypophysectomy and radiotherapy were performed with reversal of the third nerve palsy (E). Note the advancing skin pigmentation of Nelson's syndrome.
Figure 14-28  Selective removal of a microadenoma and its effect on the hypothalamic-pituitary-adrenal axis. Because the surrounding normal pituitary corticotrophs are suppressed in a patient with an adrenocorticotropic hormone (ACTH)-secreting pituitary adenoma, successful removal of the tumor results in ACTH and hence adrenocortical deficiency with an undetectable (<50 nmol/L [2 µg/dL]) plasma cortisol level. A plasma cortisol level higher than 50 nmol/L (2 µg/dL) postoperatively implies that the patient is not cured. (Courtesy of Dr. Peter Trainer.)
Figure 14-29 Gradual recovery of function of the hypothalamic-pituitary-adrenal axis after removal of a pituitary adrenocorticotropic hormone-secreting microadenoma. The insulin hypoglycemia test (I.H.T.) eventually demonstrated the return of a normal stress response.
Figure 14-30 Pigmentation in Addison's disease. **A**, Hands of an 18-year-old woman with autoimmune polyendocrine syndrome and Addison's disease. Pigmentation in a patient with Addison's disease before (**B**) and after (**C**) treatment with hydrocortisone and fludrocortisone. Note the additional presence of vitiligo. **D**, Similar changes also seen in a 60-year-old man with tuberculous Addison's disease before and after corticosteroid therapy. **E**, Buccal pigmentation in the same patient. (**B** and **C**, courtesy of Professor C.R.W. Edwards.)
Figure 14-31  Computed tomographic (CT) scans of patients with primary adrenal insufficiency. The affected adrenal glands are indicated by arrows. A, CT scan of a 59-year-old man with histoplasmosis. Note the subcapsular calcium in both glands. B, CT scan of a 59-year-old man with metastatic melanoma. C, CT scan of an 80-year-old man with bilateral adrenal hemorrhage resulting from anticoagulation for pulmonary emboli. D, Bilateral adrenal tuberculomas in a 79-year-old man with tuberculosis affecting the urogenital tract. (A and B courtesy of Dr. William D. Salmon, Jr.; C courtesy of Dr. Craig R. Sussman.)
Figure 14-32 Congenital adrenal hyperplasia related to 21-hydroxylase deficiency. The normal synthesis of cortisol is impaired, and adrenocorticotropic hormone (ACTH) levels increase because of loss of normal negative feedback inhibition resulting in an increase in adrenal steroid precursors proximal to the block. The results are cortisol deficiency, variable mineralocorticoid deficiency, and excessive secretion of adrenal androgens. DHEA, dehydroepiandrosterone; DOC, deoxycorticosterone; HSD, hydroxysteroid dehydrogenase; StAR, steroidogenic acute regulatory protein.
Figure 14-33 Map of the short arm of human chromosome 6 (upper bar), showing the relative positions of the genes encoding the major histocompatibility proteins A, C, B, DR, DQ, and DP. The detail (lower bar) shows the approximately 120-kilobase region containing the genes for complement component C2, properdin factor B (Bf), and the duplicated complement C4 gene (C4A and C4B). The pseudogene CYP21A1 and the functional gene CYP21A2 are in tandem array with the two C4 genes. HLA, human leukocyte antigen.
Figure 14-34  Basal and stimulated plasma 17-hydroxyprogesterone (17OHP) concentrations in patients with CYP21A2 (21-hydroxylase) deficiency. To convert values to nmol/L, multiply by 0.0303. The mean for each group is indicated by a large cross and the adjacent letter: c, patients with classical CYP21A2 deficiency; v, patients with nonclassical (acquired and cryptic) CYP21A2 deficiency; h, heterozygotes for all forms of CYP21A2 deficiency; p, general population; u, known unaffected persons (e.g., siblings of patients with CYP21A2 deficiency who carry neither affected parental haplotype as determined by human leukocyte antigen typing). (From White PC, New MI, Dupont B. Congenital adrenal hyperplasia: part 1. Reprinted by permission of The New England Journal of Medicine 1987; 316:15191524.)
Congenital adrenal hyperplasia related to 11-hydroxylase deficiency. The normal synthesis of cortisol is impaired, and adrenocorticotropic hormone (ACTH) levels increase because of loss of normal negative feedback inhibition resulting in an increase in adrenal steroid precursors proximal to the block. The results are cortisol deficiency, mineralocorticoid excess related to excessive deoxycorticosterone (DOC) secretion, and excessive secretion of adrenal androgens. DHEA, dehydroepiandrosterone; SIAR, steroidogenic acute regulatory protein.
Figure 14-36 Congenital adrenal hyperplasia related to 17-hydroxylase deficiency. The normal synthesis of cortisol is impaired, and adrenocorticotropic hormone (ACTH) levels increase because of loss of normal negative feedback inhibition resulting in an increase in adrenal steroid precursors proximal to the block. The result is cortisol deficiency and mineralocorticoid excess usually related to deoxycorticosterone (DOC) excess. Because gonadal 17-hydroxylase activity is also absent, sex steroid secretion in addition to adrenal androgen secretion is severely impaired, resulting in hypogonadism. DHEA, dehydroepiandrosterone; STAR, steroidogenic acute regulatory protein.
Figure 14-37 Congenital adrenal hyperplasia related to 3-hydroxysteroid dehydrogenase (3-HSD) deficiency resulting in cortisol deficiency and variable mineralocorticoid deficiency. Gonadal 3-HSD activity is also absent, resulting in male pseudohermaphroditism and hypogonadism or primary amenorrhea in females. ACTH, adrenocorticotropic hormone; DOC, deoxycorticosterone; DHEA, dehydroepiandrosterone; STAR, steroidogenic acute regulatory protein.
Figure 14-38  A. Adrenal incidentaloma discovered in a woman undergoing investigation for abdominal pain.  B. Incidentally discovered right adrenal myelolipoma.
Figure 15-1 Organization of the sympathoadrenal system. Sympathetic preganglionic axons arise in large part from cells located in the thoracolumbar spinal cord. These preganglionic sympathetic neurons in turn have synapses with descending tracks from neurons in the pons, medulla, and hypothalamus, allowing regulation of sympathetic activity by the brain. In turn, these central nervous system neurons that influence sympathetic activity are regulated by a variety of factors, including substrates (glucose) and hormones (corticotropin-releasing hormone). (From Landsberg L, Young JB. Catecholamines and the adrenal medulla. In Bondy PK, Rosenberg JE (eds). Metabolic Control and Disease, 8th ed. Philadelphia: WB Saunders, 1980:1621693.)
Figure 15-2 Sympathoadrenomedullary efferent autonomic pathways. The axons of the preganglionic neurons synapse with postganglionic cell bodies located in the paravertebral and preaortic ganglia as well as neurons in the celiac and superior and inferior mesenteric ganglia. Postganglionic axons from the cell bodies located in these ganglia in turn innervate the visceral organs. The splanchnic outflow of the lower thoracic and lumbar preganglionic axons also directly innervates the cells of the adrenal medulla.
Figure 15-3 Biosynthetic pathway for catecholamines (left to right). All catecholamines contain the catechol nucleus. L-Tyrosine is converted to L-3,4-dihydroxyphenylalanine (L-dopa) in the rate-limiting step by tyrosine hydroxylase (TH). Aromatic L-amino acid decarboxylase (AADC) converts L-dopa to dopamine. Dopamine is hydroxylated to L-norepinephrine by dopamine-beta-hydroxylase (DBH). L-Norepinephrine is converted to L-epinephrine by phenylethanolamine N-methyltransferase (PNMT).
Figure 15-4  Catecholamine metabolism. Metabolism of catecholamines occurs through two enzyme pathways. Catechol-O-methyltransferase (COMT) converts epinephrine to metanephrine and converts norepinephrine to normetanephrine by meta-O-methylation. Metanephrine and normetanephrine are oxidized by monoamine oxidase (MAO) to vanillylmandelic acid (VMA) by oxidative deamination. MAO also may oxidize epinephrine and norepinephrine to dihydroxymandelic acid (DOMA), which is then converted by COMT to VMA.
Figure 15-5 Algorithm for diagnosis of pheochromocytoma. CT, computed tomography; MRI, magnetic resonance imaging.
Figure 15-6. A, Axial in-phase gradient-echo image demonstrates left adrenal mass (arrow). B, Axial image from the out-of-phase gradient-echo sequence demonstrates the left adrenal mass. In comparison to the in-phase image, no suppression of the adrenal mass is present. Suppression is the rule in lipid-containing cortical adenomas.
Figure 15-7 Components of the renin-angiotensin system. (Redrawn from Williams GH, Chao J, Chao L. Kidney hormones. In Conn PM, Melmed S [eds]. Endocrinology: Basic and Clinical Principles. Totowa, NJ, Humana Press, 1997, pp 393404.)
Figure 15-9 Modification of vascular and aldosterone response to angiotensin II by dietary salt intake. Sodium intake has a reciprocal influence on vascular and adrenal responses to angiotensin II. On a high salt intake the vascular response is enhanced while the adrenal response is suppressed. Sodium restriction has the opposite effect. (Redrawn from Williams GH, Hollenberg NK. "Sodium-sensitive" essential hypertension: emerging insights into pathogenesis and therapeutic implications. In Klahr S, Massry SG [eds]. Contemporary Nephrology. Vol 3. New York, Plenum Press, 1985, p 303, with permission.)
Figure 15-10 Relationship between the activity of the renin-angiotensin system and the mechanisms underlying the hypertension. (Redrawn from Laragh JH, Sealey JE, Niarchos AP, et al. The vasoconstrictor volume spectrum in normotension and in the pathogenesis of hypertension. Fed Proc 41:2415-2423, 1982.)
Figure 15-11 Effect of angiotensinogen genotype on renal blood flow responses to angiotensin II infusions. Subjects were classified according to their alleles at the 235 codon of the angiotensinogen gene as to whether they were homozygous for the wild type (MM), heterozygous (MT), or homozygous for the hypertensive-link (TT) alleles. The subjects with the TT235 genotype had a renal blood flow response to angiotensin II similar to that of nonmodulators. (Redrawn from Hopkins P, Lifton RP, Hollenberg NK, et al. Blunted renal vascular response to angiotensin II is associated with a common variant of the angiotensinogen gene and obesity. J Hypertens 1996; 14:199207. Copyright 1996, Rapid Science Publishers.)
Figure 15-12 Pathogenesis of nonmodulating hypertension. ANG II, angiotensin II.
Figure 15-13 Relationship of arterial pressure to insulin resistance in normotensive ethnic subgroups. An index of insulin sensitivity, the glucose disposal rate, was determined with an insulin clamp and fasting insulin levels were measured in normotensive blacks, whites, and Pima Indians. There was a significant negative correlation between arterial blood pressure and glucose disposal rates in whites but not in the other subgroups.  

Figure 15-14 Mechanisms by which insulin resistance may produce hypertension.
**Figure 15-15** Diagnostic and therapeutic flow chart for evaluating patients for renovascular hypertension. BP, blood pressure; MR, magnetic resonance.* Stenting of renal arteries is commonly done following angioplasty.
Figure 15-16 Diagnostic flow chart for evaluating the hypokalemic hypertensive patient. Differentiation of APA versus IHA. PRA, plasma renin activity; PAC, plasma aldosterone concentration; APA, aldosterone-producing adenoma; IHA, idiopathic hyperaldosteronism; CT, computed tomography. *Suppressed values: PAC < 280 pmol/L (< 10 ng/dL) following IV saline; aldosterone excretion rate 39 nmol/d (< 14 µg/d) after oral salt loading.
Figure 16-1 Endocrine interactions in the female reproductive axis. Some of the well-characterized endocrine interactions between the hypothalamus, pituitary, ovary, and endometrium for regulation of the menstrual cycle are depicted. \( E_2 \), estradiol; FSH, follicle-stimulating hormone; LH, luteinizing hormone; LHRH, luteinizing hormonereleasing hormone; P, progesterone.
**Figure 16-2** Luteinizing hormone-releasing hormone (LHRH) production.  

A. The LHRH gene encodes a precursor protein named pre-pro-LHRH in the neuronal body. LHRH is released from this protein by proteolytic processing, which gives rise to LHRH and LHRH-associated protein (GAP) within the neuronal body. Both LHRH and GAP are transported in an axon to the nerve terminal and secreted into the portal circulation.  

B. Pre-pro-LHRH is a 92-amino-acid (aa) protein. The biologically active decapeptide (aa 1 to 10) is sandwiched between the 23-aa signal peptide and the Gly-Lys-Arg sequence. The arrow indicates the site of proteolytic processing. The C-terminal 56-aa peptide is cleaved to produce GAP. (From Yen SSC. Endocrine regulation of the reproductive system. In Yen SSC, Jaffe RB, Barbieri RL [eds]. Reproductive Endocrinology, 4th ed. Philadelphia, WB Saunders, 1999, p 44.)
Figure 16-3 Effect of pulsatile or continuous administration of luteinizing hormonereleasing hormone (LHRH) to ovariectomized monkeys rendered them LHRH-deficient by placement of a lesion in the hypothalamus. Release of LH and follicle-stimulating hormone (FSH) was restored by hourly LHRH infusion, inhibited during a continuous infusion, and again restored after reinstitution of pulsatile LHRH administration. (Adapted from Belchetz PE, Plant TM, Nakai Y, et al. Hypophysial responses to continuous and intermittent delivery of hypothalamic gonadotropin releasing hormone. Science 1978; 202:631633. Copyright © 1978, by American Association for the Advancement of Science.)
Figure 16-4 Regulation of luteinizing hormone-releasing hormone (LHRH), LH, and follicle-stimulating hormone (FSH) secretion. Locally synthesized and systemic hormones regulate the pulsatile secretion of LHRH from the hypothalamus into the portal circulation. In turn, LHRH together with a number of steroid and peptide hormones regulate the synthesis of and gonadotropin subunits and the formation and secretion of FSH and LH. CRH, corticotropin-releasing hormone; E₂, estradiol; P₄, progesterone.
Figure 16-7 Transverse section of the caudal region of a 5-week embryo showing the location of gonadal ridges, the primordium of the adrenal glands, and the migration path of primordial germ cells. From the third week on, germ cells arising from the yolk sac cross the dorsal mesentery of the hindgut and migrate to the gonadal ridges. By the end of the fifth week, rapid division of primordial germ cells, gonadal epithelium, and mesenchyme starts the early gonad that differentiates subsequently to the ovary in a 46,XX fetus. CC, coelomic cavity. (Modified from Moore K. The Developing Human. Philadelphia, WB Saunders, 1983.)
Figure 16-8 Age-dependent changes in germ cell number in the human ovary. The highest number of oocytes is found in the ovaries of a human fetus at midgestation. This number decreases sharply during the third trimester. After birth, the progressive decline in the number of ovarian follicles containing oocytes continues until complete depletion at the menopause. (From Baker TG. A quantitative and cytological study of germ cells in the human ovaries. Proc R Soc Biol Sci 1963; 158:417433.)
Figure 16-9 Meiotic cell division. Meiosis occurs exclusively in germ cells and serves two critical purposes: (1) generation of germ cells genetically distinct from the somatic cells and (2) generation of a mature egg with a reduction in the number of chromosomes from 46 to 23. Genetic recombination through crossing over of genes between homologous chromosomes and random assortment of (original) maternal and paternal chromosomes into daughter cells during the first meiotic division are responsible for the first function of meiosis, maintenance of genetic diversity. The second function is provided by a reduction in the number of chromosomes so that each daughter cell, or ovum, receives randomly one chromosome from each of the 23 pairs. During fertilization, the fusion of ovum and sperm, each of which has 23 chromosomes, produces a genetically novel individual with 46 chromosomes. The chromosome marked as white in the oogonium (upper left corner) originates from the father of the fetus, whereas the black chromosome comes from the mother of the fetus. The random exchange of genes (alleles) between homologous chromosomes (crossing over) takes place before the meiotic arrest in the prophase I stage before birth. During postnatal life, these oocytes remain in meiotic arrest until puberty. In the developing oocyte in the graafian follicle, meiosis I is resumed immediately after the preovulatory luteinizing hormone (LH) surge during each ovulatory cycle. Meiotic maturation is defined as the period from the breakdown of the oocyte's nucleus (germinal vesicle, GV) until the oocyte reaches metaphase II (i.e., transition from oocyte to egg). A second and short meiotic arrest occurs at metaphase II until the oocyte is fertilized by a sperm. DNA, deoxyribonucleic acid; GVBD, germinal vesicle breakdown; mat, maternal; n, the amount of DNA material in haploid number (23) of chromosomes; pat, paternal.
Figure 16-14 Complete follicular growth trajectory. Class 1 follicle is a secondary follicle with theca cells and is presumed to become responsive to gonadotropins. Although the tonic (early) stage of follicle development (class 1 to 4) is likely to be gonadotropin-dependent (albeit to a lesser extent), the final stages of follicular development (class 5 to 8) are the ones heavily dependent on gonadotropins. According to this view, late luteal phase, class 5 follicles constitute the cohort from which the follicle destined to ovulate in the following cycle is recruited. The exponential gonadotropin-dependent growth phase (class 5 to 8) takes place during the follicular phase of the cycle following the third menses from initiation of the growth phase. During this time, follicular selection and dominance are accomplished. The total duration of the process wherein a class 1 follicle is converted into preovulatory class 8 follicle is estimated to be 85 days and spans three ovulatory cycles. M, menses; Ovul, ovulation; Gn, gonadotropin. (Courtesy of A. Gougeon.)
**Figure 16-15** Gonadotropin-dependent (exponential) follicular growth phase: recruitment, selection, and the attainment of dominance. At and At, atretic follicle; H, healthy follicle. (Courtesy of A. Gougeon.)
**Figure 16-19** Schematic representation of the human gonadotropin receptor genes. The structure of the genes is depicted at the top of the drawings. The open bars indicate sections of the exons that encode untranslated regions of the messenger ribonucleic acid; the closed bars indicate the sequences that encode the protein. Both genes are at least 80 kb in size. The relation between the intron-exon structure of the gene and the domains on the protein is indicated by the lines connecting the gene to the protein. The horizontally hatched part of the protein indicates the signal peptide, and the crosshatched bars signify the seven segments of the transmembrane domain. The numbers below the protein indicate the start and end of the signal peptide and the length of the total protein product including the signal peptide. Note that the receptor genes are similar in structure with the exception of an additional exon in the luteinizing hormone (LH) receptor gene. Exon 1 encodes the signal peptide and a small part of the extracellular domain; the following eight or nine exons encode the rest of the extracellular domain, including the leucine-rich repeat motifs. In both receptor genes, the final exon is the largest and contains the information for the transmembrane signal transduction domain. FSH, follicle-stimulating hormone. (From Themmen APN, Huhtaniemi IT. Mutations of gonadotropins and gonadotropin receptors: elucidating the physiology and pathophysiology of pituitary-gonadal function. Endocr Rev 2000; 21:455-1583. Copyright © 2000 by The Endocrine Society.)
Figure 16-20  Two-cell hypothesis for ovarian steroidogenesis.  A, The preovulatory follicle produces estradiol through a paracrine interaction between theca and granulosa cells. In response to stimulation with a gonadotropin, steroidogenic factor-1 (SF-1, a member of the orphan nuclear receptor family) acts as a master switch to initiate transcription of a series of steroidogenic genes in each cell type. Because granulosa cells do not have a direct connection to the circulation, aromatase (P450arom) in granulosa cells is dependent for substrate on androstenedione that diffuses from theca cells. Two critical steps in estradiol formation seem to be the entry of cholesterol into mitochondria facilitated by steroidogenic acute regulatory protein (StAR) in theca cells and the conversion of androstenedione to estrone catalyzed by P450arom in granulosa cells.  B, In the corpus luteum, granulosa-lutein cells are heavily vascularized, which is critical for the entry of cholesterol into this cell type through primarily low-density lipoprotein receptors and for secretion of large amounts of progesterone into the circulation. The entry of cholesterol into mitochondria (by StAR) is likely to be the most critical steroidogenic step for progesterone formation in granulosa lutein cells. Androstenedione produced in theca-lutein cells serves as a substrate for estradiol produced in granulosa-lutein cells. Gonadotropins and the transcription factor SF-1 play key roles for important steroidogenic steps in both cell types. ATP, adenosine triphosphate; cAMP, cyclic adenosine monophosphate; FSH-R, follicle-stimulating hormone receptor; HSD, hydroxysteroid dehydrogenase; LH-R, luteinizing hormone receptor.
Figure 16-21 Steroidogenic pathway in the ovary. Biologically active steroids progesterone and estradiol are produced primarily in the ovary of a woman of reproductive age. Estradiol production requires the activity of six steroidogenic proteins including StAR and six enzymatic steps. P450c17, product of the CYP17 gene, catalyzes two enzymatic reactions. The four rings of the cholesterol molecule and its derivative steroids are identified by the first four letters in the alphabet, and the carbons are numbered in the sequence shown in the insert. 3-HSD-II, 3-hydroxysteroid dehydrogenase [5][4] isomerase type II; 17-HSD-1, 17-hydroxysteroid dehydrogenase type 1; P450arom, aromatase; P450c17, 17-hydroxylase/17,20-lyase; StAR, steroidogenic acute regulatory protein.
Figure 16-22 Changes in the ovarian follicle, endometrial thickness, and serum hormone levels during a 28-day menstrual cycle. Menses occur during the first few days of the cycle. E$_2$, estradiol; FSH, follicle-stimulating hormone; LH, luteinizing hormone; Inh, inhibin; P, progesterone.
Figure 16-23 Estrogen biosynthesis in women. The biologically active estrogen estradiol (E$_2$) is produced in at least three major sites: (1) by direct secretion from the ovary in reproductive-age women; (2) by conversion of circulating androstenedione (A) of adrenal or ovarian origins, or both, to estrone (E$_1$) in peripheral tissues; and (3) by conversion of A to E$_1$ in estrogen target tissues. In the latter two instances, estrogenically weak E$_1$ is further converted to E$_2$ within the same tissue. The presence of the enzyme aromatase and 17-hydroxysteroid dehydrogenase (17-HSD) is critical for E$_2$ formation at these sites. E$_2$ formation by peripheral and local conversion is particularly important in post-menopausal women and in estrogen-dependent diseases such as breast cancer, endometriosis, and endometrial cancer.
Figure 16-24 Functional anatomy of the endometrium. Endometrium is a multilayered mucosa specialized for implantation and support of pregnancy. A single, continuous layer of epithelial cells lines the surface of the stroma and penetrates the stroma with deep invaginations almost all the way down to the myometrium-endometrium junction. The entire thickness of the endometrium is penetrated by the spiral arteries and their capillaries. Spiral arteries originate from the radial branches of arcuate arteries, which in turn arise from uterine arteries. The superficial layer (functionalis) is shed during menstruation, whereas the permanent bottom layer (basalis) gives rise to the regeneration of endometrium after each menstruation. The striking changes in the spiral arteries (coiling, stasis, vasodilatation followed by intense vasoconstriction) are consistently observed before the onset of every menstruation episode. (Courtesy of Kristof Chwalisz.)
Figure 16-25 Cyclic changes in thickness and morphology of endometrium and the relation of these changes to those of the ovarian cycle. (From Cunningham FG, MacDonald PC, Gant NF, et al. The endometrium and decidua: menstruation and pregnancy. In Williams Obstetrics, 19th ed. Stamford, Conn, Appleton & Lange, 1993, pp 81109.)
Figure 16-26 Dating the endometrium by morphology. (From Noyes RW, Hertig AW, Rock J. Dating the endometrial biopsy. Fertil Steril 1950; 1:325.)
Critical epithelial effects of estrogen (e.g., deoxyribonucleic acid [DNA] synthesis, proliferation, and gene expression) are mediated primarily by estrogen receptor (ER) in stromal cells in a paracrine manner in the endometrium.
**Figure 16-28** The antiestrogenic effects of progesterone on epithelial cells (e.g., decreased proliferation and enhanced differentiation) are mediated primarily by progesterone receptors (PRs) in stromal cells in a paracrine manner in the endometrium. DNA, deoxyribonucleic acid; ER, estrogen receptor.
Figure 16-29 Diagrammatic representation of donation of excess oocytes by a woman undergoing in vitro fertilization (IVF) to a woman with ovarian failure treated with exogenous estrogen and progesterone. A, Woman with ovarian failure treated with increasing doses of estrogen during days 1 through 14 of the cycle. Exogenous progesterone was added to the estrogen treatment on days 15 through 28 and continued if pregnancy was diagnosed. Seven donor eggs were fertilized with sperm from the recipient's husband, and five embryos were transferred to the uterus on day 16 to 18. B, The IVF patient-donor was treated with human menopausal gonadotropin (hMG) until day 8, when human chorionic gonadotropin (hCG) was given, and oocytes were harvested 32 to 36 hours later. Half of the eggs were donated to the recipient, and the other half were fertilized with sperm from the donor's husband; the five fertilized eggs were transferred to the uterus of the IVF donor. Serum levels of E$_2$ (estradiol) and P$_4$ (progesterone) in both women are shown. To convert estradiol values to picomoles per liter, multiply by 3.671. To convert progesterone values to nanomoles per liter, multiply by 3.180. (Adapted from Rosenwaks Z. Donor eggs: their application in modern reproductive technologies. Fertil Steril 1987; 47:895909.)
Figure 16-31 Severe clitoromegaly resulting from a testosterone-secreting ovarian tumor.  

A. The entire length of the clitoris is approximately 4 cm (normal < 1 cm).  
B. The transverse diameter of the clitoris measures 1.5 cm (normal < 0.7 cm).
Androgen biosynthesis in women. There are two biologically active androgens, testosterone (T) and dihydrotestosterone (DHT). Depending on the menstrual cycle phase or postmenopausal status, 20% to 30% of T is secreted by the ovary. The rest of T production (blood) is accounted for by the conversion of circulating androstenedione (A) to T in various peripheral tissues. Both the adrenal and ovary contribute to circulating A directly or indirectly depending on the cycle phase, reproductive-age versus postmenopausal status, and chronologic age. Moreover, T may also be formed locally in androgen target tissues. Finally, T is converted to the more potent androgen DHT within the target tissues and cells. For example, local conversion of T to DHT in sex skin fibroblasts and hair follicles amplifies androgenic action for clitoral enlargement and hirsutism. DHEA, dehydroepiandrosterone; HSD, hydroxysteroid dehydrogenase.
Figure 16-33 Polycystic ovaries. A, Operative findings of classical enlarged polycystic ovaries. The uterus is located adjacent to the two enlarged ovaries. B, Sectioned polycystic ovary with numerous follicles. C, Histologic section of a polycystic ovary with multiple subcapsular follicular cysts and stromal hypertrophy (low power, left). At higher power (× 100), islands of luteinized theca cells are visible in the stroma (right). This morphologic change is called stromal hyperthecosis and appears to be directly correlated with circulating insulin levels. (C, From Dunaif A. Insulin resistance and the polycystic ovary syndrome: mechanism and implications for pathogenesis. Endocr Rev 1997; 18:774800. Copyright © 1997 by The Endocrine Society.)
Figure 16-34 Acanthosis nigricans. A, Moderate acanthosis nigricans (darkening and thickening of skin) at the lateral lower fold of the neck. Note facial hirsutism (sideburns) in the same patient. B, Severe acanthosis nigricans in another patient with severe insulin resistance. (B, courtesy of Dr. R. Ann Word.)
Figure 16-36 Pathologic mechanisms in polycystic ovary syndrome (PCOS). A deficient in vivo response of the ovarian follicle to physiologic quantities of follicle-stimulating hormone (FSH), possibly because of an impaired interaction between signaling pathways associated with FSH and insulin-like growth factors (IGFs) or insulin, may be an important defect in PCOS. This ovarian defect may be the key event responsible for anovulation in PCOS. Insulin resistance associated with increased circulating and tissue levels of insulin and bioavailable estradiol (E₂), testosterone (T), and IGF-I gives rise to abnormal hormone production in a number of tissues. Oversecretion of luteinizing hormone (LH) and decreased output of FSH by the pituitary, decreased production of testosterone-binding globulin (TeBG) and IGF-binding protein 1 (IGFBP-1) in the liver, increased adrenal secretion of dehydroepiandrosterone sulfate (DHEAS), and increased ovarian secretion of androstenedione (A) all contribute to the vicious circle that maintains anovulation and androgen excess in PCOS. Excessive amounts of E₂ and T arise primarily from the conversion of A in peripheral and target tissues. 17-HSD, 17-hydroxysteroid dehydrogenase; 5-red, 5-reductase.
Figure 16-37 Extraovarian conversion of androstenedione to androgen and estrogen. Androstenedione of adrenal or ovarian origin, or both, acts as a dual precursor for androgen and estrogen. Five percent of circulating androstenedione is converted to circulating testosterone, whereas 1.3% of circulating androstenedione is converted to circulating estrone in peripheral tissues. Testosterone and estrone are further converted to biologically potent steroids, dihydrotestosterone and estradiol, in peripheral and target tissues. Biologically active amounts of estradiol in serum are measured in pg/mL (pmol/L), whereas biologically active levels of testosterone in serum are measured in ng/mL (nmol/L). Thus, 1.3% conversion of normal quantities of androstenedione to estrone may have a critical biologic impact in settings such as postmenopausal endometrial or breast cancer. Furthermore, significant androgen excess is observed in conditions with abnormally increased androstenedione formation (e.g., polycystic ovary syndrome).
Figure 16-38 Hormonal monitoring in clomiphene citrate-initiated ovulation: use of the triple-7 regimen. Please see text. 
(From Adashi EY. Clomiphene citrate-initiated ovulation: a clinical update. Semin Reprod Endocrinol 1986; 4:255276.)
Figure 16-39 Variation of the duration of the menstrual cycle in women with regular cycles. (From Cunningham FG, MacDonald PC, Gant NF, et al. The endometrium and decidua: menstruation and pregnancy. In Williams Obstetrics, 19th ed. Stamford, Conn, Appleton & Lange, 1993, pp 81109.)
Figure 16-40 Tissue sources of estrogen in postmenopausal breast cancer. This figure exemplifies the important pathologic roles of extraovarian (peripheral) and local estrogen biosynthesis in an estrogen-dependent disease in postmenopausal women. The estrogen precursor androstenedione (A) originates primarily from the adrenal in the postmenopausal woman. Aromatase expression and enzyme activity in extraovarian tissues such as fat increase with advancing age. The aromatase activity in skin and subcutaneous adipose fibroblasts gives rise to formation of systemically available estrone (E₁) and to a smaller extent estradiol (E₂). The conversion of circulating A to E₁ in undifferentiated breast adipose fibroblasts compacted around malignant epithelial cells and subsequent conversion of E₁ to E₂ in malignant epithelial cells provide high tissue concentrations of E₂ for tumor growth. The clinical relevance of these findings is exemplified by the successful use of aromatase inhibitors to treat breast cancer.
Figure 16-41 Regimens of hormone replacement therapy (HRT). Estrogen (E) is replaced in a postmenopausal woman to prevent osteoporosis, urogenital atrophy, and hot flashes. In the postmenopausal woman with a uterus, a progestin (P) is added to estrogen to prevent endometrial hyperplasia and cancer. E and P can be administered in a number of ways. A and B. The postmenopausal women receiving hormone replacement have predictable withdrawal bleeding episodes after each P course. C. These women take E and P together continuously. After a year of continuous combination therapy, the rate of unpredictable breakthrough spotting is 20%.
Figure 17-1 The growth of the world population (billion) and increments by decades (million).  () Publication of Malthus' essay on population. Since then, the world population has increased six times.  () Publication of Paul Ehrlich's *The Population Bomb*. Since then, the world population has increased from 4 billion to 6 billion. (Modified from Raleigh VS. Trends in world population: how will the millennium compare with the past? Hum Reprod Update 1999; 5:500.)
**Figure 17-2** The population growth in individual continents. *(From United Nations Population Division. World Population Prospects: The 2000 Revision. New York, United Nations, 2000.)*
**Figure 17-3** Birth rates from 1950 to 2000 and projections for the first half of the 21st century. *(From United Nations Population Division. World Population Prospects: The 2000 Revision. New York, United Nations, 2000.)*
Figure 17-4 The decrease of the birth rate is inversely proportional to the percentage of the population practicing contraception and positively related to the standard of living. (Modified from Potts M. The unmet need for family planning. Sci Am 2000; 282[1]:69.)
Figure 17-5 Intended and unintended pregnancies in the United States, 1994. (Data adapted from Henshaw SK. Unintended pregnancy in the United States. Fam Plann Perspect 1998; 30:2429, 46.)
Figure 17-6  Attachment of the ethinyl group to C-17 of estradiol creates an orally highly active estrogen, ethinylestradiol (EE). Mestranol has a methyl group on C-3 of EE. Mestranol is a prohormone because it must be metabolically converted to EE to be able to bind to estrogen receptors. Modification of the estradiol molecule on C-17 provides long-acting 17-cypionate, used in injectable preparations. For numbering of the steroid molecule, see Figure 17-8.
Figure 17-7 Development of norethindrone from testosterone. Splitting off the C-19 radical from the testosterone molecule changes this androgen to a progestagen. Attachment of the ethinyl group to C-17 enhances the progestagenic activity of the compound and makes it orally active. For numbering of the steroid molecule, see Figure 17-8.
Figure 17-8 Contraceptive progestagens are derived from three skeleton structures, pregnane, estrane, and gonane (see details in the text). The pregnane molecule shows the numbering system of contraceptive steroids.
Figure 17-9 Classification of contraceptive steroids. Hybrid progestagens are in gray boxes. In parentheses are the active progestagenic metabolites of norgestimate and desogestrel.
**Figure 17-10** Progestagens derived from progesterone. Progesterone loses its activity when a hydroxyl group is attached to C-17. Formation of an acetate restores the progestational activity, which is further enhanced by manipulations at C-6. Derivatives of hydroxyprogesterone acetate are potent progestagens as well as antiandrogens.
Figure 17-11 Norethindrone (NET) and its derivatives. In order to become biologically active, the individual derivatives must be converted into NET.
Figure 17-12 Early gonanes: levonorgestrel and norgestrel.
Figure 17-13 Progestagens of the advanced gonane group. Norgestimate and desogestrel are prohormones, metabolically converted into the active progestagenic substances, norelgestromin and etonogestrel, respectively. Gestodene is active without metabolic conversion.
Figure 17-14 Hybrid progestagens: nonpregnanes nomegestrol and nesterone, dienogest, and drospirenone. Drospirenone is a spironolactone derivative.
**Figure 17-15** Relative binding affinities of contraceptive progestagens for progesterone receptors. The assay measures displacement of $^3$H-labeled R5020 from progestagen receptors isolated from the rabbit uterus. The $[^3]$H$^3$H]R5020 is a radiolabeled synthetic progestagen used in in vitro studies. (Data from Phillips A, Demarest K, Hahn DW, et al. Progestational and androgenic receptor binding affinities and in vivo activities of norgestimate and other progestins. Contraception 1990; 41:399.)
Figure 17-16 Relative binding affinity of contraceptive progestagens for androgen receptors. The assay measures displacement of $^{3}$H-labeled dihydrotestosterone from rat prostatic androgen receptors. (Data from Phillips A, Demarest K, Hahn DW, et al. Progestational and androgenic receptor binding affinities and in vivo activities of norgestimate and other progestins. Contraception 1990; 41:399.)
Figure 17-17 Contrasting effects of oral contraceptives containing levonorgestrel (LNG) and norgestimate (NGM) on high-density and low-density lipoproteins (HDL and LDL). (Modified from Henzl M. Norgestimate: from the laboratory to three clinical indications. J Reprod Med 2001; 46:647-661.)
Figure 17-18 Menstrual patterns during 5 years of using contraceptive hormonal implant with levonorgestrel. (From Speroff L, Darney PD. A clinical guide for contraception, 3rd ed. Philadelphia: Lippincott Williams & Wilkins, 2001.)
Figure 17-19 The structure of mifepristone. The major modification of the 19-nor steroid molecule is on C-11.
**Figure 17-20** Types of intrauterine devices (IUDs). The copper T380 and the levonorgestrel-releasing IUD are clinically used in the United States. The frameless copper IUD is available in Europe.
Figure 18-1 Schematic diagram of sexual differentiation beginning with formation of the bipotential gonads from intermediate mesoderm, differentiation of the testes and ovaries from the bipotential gonads, and formation of the sexual phenotypes. The genes known to be involved and their presumed sites of action are indicated. 
5-red, steroid 5-reductase 2; AMH, antimüllerian hormone; AMHR, antimüllerian hormone receptor; AR, androgen receptor; DAX-1, dosage-sensitive sex reversaladrenal hypoplasia congenital gene; DHT, dihydrotestosterone; SF-1, steroidogenic factor 1; SOX3, SRY-related gene HMGbox 3; SOX9, SRY-related gene HMGbox 9; SRY, sex-determining region of the Y chromosome; WT1, Wilms' tumorrelated gene 1.
Figure 18-2 Germ cells. A, Schematic drawing of a 3-week-old embryo showing the site of origin of germ cells in the wall of the yolk sac. B, Migration path of primordial germ cells along the wall of the yolk sac and along the dorsal mesentery into the genital ridge. (From George FW, Wilson JD. Embryology of the genital tract. In Walsh PC, Gittes RF, Perlmutter AD, et al [eds]. Campbell's Urology, 5th ed. Philadelphia, WB Saunders, 1986, pp 1804-1818.)
Figure 18-3 Descent of the testis. (From George FW, Wilson JD. Embryology of the genital tract. In Walsh PC, Gittes RF, Perlmutter AD, et al [eds]. Campbell's Urology, 5th ed. Philadelphia, WB Saunders, 1986, pp 18041818.)
Figure 18-4 Diagram of the Sertoli cell showing the relation between Sertoli cell cytoplasm and developing spermatocytes.
Figure 18-7 Pathways of testosterone synthesis in human testis. Cholesterol for testosterone synthesis can arise from three sources: (1) plasma cholesterol, (2) cholesterol newly synthesized within the Leydig cell, and (3) cholesterol stored as cholesterol esters in Leydig cells. The delivery of cholesterol to the inner mitochondrial membrane for side-chain cleavage to pregnenolone is the rate-limiting reaction and under control of luteinizing hormone. Conversion of pregnenolone to testosterone can take place by two theoretical pathways—one in which side-chain cleavage and reduction of the 17-keto group are accomplished before A-ring oxidation and the other in which this sequence is reversed. As indicated by bold arrows, the former pathway is predominant in the human. DHEA, dehydroepiandrosterone; HSD, hydroxysteroid dehydrogenase; StAR, steroidogenic acute regulatory protein. (Modified from Griffin JE, Wilson JD. The testis. In Bondy PK, Rosenberg LE [eds]. Metabolic Control and Disease, 8th ed. Philadelphia, WB Saunders, 1980, pp 1535-1578.)
Figure 18-9 Metabolism of plasma testosterone in extraglandular tissues. Testosterone can be metabolized to either active or excretory metabolites. Active metabolites such as dihydrotestosterone may be further metabolized to excretory metabolites. HSD, hydroxysteroid dehydrogenase. (From Griffin JE, Wilson JD. The testis. In Bondy PK, Rosenberg LE [eds]. Metabolic Control and Disease, 8th ed. Philadelphia, WB Saunders, 1980, pp 1535-1576.)
Figure 18-10 Schematic diagram of androgen action. Testosterone, secreted by the testis, binds to the androgen receptor in a target cell, either directly or after conversion to dihydrotestosterone. Dihydrotestosterone binds more tightly than testosterone. The major actions of androgens, shown on the right, are mediated by testosterone (solid lines) or by dihydrotestosterone (broken lines). (From Griffin JE. Androgen resistance: the clinical and molecular spectrum. N Engl J Med 326:611618, 1992. Copyright 1992, Massachusetts Medical Society. All rights reserved.)
Figure 18-11 Schematic diagram of the normal androgen receptor.
Figure 18-12 Cell divisions during spermatogenesis. The overall number of cell divisions is much higher than that during oogenesis.
Figure 18-13 Schematic diagram illustrating conversion of spermatocyte to spermatid to spermatozoon.
Figure 18-14 Schematic diagram of the different phases of male sexual function during life as indicated by mean plasma testosterone level and sperm production at different ages. (From Griffin JE, Wilson JD. The testis. In Bondy PK, Rosenberg LE [eds]. Metabolic Control and Disease, 8th ed. Philadelphia, WB Saunders, 1980, pp 1535-1578.)
Figure 18-15 Ontogeny of luteinizing hormone (LH) secretion. Plasma LH concentrations were sampled every 20 minutes for 24 hours in three normal males at different stages of development. Top, Pattern in an adult man with frequent secretory episodes throughout the 24-hour period and no significant sleep-related augmentation. Middle, Secretory pattern in midpuberty in which marked secretory episodes occur during sleep. Bottom, Pattern in prepuberty in which there are no significant secretory episodes at any time throughout the sampling period. (From Griffin JE, Wilson JD. The testis. In Bondy PK, Rosenberg LE [eds]. Metabolic Control and Disease, 8th ed. Philadelphia, WB Saunders, 1980, pp 1535-1578. Courtesy of R.M. Boyar.)
Figure 18-17 Diagram of sequence of events at puberty. The ranges of age during which changes occur in normal boys are indicated by figures below each bar.  (Data from Marshall WA, Tanner JM. Variations in the pattern of pubertal changes in boys. Arch Dis Child 1970: 45:1323.)
Four different patterns of abnormal androgen-estrogen dynamics can result in the development of gynecomastia. The altered component in each pattern is highlighted in black, and specific examples of each type of abnormality are listed at the bottom of each panel. Details of normal androgen-estrogen dynamics are shown in Figure 18-8 (Figure Not Available). HCG, human chorionic gonadotropin; HSD, hydroxysteroid dehydrogenase.
Figure 18-19 Types of androgen preparations available for clinical use. Type A derivatives are esterified in the 17 position. Type B steroids have alkyl substitutions in the 17 position. Type C derivatives involve a variety of alterations of ring structure that enhance activity, impede catabolism, or influence both functions. Most androgen preparations involve combinations of type AC or type BC changes.
Figure 18-20 Effect on penile size of 200 mg of testosterone cypionate intramuscularly every 2 weeks for 11 months in a previously untreated 22-year-old man with microphallus caused by hypogonadotropic hypogonadism. (From Griffin JE, Wilson JD. Disorders of sexual differentiation. In Walsh PC, Retik AB, Stamey TA, et al [eds]. Campbell’s Urology, 6th ed. Philadelphia, WB Saunders, 1992, pp 1509-1542.)
Figure 19-2 Biochemical mechanisms of penile smooth muscle relaxation. A, Relaxation of the cavernosal smooth muscle is regulated by intracellular cyclic adenosine monophosphate (cAMP) and cyclic guanidine monophosphate (cGMP). The intracellular second messengers, by activation of specific protein kinases, cause sequestration of intracellular calcium (Ca$^{2+}$), closure of calcium channels, and opening of potassium (K$^+$) channels. This results in a net decrease in intracellular calcium, causing smooth muscle relaxation. Nitric oxide, released by noradrenergic, noncholinergic nerve endings (NANC), stimulates guanyl cyclase. By inhibiting phosphodiesterase type 5 (PDE5), sildenafil increases the amount of intracellular cGMP. Prostaglandin E$_1$ (PGE$_1$) stimulates generation of cAMP. Papaverine inhibits PDE$_2$, PDE$_4$, and PDE$_5$ and thereby increases the amount of intracellular cAMP. AMP, adenosine monophosphate; GTP, guanosine triphosphate; IP$_3$, inositol 1,4,5-triphosphate; NO, nitric oxide; PKA, protein kinase A; PKG, cGMP-specific protein kinase. B, Interconnection of cavernosal smooth muscle cells in the penis. The smooth muscle cells in the corpora cavernosa are interconnected through connexin 43-derived gap junctions. Therefore, alterations in action potential and potassium channel activity in any myocyte affect the adjacent myocytes. (A and B, Adapted and redrawn from Melman A, Christ GJ. Integrative erectile biology: the effects of age and disease on gap junctions and ion channels and their potential value to the treatment of erectile dysfunction. Urol Clin North Am 2001; 28:28:217231.)
Figure 19-3 Arterial supply of the perineum and clitoris in prearousal (A) and arousal (B) states.
Figure 20-1 A, Serum thyrotropin (hTSH) and human chorionic gonadotropin (hCG) concentrations throughout pregnancy. Between 8 and 14 weeks of gestation, there is a significant negative correlation between the individual hTSH and hCG levels (P < .001). Each point represents the mean (±SE). SE, standard error. B, Linear regression of maternal serum free thyroxine (T₄) and hCG concentrations during the first half of gestation (P < .001). (From Glinoer D, de Nayer P, Bourdoux P, et al. Regulation of maternal thyroid during pregnancy. J Clin Endocrinol Metab 1990; 71:276.)
Figure 20-2 Mean plasma concentrations of estrone (E₁), estradiol (E₂), estriol (E₃), and progesterone (P) during pregnancy. (Data from Tulchinsky D, Hobel CJ, Yeager E, et al. Plasma estrone, estradiol, estriol, progesterone, and 17-hydroxyprogesterone in human pregnancy: I. Normal pregnancy. Am J Obstet Gynecol 1972; 112:1095–1096; and Levitz M, Young BK. Estrogens and pregnancy. Vitam Horm 1977; 35:109.)
Figure 20-3 Steroidogenesis in the maternal-fetal-placental unit. SCC, cholesterol side-chain cleavage enzyme; DHEA, dehydroepiandrosterone; HSD, hydroxysteroid dehydrogenase; AROM, aromatase-enzyme complex.
Figure 20-4 Mean (±SE) maternal serum human chorionic gonadotropin (hCG) levels throughout normal pregnancy. (From Braunstein GD, Rasor J, Danzer H, et al. Serum human chorionic gonadotropin levels throughout normal pregnancy. Am J Obstet Gynecol 1976; 126:678.)
Figure 20-5 Proposed pathways for metabolism of human chorionic gonadotropin: , intact hCG; : N, hCG with nicked subunit; N, free nicked subunit; CTP fragment, carboxy-terminal fragment; mRNA, messenger RNA. (From Braunstein GD. Physiologic functions of human chorionic gonadotropin during pregnancy. In Mochizuki M, Hussa R [eds]. Placental Protein Hormones. Amsterdam, Elsevier Science, 1988, p 33.)
Figure 20-6  Placental weight (Pl. wt.) and maternal serum concentrations of human placental lactogen (hPL) during pregnancy.  (From Selenkow HA, Saxena BN, Dana CL. Measurement and pathophysiologic significance of human placental lactogen. In Pecile A, Finzi C [eds]. The Feto-Placental Unit. Amsterdam, Excerpta Medica, 1969, p 340.)
Figure 20-7. Mean (± standard error) of plasma human growth hormone (hGH) (A) and insulin-like growth factor-1 (IGF-1) (B) levels throughout pregnancy. The number of individual assays of growth hormone (GH) and IGF-1 at each gestational stage is indicated in A on top of vertical bars. GH 5 B4 indicates placental GH (hGH-V); GH 5 24 indicates pituitary GH. (From Mirlesse V, Frankenne F, Alsat E, et al. Placental growth hormone levels in normal pregnancy and in pregnancies with intrauterine growth retardation. Pediatr Res 1993; 34:39.)
Figure 21-1 Diagrammatic representation of a chorionic villus extending into the maternal blood lake and showing fetal capillaries in the fetal mesenchyme. The villus is sheathed by the syncytiotrophoblast. The residual sparse areas of cytotrophoblast provide cells to renew and maintain the syncytiotrophoblast layer. The villus is surrounded by maternal blood in the maternal intervillous space. The placenta serves as an important endocrine organ. Hormones are produced by cytotrophoblast and syncytiotrophoblast cells. Neuropeptides appear to modulate syncytiotrophoblast production of placental protein hormones, and decidual prostaglandins and cytotrophoblast growth factors may participate in regulation of syncytiotrophoblast steroid hormone production. See text for details.
Figure 21-2 Diagrammatic representation of the fetoplacental unit composed of the fetal adrenal cortex and the placenta. The placenta is deficient in 17-hydroxylase activity and cannot synthesize estrogens from progesterone. The fetal adrenal cortex has low 3-hydroxysteroid dehydrogenase (3HSD) and 17β isomerase activity and cannot synthesize progesterone. Sulfokinase activity is high in fetal adrenal tissue, and steroid sulfatase activity is high in placental tissue. Thus, the placenta produces progesterone, which is predominantly converted to dehydroepiandrosterone (DHEA) by the fetal adrenal cortex; the DHEA can be sulfated to form DHEA sulfate (DHEAS). Part of this is 16-hydroxylated by the fetal liver, and both DHEA and DHEAS are used by the placenta as substrates for estrone (E₁) and estradiol (E₂) synthesis, respectively. Placental sulfatase converts DHEAS and 16-hydroxy-DHEAS to DHEA and 17-hydroxy-DHEA. The 16-hydroxy-DHEA is used for estriol (E₃) synthesis. See text and references for details.
Figure 21-3 Cartoon illustrating the homeobox genes programming hypothalamic and pituitary embryogenesis and function. SHH and ZIC2 are the sonic hedgehog and Drosophila odd-paired homologues, mutations of which have been shown to cause human holoprosencephaly. RPX is the Rathke’s pouch homeobox gene involved in anterior pituitary gland embryogenesis. LHX3 and LHX4 are LIM class homeodomain transcription factors also essential for normal pituitary embryogenesis. PROP1 and PIT1 defects in mice and humans lead to growth hormone (GH), prolactin (PRL), and thyroid-stimulating hormone (TSH) deficiency. FSH, follicle-stimulating hormone; LH, luteinizing hormone; POMC, pro-opiomelanocortin. See text for details.
Figure 21-4 Patterns of change of fetal plasma human placental lactogen (hPL), growth hormone (GH), prolactin (PRL), insulin-like growth factor I (IGF I), and insulin-like growth factor II (IGF II) during gestation and in the neonatal period. The range of fetal plasma hPL concentrations is shown as the hatched area. (Data from Bennett A, Wilson DM, Liu R, et al. J Clin Endocrinol Metab 1983; 57:609612; Kaplan SL, Grumbach MM, Aubert ML. Recent Prog Horm Res 1976; 32:161243; Bala RM, Lopatka J, Leung A, et al. Clin Endocrinol Metab 1981; 52:508512.)
Figure 21-5 Hemi-cross-section of a 5-week human embryo with location of the adrenal primordia (suprarenal cortices) and gonadal ridges. The homeobox genes programming adrenal and gonadal embryogenesis are indicated. SF1 (steroidogenic factor 1) is involved in testicular and ovarian development, whereas SRY is the single critical regulator of testicular embryogenesis. Inactivation of the DAX1 gene leads to adrenal hypoplasia. The steroidogenic acute regulatory protein (StAR) is the rate-limiting factor for adrenal steroidogenesis. See text for details.
Figure 21-7 Cartoon showing the homeobox genes programming development of the thyroid and parathyroid glands. HOXB3 may be responsible for activation of thyroid transcription factor 1 (TTF1) during early embryogenesis, with TTF2 and PAX8 involved in a synergistic cascade programming thyroid gland embryogenesis. These factors are also involved in thyroid follicular cell function, promoting thyroglobulin (TG), thyroid peroxidase (TPO), and thyroid-stimulating hormone receptor (TSHR) gene transcription. HOX15 gene knockout in mice causes parathyroid gland aplasia. See text for details.
Figure 21-8 Patterns of change of fetal plasma thyroid-stimulating hormone (TSH), thyroxine (T₄), triiodothyronine (T₃), reverse T₃ (rT₃), and iodothyronine sulfate (T₄S, rT₃S, and T₃S) levels during gestation and in the neonatal period. The patterns for T₄S and rT₃S are based on limited 30-week data. (Data from Fisher DA, Klein AH. N Engl J Med 1981; 304:702712; Santini F, Chiovato L, Ghirri P, et al. J Clin Endocrinol Metab 1999; 84:493498; Burrow GN, Fisher DA, Larsen PR. N Engl J Med 1994; 331:10721078.)
Figure 21-9 Patterns of change of plasma levels of human chorionic gonadotropin (hCG), luteinizing hormone (LH), testosterone (T), and estradiol (E₂) in a male fetus during gestation and in the neonatal period. (Data from Reyes FI, Boroditsky RS, Winter JS, et al. J Clin Endocrinol Metab 1974; 38:612617; Kaplan SL, Grumbach MM, Aubert ML. Recent Prog Horm Res 1976; 32:161243; Winter JS, Faiman C, Hobson WC, et al. J Clin Endocrinol Metab 1975; 40:545551; Forest MG, Cathiard AM. J Clin Endocrinol Metab 1975; 41:977980; and Penny R, Parlow AF, Frasier SD. Pediatrics 1979; 64:604608.)
**Figure 21-10** Proposed actions of parathyroid hormone (PTH), PTH-related protein (PTHrP), and calcitonin (CT) in the fetus. PTHrP and perhaps PTH from the parathyroid glands and PTHrP from the placenta act on the placenta to promote calcium (Ca) and phosphate (PO$_4$) transport from the maternal to the fetal circulation to maintain the relative fetal hypercalcemia and the high rate of fetal bone formation during the last half of gestation. PTHrP also acts on the kidney to promote 1-hydroxylation of 25-hydroxycholecalciferol to 1,25-dihydroxyvitamin D, 1,25(OH)$_2$D, which augments placental calcium transport and promotes fetal bone growth. High fetal CT levels tend to promote bone accretion. See text for details.
Figure 21-11 Cartoon showing the homeobox genes programming development of the pancreas. IDX1 gene knockout in the mouse leads to pancreatic agenesis, whereas ISL1 knockout produces islet cell agenesis. Knockout of PAX4 and PAX6 leads to beta cell or alpha cell agenesis or hypogenesis; Beta2 gene disruption also produces beta cell hypoplasia. See text and Koshimizu et al. J Clin Endocrinol Metab 1995; 61:7882.
Figure 21-12 Actions of cortisol and catecholamines during fetal adaptation to the extrauterine environment. The prenatal cortisol surge acts to promote functional maturation of several organ systems as indicated. The neonatal catecholamine surge triggers or potentiates a number of the extrauterine cardiopulmonary and metabolic functional adaptations that are critical to extrauterine survival. See text for details. BAT, brown adipose tissue; E, epinephrine; NE, norepinephrine; T3, triiodothyronine; T4, thyroxine.
Figure 22-1 Typical G-banded karyotypes of patients with abnormal gonadal differentiation.  A, The 45,X karyotype of a patient with streak gonads, short stature, and physical stigmata of Turner's syndrome.  B, The 47,XXY karyotype of a phenotypic male with seminiferous tubule dysgenesis (chromatin-positive Klinefelter's syndrome).
Figure 22-2 A partial karyotype of C group (chromosome numbers 6 to 12) and X and Y in a patient with a 46,X,i(Y;7)(q11;q36) karyotype. Standard Giemsa staining, autoradiography, fluorescent (Q), and Giemsa (G) banding techniques were used to identify the chromosome anomaly. A, The standard staining technique for karyotype analysis revealed an enlarged C group chromosome and a deleted G group chromosome. B, Autoradiography after incubation of lymphocyte culture with tritiated thymidine showed a late-labeling segment on the distal arms of the C chromosome and absence of a late-labeling segment on the deleted long arm of the presumptive Y. C, Quinacrine hydrochloride staining and fluorescence microscopy demonstrated a translocation of the brightly fluorescent segment of the long arm of the Y chromosome to the long arm of chromosome 7. D, Giemsa banding confirmed that the C group chromosome involved in the translocation was chromosome 7.
**Figure 22-3** FISH for SRY analysis in metaphase and interphase 46,XY cells. FISH images illustrating localization of the Vysis SRY probe on the distal short arm of the Y chromosome (Yp11.3) shown in SpectrumOrange and Vysis CEPX, a probe for the X centromere shown in SpectrumGreen. Note the localization of the X (green) and Y (orange) in an interphase nucleus (right). (Courtesy of Philip Cotter and Helen Jenks.)
Figure 22-4 Daughter cell lines can arise from mitotic nondisjunction or anaphase lag during first mitotic division in the zygote. More complex mosaicism can result if the zygote is aneuploid or if replication errors arise beyond the one cell stage. In females, nondisjunction or anaphase lag may involve either the maternal or paternal X chromosome. Deductions regarding the origin of X chromosomes in aneuploid patients can be made by correlating sex-linked traits with those in parents and by using specific DNA probes for analysis.
Figure 22-5 A diagram of chromosome breakage and recombination to form long and short arm deletions and ring chromosomes. Deleted segments may also be transposed to terminal portions of other chromosomes as additions, or there may be reciprocal translocations of deleted segments with those from another chromosome.
Long arm isochromosomes of the X,Xqi, were postulated to result from centric fission, that is, transverse rather than longitudinal division of the centromere. A more likely mechanism is shown. A deletion occurs at the centromere on the short arm or above the centromere. Fusion of remaining chromatids of the short arm followed by division of the centromere and duplication of entire chromatid and centromere(s) results in an isochromosome with either one or two centromeres. The acentric fragment is lost.
Figure 22-7 Diagram of a G-banded Y chromosome. Y-linked genes are shown. SHOX/PHOG, short stature/pseudoautosomal homeobox-containing osteogenic gene on the X; MIC2, a cell-surface antigen recognized by the monoclonal antibody 12E7; SRY, sex-determining region Y; RPS4Y, ribosome protein S4Y; ZFY, zinc finger Y; TSPYA, TSPYB, testes-specific protein Y; PRKY, a member of the cyclic adenosine monophosphate-dependent serine threonine protein kinase gene family, homologous to PRKX. DAZ, deleted in azoospermia; AZF, azoospermic factor.
**Figure 22-8** A, Q staining and fluorescence microscopy of interphase cells from a normal male, illustrating typical Y bodies. B, An enlarged photograph of one cell, showing a fluorescent Y body at the periphery of the nucleus. C, Metaphase chromosomes from a normal male, illustrating the brightly fluorescent distal segment of the long arm of the Y chromosome. D, An interphase nucleus in a buccal smear of a patient with a 47,XXY karyotype. A brightly fluorescent Y body and an X chromatin body (which exhibits much weaker fluorescence) were identified by Q staining and fluorescence microscopy.
Figure 22-9 Diagram of a G-banded X chromosome. Selected X-linked genes are shown. SHOX/PHOG, short stature/pseudoautosomal homeobox-containing osteogenic gene; MIC2, a cell-surface antigen recognized by monoclonal antibody 12E7; PRKX, a member of the cyclic adenosine monophosphate-dependent serine threonine protein kinase gene family. Illegitimate X-Y interchange occurs frequently between PRKX and PRKY. DAX1, dosage-sensitive sex reversal congenital adrenal hypoplasia critical region on the X chromosome-1; GK, glycerol kinase; DMC, Duchenne's muscular dystrophy; USP9X, human X-linked homologue of the DFFRX (Drosophila fat facets related X gene); RPS4X, ribosomal protein S4X; Xist, Xi specific transcripts; XIC, X-inactivation center; ATRX, -thalassemia, X-linked mental retardation; DIAP2, human homologue of the Drosophila diaphanous gene, mutations of which affect oogenesis and spermatogenesis.
Figure 22-10  A and B, Photomicrographs showing the X chromatin body (Barr body, arrow) in the nucleus of buccal mucosal cells from a normal female (thionine stain, ×2000). Such cells are found in about 25% of well-preserved nuclei.  C, A buccal mucosal cell from a normal male, illustrating absence of this body (thionine stain, ×2000).  D, A typical “drumstick” nuclear appendage (arrow) found in a variable proportion of leukocytes of female subjects.
Figure 22-11 Characteristics of heterochromatin formation as exemplified by differential behavior of the two X chromosomes of the female in somatic cells. 1. Precocious condensation of a large part of one of the two X chromosomes in prophase and formation of the X chromatin body in interphase nuclei. 2. Delayed replication of DNA in one of the X chromosomes (arrow indicates silver grains overlying one X chromosome in the autoradiogram of metaphase chromosomes from a normal female exposed to tritiated thymidine late in the synthetic period). With some exceptions (PAR region, etc.) gene activity on the heterochromatic late-replicating X chromosome is silenced or modified. (From Grumbach MM. On the significance of sex chromatin. In Second International Conference on Congenital Malformations. New York, International Medical Congress, 1964, pp 6267.)
Figure 22-12 Diagram of the fixed differentiation or Lyon hypothesis of X chromosome behavior in somatic cells of the human female. At the late blastocyst stage (the time when X chromatin can first be identified), one of the two X chromosomes becomes heterochromatinized in each cell and gives rise to an X chromatin body; it is by chance in each cell whether this differentiation involves a maternally derived X (M) or a paternally derived X (P). Once differentiation has occurred, this characteristic is fixed in succeeding generations of somatic cells. Most of the genes on the heterochromatic portion of an X chromosome are suppressed or inactivated, thus serving as a means of "dosage compensation" for the greater number of X-linked genes in the female than in the male. This mechanism has an important bearing on expressivity and penetrance of an X-linked mutant gene in a heterozygous female. In the diagram, the maternally derived X carries a mutant gene (a) that is expressed only in cells in which this X is the isopyknotic, euchromatic active X (white X_M). Although the heterochromatinized X (black X) in this diagram is represented as wholly inactive, some loci on the heterochromatinized X do remain active and exert genetic effects. The female germ cell line beyond the oogonia stage is exempted from heterochromatinization.
Figure 22-13 Diploid somatic cells from a girl with a 49,XXXX karyotype. A, Four X chromatin bodies (arrows) in an interphase nucleus from a culture of skin fibroblasts. B, Autoradiogram of metaphase chromosomes, illustrating four areas of high grain density overlying four of the five X chromosomes. C, An autoradiogram of an interphase nucleus in a culture of skin fibroblasts; four peripheral "hot" areas (arrows) of high grain density overlie four X chromatin bodies and provide direct evidence that each X chromatin body is derived from one late-labeling X chromosome. (Modified from Grumbach MM, Morishima A, Taylor JA. Human sex chromosome abnormalities in relation to DNA replication and heterochromatization. Proc Natl Acad Sci USA 1963; 49:581589; and Grumbach MM. On the significance of sex chromatin. In Second International Congress on Congenital Malformations. New York: International Medical Congress, 1964, pp. 62-67.)
Figure 22-14 Diagram of the historical search for the testis-determining factor. The shaded area on the Y chromosome is the region to which this factor has been localized. ZFY, zinc finger Y; SRY, sex-determining region Y; numbers 1 to 4A indicate arbitrary deletion segments on the Y chromosome. (Modified from McLaren A. What makes a man a man? Reprinted by permission from Nature, vol. 346, pp. 216-217. © 1990 Macmillan Magazines Ltd.)
Figure 22-15 Localization of the putative sex-determining region, SRY, on the short arm of the Y chromosome. The zinc finger locus ZFY (the suggested site of the testis-determining factor in 1987) is shown, as well as the break points observed in four 46,XX males described by Palmer and co-workers. The break points of one 46,XX male and one 46,XY female studied by Page and colleagues are also indicated. Note that the 46,XY female has a noncontinuous deletion that involves both ZFY and SRY. (From Page DC, Fisher EMC, McGillivray B, et al: Additional deletion in the sex-determining region of the human Y chromosome resolves the paradox of X[t(y;22)] female. Nature 1990; 346:279-281.)
**Figure 22-16** Diagram of the human SRY protein. The HMG box is an 80 amino acid DNA-binding domain with two nuclear localization signals at either end: CaM, calmodulin and imp, importen. The last seven amino acids of SRY can bind to either of the PDZ domains found in SRY-interacting protein 1 (SIP-1). The solid circles indicate selected mutations reported in the SRY protein affecting testicular differentiation and consequent male development.
Figure 22-17 Diagram of the SOX9 protein. Unlike SRY, SOX9 has two introns. Like SRY it has two nuclear localization signals at the ends of the HMG box. SOX9, unlike SRY, binds to heat shock protein 70 (HSP70) and has a trans-activation domain at the carboxyl-terminal end. Selected mutations causing sex reversal and campomelic dysplasia are indicated by the solid circles, those causing only campomelia are indicated by the open triangles, whereas SRY mutations occur primarily on the HMG box. SOX9 mutations appear to occur throughout the gene.
Figure 22-18  Diagram of the SF1 protein. Mutations causing SF1 deficiency are shown. Gly35Glu (P box) and Arg255Leu are heterozygous mutations in a male and a female. Arg92Gln (A-Box) is a homozygous mutation in a male; heterozygotes are unaffected. P box, proximal box; D box, distal box; A-box is the FTZ-F1 box; PRR is a proline-rich region; regions II and III are conserved regions among SF1 proteins; AF-2 is an activation domain.
Figure 22-19 A and B, Hypothetical linear cascade and network of genes involved in human sex determination and differentiation (refer to manuscript for genes described in the mouse). WT1, Wilms' tumor suppressor gene; SF1, steroidogenic factor 1; DAX1, dosage-sensitive sex reversal adrenal hypoplasia critical region on the X gene 1; SOX9, autosomal gene containing SRY-like HMG box; SRY, sex-determining region Y; DMRT1, double sex, mab3-related transcription factor 1; WNT4, a member of the vertebrate homolog family of the Drosophila segment polarity gene, "wingless"; AMH, antimüllerian hormone; GATA4, transcription factor.
Types of cell division. A female somatic cell undergoing mitosis is represented. At the metaphase plate are two X chromosomes and two homologous autosomes of group 21 to 22. Division occurs through the centromere, giving rise to two daughter cells of identical chromosomal composition. Replication of each arm into two chromatids takes place while the chromosomes are extended and before the next metaphase. The first meiotic division involves pairing of homologous chromosomes. The centromere does not divide in this cell division. It is by chance whether the maternal (X<sup>M</sup>) or paternal (X<sup>P</sup>) member of each pair goes to the respective daughter cells. During the complex prophase of first meiotic division (not shown), multiple chiasmata are formed between the chromosomes of each pair, facilitating exchanges of chromosomal segments (crossing over) between them. During the second meiotic division, the centromere again divides, giving rise to daughter cells identical to the parent cell. This division more nearly resembles mitosis than the first meiotic division. Nondisjunction can take place in mitosis or in the first or second meiotic division; representative examples are illustrated.
Figure 22-21 Anatomic and schematic representations of gonadal differentiation. A and B. Transverse section through the urogenital ridge at the stage of the indifferent gonad. Note the proximity of a large fetal adrenal to the hilar portion of gonad. C and D. Transverse section through the fetal testis at 56-mm stage. E and F. Transverse section through the fetal ovary at 60-mm stage. In ovarian development, coelomic epithelium continues to proliferate for a much longer period. (Modified from Arey LB. Developmental Anatomy, 7th ed. Philadelphia, WB Saunders, 1965; and Witschi E. Development of Vertebrates. Philadelphia, WB Saunders, 1956.)
Figure 22-22 Comparison of the pattern of change of serum testosterone and human chorionic gonadotropin (hCG) and serum and pituitary luteinizing hormone (LH) (LER960) and follicle-stimulating hormone (FSH) (LER869) in the human male fetus during gestation with morphologic changes in the fetal testis. (Adapted from Kaplan SL, Grumbach MM. Pituitary and placental gonadotropins and sex steroids in the human and subhuman primate fetus. Clin Endocrinol Metab 1978; 7:487-511.)
**Figure 22-23** Comparison of the pattern of serum follicle-stimulating hormone (FSH), luteinizing hormone (LH), and human chorionic gonadotropin and pituitary FSH and LH in the human female fetus during gestation with the developmental histology of the fetal ovary. (Adapted from Kaplan SL, Grumbach MM. Pituitary and placental gonadotropins and sex steroids in the human and subhuman primate fetus. Clin Endocrinol Metab 1978; 7:487511.)
Figure 22-24 The sequence of sexual differentiation in the human fetus. The sequence as schematically depicted here emphasizes that testicular development in the male fetus precedes all other forms of sexual dimorphism. There is an inherent propensity of the gonads, genital ducts, and external genitalia to feminize, whereas masculinization requires Y chromosomemediated (SRY) differentiation of the fetal testes. (Modified from Jost A. Hormonal factors in the sex differentiation of the mammalian foetus. Philos Trans R Soc Lond [Biol] 1970; 259:119130.)
Figure 22-25 Embryonic differentiation of male and female genital ducts from wolffian and müllerian primordia. Left, An indifferent stage showing large mesonephric body. Middle, Female ducts. Remnants of mesonephros and wolffian ducts are now termed the epoöphoron, paroöphoron, and Gartner duct. Right, Male ducts before descent into scrotum. The only müllerian remnant is the testicular appendix. Prostatic utricle (vagina masculina) is derived from urogenital sinus. (Modified from Corning HK. Lehrbuch der Entwicklungsgeschichte des Menschen. Munich, JF Bergmann, 1921; and Wilkins L. The Diagnosis and Treatment of Endocrine Disorders in Childhood and Adolescence, 3rd ed. 1965. Courtesy of Charles C Thomas, Publisher, Springfield, IL.)
Figure 22-26 A schematic summary of Jost's experiments with rabbit embryos. The fetal testis plays a decisive role in determining the differentiation of the genital ducts. Testosterone stimulates wolffian development but fails to effect involution of müllerian structures. (Data from Jost A. Embryonic sexual differentiation [morphology, physiology, abnormalities]. In Jones HW, Scott WW [eds]. Hermaphroditism, Genital Anomalies and Related Endocrine Disorders, 2nd ed. Baltimore, Williams & Wilkins, 1971, p 16.)
Figure 22-27 Mechanism of action of anti-müllerian hormone (AMH/MIS). AMH/MIS binds to the ubiquitous AMH type I (ALK2) and the müllerian ductspecific AMH type II receptor, forming a heterodimer that induces signaling through the SMAD pathway and binding to DNA, resulting in the production of MMP2, matrix metalloprotein 2 in the müllerian duct mesenchyme. Pro MMP2 is secreted into the extracellular space, where it is activated by a membrane based metalloproteinase. Apoptosis (cell death) may occur as a result of activation of a death factor or inactivation of a survival factor. An alternate hypothesis is that MMP2 acts to cleanse the epithelial cell basement membrane. (Redrawn from Roberts LM, Visser JA, Ingraham HA. Involvement of a matrix metalloproteinase in MIS-induced cell death during urogenital development. Development 2002; 129:14571496).
Figure 22-28 Differentiation of male and female external genitalia from indifferent primordia. Male development occurs only in the presence of androgenic stimulation during the first 12 fetal weeks. (Adapted from Spaulding MH. The development of the external genitalia in the human embryo. Contrib Embryol Carnegie Inst 1921; 13:6968.)
**Figure 22-30** Linear representations of the steroid/thyroid hormone receptor superfamily are shown to illustrate sequence homology. hGR is the glucocorticoid receptor; hMR, the mineralocorticoid receptor; hAR, the androgen receptor; hPR, the progesterone receptor; hER, the estrogen receptor; hERR 1 and hERR 2, estrogen-related receptors; hRAR, the retinoic acid receptor; hTR, the thyroid hormone receptor; hVDR, the vitamin D receptor; and hCOUP, the chicken ovalbumin upstream promoter. The DNA-binding site (region I) and the hormone-binding regions (II and III) are shown. *(From O'Malley B. The steroid receptor superfamily: more excitement predicted in the future. Mol Endocrinol 1990; 4:363369. © by The Endocrine Society.)*
Figure 22-31 Type II zinc fingers. +++ indicates the amino acid skeleton of the zinc fingers, which specifies DNA binding in a sequence-specific manner.
Figure 22-32 A simplified scheme of male sex differentiation. (Modified from Grumbach MM. Genetic mechanisms of sex development. In Vallet HL, Porter IH [eds]. Genetic Mechanisms of Sexual Development. New York, Academic, 1979, pp 3373.)
Figure 22-34 A, A 19-year-old phenotypic male with chromatin-positive seminiferous tubule dysgenesis (Klinefelter's syndrome). The karyotype was 47,XXY, gonadotropin levels were elevated, and testosterone levels were low normal. Note normal virilization with long legs and gynecomastia (B, C). The testes were small and firm and measured 1.8 × 0.9 cm. Testicular biopsy revealed a severe degree of hyalinization of the seminiferous tubules and clumping of Leydig cells. D, A 48-year-old male with 47,XXY Klinefelter's syndrome with severe leg varicosities.
Figure 22-35  A. An 8½-year-old boy with a 48,XXXY chromosome constitution, mental retardation, precocious sexual development, and accelerated growth. The appearance of pubic hair was noted at age 6. By 8 years, acne, a deep voice, tall stature, and axillary hair were present. Height was 148 ± 2.9 cm, weight was 47.7 ± 3.9 kg, span was 140 cm, and upper segment/lower segment ratio was 0.87. Testes measured 2.1 × 1.3 cm. Note the long legs, prognathism, small hands and feet, gynecomastia, and secondary sexual characteristics. His I.Q. was 62. The urinary 17-ketosteroid level was 3.2 mg/day and urinary gonadotropin levels were elevated. Bone age was 13.5 years. Roentgenogram of chest was normal. The buccal smear contained nuclei with a maximum of two X chromatin bodies. Karyotype of cells derived from skin and blood was 48,XXXY. B. Testicular biopsy showed hyalinized tubules and clumping of Leydig cells; germ cells were absent. The findings suggest that true precocious puberty, with stimulation of juvenile testes by pituitary gonadotropin, led to premature appearance of typical histologic changes of seminiferous tubule dysgenesis. (Courtesy of M. M. Grumbach and A. Morishima, unpublished data.)
Figure 22-36 Origin of the 47,XXY karyotype. Superscripts M and P designate, respectively, maternal and paternal X chromosomes. The dashed circle indicates a nonviable cell line. (From Grumbach MM. The testes. In Beeson PB, McDermott W [eds]. Cecil Loeb Textbook of Medicine, 13th ed. Philadelphia, WB Saunders, 1971, pp 16041818.)
Figure 22-37 Diagram of the short arms of the X and Y chromosomes during meiotic pairing. A, A crossover (dashed lines) usually occurs between the pseudoautosomal regions of the X and Y chromosomes. Anomalous but equal crossovers (solid lines) can occur that result in an X chromosome with the sex-determining region (SRY) and a Y chromosome deficient in the SRY. It is estimated to occur at the PRKX and PRKY locus in 40% of SRY+XX males. Zygotes with these sex chromosomes will become XX males or XY females as indicated. B, Anomalous unequal crossovers (solid lines) during male meiosis can result in an X chromosome with an SRY gene as well as the pseudoautosomal regions of both the X and Y chromosomes. SRY, sex-determining region Y; ZFY, zinc finger Y.
Figure 22-38 Variation in physical appearance in five patients with the typical form of the syndrome of gonadal dysgenesis. All of these patients had a 45,X karyotype, and all had differences between height age and chronologic age of 3 years or more. (Modified from Grumbach MM. Some considerations of the pathogenesis and classification of anomalies of sex in man. In Astwood EB [ed]: Clinical Endocrinology. New York, Grune & Stratton, 1960, pp 407-436.)
A patient aged 14 years, 10 months with the typical form of the syndrome of gonadal dysgenesis. The X chromatin pattern was negative and the karyotype was 45,X. She was short (height, 134.5 cm; height age, 9 years, 5 months), was sexually infantile except for the appearance of sparse pubic hair, and exhibited characteristic stigmata of the syndrome: a short webbed neck, shield-like chest with widely separated nipples, bilateral short fourth metacarpals, puffiness over dorsum of fingers, cubitus valgus, and an increased number of pigmented nevi. The facies were characteristic and the ears low set. The bone age was 13.5 years. Plasma gonadotropin levels were elevated. Vaginal smears and urocytogram showed an immature pattern in which cornified squamous cells were absent. With estrogen therapy, female secondary sexual characteristics were induced; cyclic estrogen administration resulted in periodic estrogen-withdrawal bleeding.
Figure 22-40 A and B. An infant with the syndrome of gonadal dysgenesis (karyotype 45,X) and associated lymphedema of extremities. The term Bonnevie-Ullrich syndrome is applied when this characteristic swelling of the feet or hands or both is associated with other features of gonadal dysgenesis. (From Grumbach MM. Chromosomal sex and the prepubertal diagnosis of gonadal dysgenesis. Reproduced by permission of Pediatrics 20:740. Copyright 1957.)
Figure 22-41 A. The mean height (50th percentile) in untreated patients with gonadal dysgenesis (mainly 45,X karyotype), compared with the growth curve of normal females. B. The mean height velocity in patients with gonadal dysgenesis and in normal females. Data derived from various sources. Note the lack of a pubertal growth spurt. (B, Courtesy of J. Frame and K. Attie, Genentech, Inc.)
Figure 22-42. The pattern of plasma follicle-stimulating hormone (FSH) concentration in relation to age in 58 patients with the syndrome of gonadal dysgenesis. , patients with 45,X karyotype; , patients with structural abnormalities of the X chromosome and mosaics. The hatched area shows the mean to the lower limits of the assay for FSH values in normal females. (From Conte FA, Grumbach MM, Kaplan SL. A diphastic pattern of gonadotropin secretion in patients with the syndrome of gonadal dysgenesis. J Clin Endocrinol Metab 40:670674, 1975. © The Endocrine Society.)
Figure 22-43 The range of phenotypic and gonadal expression in variants of the syndrome of gonadal dysgenesis and its relation to sex chromosome constitution. Typical phenotypic and gonadal findings in monosomic 45,X gonadal dysgenesis may be modified by the presence of a mosaic chromosomal constitution or by the presence of a structurally abnormal second sex chromosome. For example, 45,X/46,XX and 45,X/47,XXX mosaicism may be associated with normal stature, minimal somatic features of gonadal dysgenesis, and varying degrees of ovarian differentiation, or with a clinical picture indistinguishable from that of classic 45,X gonadal dysgenesis. Phenotype and gonadal differentiation apparently depend on the proportion of 45,X to 46,XX or 47,XXX cells in somatic and germ cells during differentiation. Similarly, the presence of a structurally abnormal X chromosome frequently modifies some features of the classic syndrome. When 45,X/46,XY mosaicism or a structurally abnormal Y chromosome is present, varying degrees of testicular differentiation may be found. The spectrum of clinical findings may extend from phenotypic male through ambiguous genitalia to phenotypic female, depending on the degree of fetal testicular insufficiency. In addition, beneficial effects of a normal XY cell line or presence of some part of a Y chromosome may lead to normal stature and a modification of the somatic defects associated with 45,X monosomy. (From Jones HW Jr, Grumbach MM. Developmental disorders [females]. In Cooke RE [ed]. Biologic Basis of Pediatric Practice. New York, McGraw-Hill, 1968, pp 10871093. Reproduced with permission of McGraw-Hill, Inc.)
Figure 22-44 The loss of germ cells during migration to or after seeding of the indifferent gonad in a 45,X individual gives rise to a gonadal streak, because germ cells are necessary for ovarian development of the indifferent gonad; evidence suggests that loss occurs after implantation of germ cells. In the presence of 45,X/46,XX mosaicism, gonadal differentiation may vary from that of an ovary to that of a gonadal streak. Similarly, in 45,X/46,XY mosaics, depending on the sex chromosome constitution of the germ cells and gonadal blastema, gonadal differentiation may vary from that of a testis to that of a gonadal streak. In 47,XXY individuals, germ cells become implanted in the primitive testis, but a marked loss of spermatogonia seems to occur in the perinatal period and infancy. (From Jones HW Jr, Grumbach MM. Developmental disorders [females]. In Cooke RE [ed]. Biologic Basis of Pediatric Practice. New York, McGraw-Hill, 1968, pp 1087-1093. Reproduced with permission of McGraw-Hill, Inc.)
**Figure 22-45** Structural anomalies of the X chromosome. The normal X at the left is G banded. A dark band on the short arm and two major dark bands on the long arm are visible. The first Xq and the ring X chromosome (Xr) are not banded. They show late replication with tritiated thymidine. Note symmetry of the arms of the second Xq. Even with G banding, it is difficult to distinguish this chromosome from a possible short arm isochromosome. The long arm isochromosome (Xqi) appears to be dicentric. The two chromosomes to the far right are isodicentric X chromosomes. Both have two C bands but only one functional centromere. There is a mirror-like band pattern on both sides of a point between the two C bands. The first isodicentric X chromosome presumably represents a break in the long arm of X at q22 with fusion of chromatids and duplication of the entire chromatid. The second isodicentric X chromosome appears to represent a terminal break in the short arm so that reduplication of the chromatids has produced what appears to be almost two X chromosomes.
Figure 22-46 Variable gonadal function and phenotypic stigmata in three patients with a deletion of the short arm of the X chromosome (Xp) of different degrees.  

A 13-year-old phenotypic female of short stature (-3.5 SD) with low-set ears, high-arched palate, low hairline, broad chest with wide-spaced areolae, cubitus valgus, puffy hands and feet, and short fourth metacarpals. There was no evidence of secondary sexual characteristics. The plasma FSH level was elevated at 26 µg/L (LER-869); the plasma estradiol level was less than 22 pmol/L (6 pg/mL). The buccal smear contained a normal proportion of X chromatin bodies in interphase nuclei, which were conspicuously small. Karyotype analysis and autoradiography revealed a 46,XXp karyotype. The abnormal X chromosome appeared to lack the entire short arm.

B, A phenotypic female, aged 17 years, 4 months, with the stigmata of the syndrome of gonadal dysgenesis. Her height was 151 cm (-3 SD), and she had multiple nevi, cubitus valgus, and a short fourth metacarpal on the right hand. At age 13 the patient noted spontaneous onset of breast development, which did not progress. Plasma gonadotropin levels were elevated: LH 7.3 µg/L (LER-960) and FSH 53 µg/L (LER-869). The concentration of plasma estradiol was 70 pmol/L (19 pg/mL). On buccal smear, the cells had a normal proportion of X chromatin bodies, which appeared small. Karyotype analysis and autoradiography indicated an Xp chromosome that had been deleted close to the centromere, but a small segment of the short arm was visible distal to the centromere.

C, A 20-year-old phenotypic female with a chief complaint of dysfunctional uterine bleeding. She had short stature, slight puffiness of hands and feet, and short fourth metacarpals. Female secondary sexual characteristics appeared at age 11, and menarche at age 13 was followed by regular menses, which later became irregular. The buccal smear contained nuclei with a normal proportion of small sex chromatin bodies. Bilateral ovaries were identified grossly and histologically during an appendectomy. Karyotype was 46,XXp. The extent of deletion of the short arms of the abnormal X chromosome in this patient is less than that seen in patients in A and B. A segment of the short arm is readily discernible above the centromere. It appears that, in these three patients with Xp karyotypes, somatic and gonadal manifestations of the syndrome of gonadal dysgenesis correlated with the magnitude of deletion of the short arm of the X chromosome.
Figure 22-47  A. A 22-year-old tall female with a chief complaint of primary amenorrhea had a deletion of the long arm of one X chromosome, Xq. At age 12 she developed sparse pubic hair. Breast development did not occur, and she remained sexually infantile. Height was 178 cm (+2.6 SD), and weight was 70 kg (+1.2 SD). No somatic stigmata of the syndrome of gonadal dysgenesis were noted. Plasma gonadotropin levels were elevated: LH was 5.6 µg/L (LER-960), and follicle-stimulating hormone level was 36.5 µg/L (LER-869). The buccal smear showed a normal proportion of X chromatin bodies that were slightly small.  B. A Giemsa-stained Xq, which exhibited (C) the late-labeling pattern characteristic of an X chromosome.
Figure 22-48 Three patients with 45,X/46,XY sex chromosome mosaicism who illustrate the highly variable phenotype in this variant of the syndrome of gonadal dysgenesis. (Numbers of the patients refer to designation in Table 22-13.) A, Patient 1, a phenotypic female, was age 15 years, 4 months. She had short stature (-3.1 SD), an increased number of pigmented nevi, puffiness over the dorsa of fingers, and broad and short hands, and she was sexually infantile (breast development seen in photograph followed estrogen therapy) except for sparse pubic and axillary hair. The urinary gonadotropins were markedly elevated. B, Patient 3, aged 3 years, 1 month, had ambiguous external genitalia, perineal hypospadias, and undescended gonads. He was of average height and had a broad chest and a duplication of the left kidney. C, Patient 9, aged 8 years, 1 month, was a phenotypic male with a penile urethra and unilateral undescended gonad, average height, cubitus valgus, short fourth metacarpals, and puffiness of dorsa of fingers. By age 15, male secondary sexual characteristics were well advanced and a left scrotal testis, which was normal in histologic appearance, measured 4.0 × 2.4 cm.
Figure 22-49 The external genitalia of a normally differentiated male with 45,X/46,XY mosaicism. Karyotype analyses revealed 16% and 68% mosaicism for a 45,X cell line in blood and skin, respectively. Gonadotropin levels, both basal and LHRH stimulated, and plasma testosterone levels were normal. Fertility was documented in vitro and by the conception of a normal male fetus.
Figure 22-50 45,X/46,XY mosaicism with a feminizing gonadoblastoma. A, A 20-year-old female with many stigmata of the syndrome of gonadal dysgenesis, including short stature, multiple nevi, cubitus valgus, and hyperconvex, small nails. The buccal smear was X chromatin negative; on fluorescence microscopy, 30% of interphase nuclei had a single Y body. Karyotype was 45,X/46,XY. The patient had spontaneous development of pubic and axillary hair at age 12. At age 18, breast development was noted. Her height was 139 cm (-5.1 SD) and weight 39 kg (-2.5 SD). Bone age was 17 years; an intravenous pyelogram was normal. The concentration of plasma gonadotropins at 20 years of age was elevated; plasma luteinizing hormone was 6 µg/L (LER-960) and follicle-stimulating hormone was 50 µg/L (LER-869). The concentration of plasma estradiol was 95 pmol/L (26 pg/mL), and that of estrone was 117 pmol/L (32 pg/mL); the plasma testosterone level was less than 0.7 nmol/L (0.2 ng/mL). On exploratory laparotomy, normal-appearing fallopian tubes and a uterus were found. The right gonad was a typical "streak," with whorls of fibrous connective tissue. B, The left gonad was replaced by a 1.3 × 1 × 1 cm tumor mass, which, on histologic section, revealed well-defined nests and islands of Sertoli-Leydig-like cells and germ cells, as well as calcification consistent with diagnosis of gonadoblastoma. C, Higher magnification illustrates aggregates of germ cells and small epithelial cells resembling immature Sertoli cells, as well as cells indistinguishable from Leydig cells. After gonadectomy the concentration of plasma estradiol was prepubertal (<18 pmol/L [<5 pg/mL]).
Figure 22-51 Phenotypic male and female with syndrome of webbed neck, ptosis, congenital heart disease, short stature, and hypogonadism (pseudo-Turner's syndrome, Noonan's syndrome). A, A boy, aged 9 years, 7 months, exhibited characteristic abnormalities: triangular facies, prominent brow, hypertelorism, ptosis, antimongoloid slant of palpebral fissures, broad apex nasi, low-set ears, webbed neck, pectus excavatum, pulmonic stenosis and atrial septal defect, short stature (-3.5 SD), bilateral undescended testes, and high-grade mental retardation. At age 18, he was 154.0 cm in height (height age: 12 years, 5 months); the boy had Leydig cell hypofunction. Biopsy of testes showed germinal aplasia. (From Grumbach MM, Barr ML. Cytologic tests of chromosomal sex in relation to sexual anomalies in man. Recent Prog Horm Res 1958; 14:255334.) B, An 8-year-old girl with similar features. Height was 106.2 cm (height age: 4 years, 4 months). Pulmonic stenosis was present, and the karyotype was 46,XX.
Figure 22-52  A 17-year-old true hermaphrodite with bilateral scrotal ovotestes and a 46,XX sex chromosome constitution in cultures of peripheral blood and skin, perineal hypospadias (partially repaired in photograph), moderate bilateral gynecomastia and pubic hair (recently shaved in picture), sparse axillary hair, a high-pitched voice, and absent facial hair. Height was 168 cm. Urinary 17-ketosteroid level was 1.3 mg/day; urinary gonadotropin levels were elevated. A male type of urethra, bilateral scrotal fallopian tubes and ovotestes, and rudimentary bicornuate uterus and vagina attached to the posterior urethra were seen at operation. The photomicrographs show histopathology of the ovarian and testicular portion of one ovotestis.  B, Immature seminiferous tubules lined with Sertoli cells and spermatogonia and Leydig cells.  C, Ova and follicles.  (From Grumbach MM, Barr ML. Cytologic tests of chromosomal sex in relation to sexual anomalies in man. Recent Prog Horm Res 1958; 14:255-334.)
Figure 22-53 Female pseudohermaphroditism induced by prenatal exposure to androgens. Exposure after 12th fetal week leads only to clitoral hypertrophy (diagram on left). Exposure at progressively earlier stages of differentiation (depicted from left to right in drawings) leads to retention of the urogenital sinus and labioscrotal fusion. If exposure occurs sufficiently early, the labia fuse to form a penile urethra. (From Grumbach MM, Ducharme JR. The effects of androgens on fetal sexual development: androgen-induced female pseudohermaphroditism. Fertil Steril 1966; 11:157-180. Reproduced with permission of the publisher. © The American Fertility Society.)
Figure 22-54 Diagram of the steroid biosynthetic pathways and the biosynthetic defects that result in congenital adrenal hyperplasia. The defect in patients with "lipoid adrenal hyperplasia" is not (except for one reported case) in the CYP11A1 (cholesterol side-chain cleavage) enzyme but in STAR, the steroidogenic acute regulatory protein. This protein is involved in the transport of cholesterol from the outer mitochondrial membrane to the inner membrane where the CYP11A1 enzyme is located. CYP11B1 (11-hydroxylase) catalyzes 11-hydroxylation of deoxycorticosterone and 11-deoxycorticisol primarily. CYP17 (17-hydroxylase/17,20-lyase) catalyzes both 17-hydroxylation and splitting of the 17,20 bond, but for the latter it has preferential ♂-17,20-lyase activity (see text). CYP19 (aromatase) catalyzes the conversion of androstenedione to estrone and testosterone to estradiol. CYP11B2 (aldosterone synthetase) catalyzes the conversion of corticosterone to aldosterone. 3-HSD I and 3-HSD II, 3-hydroxysteroid dehydrogenase/4,5-isomerase types I and II; CYP21 (P450c21), 21-hydroxylase; 17-HSD 3, 17-hydroxysteroid dehydrogenase type 3. In the human, deletion of or a homozygous null mutation of CYP11A1 (P450scc) is probably lethal in utero but a heterogeneous mutation caused congenital lipoid adrenal hyperplasia (see text).
**Figure 22-55** Diagram of chromosome 6. Only the banding pattern of the short arm is shown. Numbers 11 to 25 delineate bands according to the Paris nomenclature. To right of the centromere, the sites of genes for the major histocompatibility complex (MHC), glyoxalase I (GLO), and phosphoglucomutase (PGM) are indicated on a recombinant unit scale. To left is a scheme of genes in the MHC complex. The gene for 21-hydroxylase is closely linked to HLA-B and resides between the HLA-B and HLA-D loci.
Figure 22-56 Pedigrees of two families with children with 21-hydroxylase deficiency. HLA haplotypes for HLA-A, HLA-B, and HLA-C are indicated for each individual. a, b indicate paternal haplotypes and c, d maternal haplotypes. Parents are heterozygotes for 21-hydroxylase deficiency. Patients with haplotype a, c are homozygous for 21-hydroxylase deficiency. Haplotype b, d indicates a child who has two normal genes for 21-hydroxylase activity. (Redrawn from Levine LS, Zachmann M, New MI, et al. Genetic mapping of the 21-hydroxylase deficiency gene within the HLA linkage group. Reprinted by permission of the New England Journal of Medicine 1978; 299:911915.)
Figure 22-58 Diagram of the CYP21B gene and the site of microconversions that cause 21-hydroxylase deficiency. The numbered black boxes are the exons. SW, salt wasting; SV, simple virilizing; NC, nondiagnostic form. AG indicates an adenine to guanine transition near the end of intron 2, which causes premature splicing, resulting in an aberrant 11-amino-acid string and a stop codon. 8nt indicates an eight-nucleotide deletion in exon 3. 306+1 indicates a thymidine insertion at codon 306, which causes a frameshift and stop. GGC is a guanine-to-cytosine transition at codon 484. As noted in the text, all of these deleterious mutations that inactivate or diminish 21-hydroxylase activity are normally present in the CYP21A pseudogene. The lengths of the arrows and the percentages in parentheses designate the magnitude of activity of the 21-hydroxylase enzyme in transfected cells.
Figure 22-59 Normal plasma 17-hydroxypregesterone levels in non-stressed infants from birth to 2 years of age. To convert 17-hydroxypregesterone values to nanomoles per liter, multiply by 0.03026. (From Jenner MR, Grumbach MM, Kaplan SL. Plasma 17-OH progesterone in maternal and umbilical cord plasma in children, and in congenital adrenal hyperplasia (CAH): application to neonatal diagnosis of CAH. Pediatr Res 1970; 4:380 [abstract].)
**Figure 22-60** A and B, An untreated girl with the nonsalt-losing form of congenital adrenal hyperplasia. Androgens caused disproportionate acceleration of bone maturation compared with stature. C, Virilized adult female with nonsalt-losing adrenal hyperplasia. The patient had a deep voice, shaved daily, and wore a toupee for baldness. After treatment with cortisone, her 17-ketosteroid levels fell to normal values, her breasts enlarged, she underwent a normal menarche, and hair regrew on her head. Note short stature and short extremities. D, Female pseudohermaphroditism caused by maternal ingestion of an oral progestational compound from the 8th to 12th week of pregnancy. Labioscrotal fusion is sufficient to obscure the vaginal orifice and create a urogenital sinus. Clitoris is enlarged. There is no progressive virilizing tendency. (C, from Wilkins L. The Diagnosis and Treatment of Endocrine Disorders in Childhood and Adolescence, 3rd ed. Springfield, IL, Charles C Thomas, 1965.)
Figure 22-61 Diagram of the CYP11B1 (P450c11) gene with representative mutations causing 11-hydroxylase deficiency. The exons are the numbered black boxes. Missense mutations causing amino acid substitutions in the enzyme are indicated by the three-letter abbreviation for the wild-type amino acid, followed by the amino acid number in the enzyme and then by the three-letter abbreviation for the substituted amino acid. X indicates a nonsense (stop) mutation; Pro32C indicates the deletion of a cytosine in the proline 32 codon, causing a frameshift; Asn394 + 2nt designates the addition of two nucleotides in the asparagine 394 codon, causing a frameshift. Mutations in CYP11B1 are distinct from those in CYP11B2 (P450as).
Figure 22-62 Diagram of the 3-hydroxysteroid dehydrogenase type 2 (HSD3B2) gene with selected mutations that result in 3-HSD deficiency. The numbered solid boxes indicate the exons. Missense mutations causing amino acid substitutions in the enzyme are indicated by the three-letter abbreviation for the wild-type amino acid, followed by the amino acid number in the enzyme and then the three-letter abbreviation for the substituted amino acid. X indicates a nonsense (stop) mutation. GA, nt6651 is a guanine-to-adenine transition at nucleotide 6651 in intron 3 that creates a new splice junction. 186/InsC/187 is a single-nucleotide cytosine insertion after codon 186 (proline) that changes the reading frame and causes an aberrant 16-amino-acid sequence before a stop codon. 273AA is a deletion of two adenines in codon 273 (AAA), which encodes lysine. This deletion causes a frameshift resulting in a premature termination at residue 279 and results in a truncated protein 94 amino acids shorter than the wild type. Mutations with less than 1% 3-HSD activity are indicated below the gene and cause salt loss. Missense and splicing mutations, indicated above the gene, result in 2% to 4.7% enzymatic activity and are associated with the nonsalt-losing phenotype.
Figure 22-63 Diagram of selected mutations in the CYP17 gene (17-hydroxylase/17,20-lyase deficiency). The exons are the numbered black boxes. Missense mutations causing amino acid substitutions in the enzyme are indicated by the three-letter abbreviation for the wild-type amino acid, followed by the amino acid number in the enzyme and the three-letter abbreviation for the substituted amino acid. X indicates a nonsense (stop) mutation. Phe53 or 54 is a deletion of phenylalanine at codon 53 or 54. 518/Ins469nt is a deletion of 518 nucleotides and an insertion of 469 nucleotides. His120 + 7nt is a seven-nucleotide duplication at codon 120 (histidine). +ILE112 is a duplication of isoleucine at codon 112. GC300,301 is a deletion of two nucleotides, guanine and cytosine, at codon 300 and 301. ILe480 + 4nt is a four-nucleotide duplication (cytosine-adenine-thymidine-cytosine) at codon 480. Asp487, Ser488, Phe489 indicates a deletion of aspartic acid (codon 487), serine (codon 488), and phenylalanine (codon 489). All of these mutations cause 17-hydroxylase deficiency. Missense mutations at codons 347 and 358 (indicated by the box) have been associated with "isolated" 17,20-lyase deficiency.
Figure 22-64 Model of the steroid-synthesizing cell (adrenal/gonadal) showing conversion of cholesterol to steroids.  

A, Cholesterol from low-density lipoprotein, from cholesterol esters stored in lipid droplets, and from endogenous synthesis in the endoplasmic reticulum is transported from the outer mitochondrial membrane to the inner membrane. This transport, which is a rate-limiting step in steroid synthesis, is facilitated by StAR (steroidogenic acute regulatory protein) as well as by other, StAR-independent mechanisms. In the mitochondria, steroid synthesis then ensues as a result of the conversion of cholesterol to 5α-pregnenolone by the enzyme CYP11A1 (P450scc).  

B, In patients with congenital lipoid adrenal hyperplasia, a mutation in the gene encoding StAR results in little or no activity of the mutant StAR, causing greatly diminished cholesterol transport into the mitochondria. Low levels of steroidogenesis via mechanisms independent of StAR can occur; however, increased ACTH (LH/FSH) secretion results in cholesterol accumulation in the cells as lipid droplets.  

C, Continued stimulation and resultant accumulation of cholesterol causes engorgement of these cells, with both mechanical and chemical perturbation of the cell function. Females with congenital lipoid adrenal hyperplasia feminize at puberty and menstruate but have progressive hypergonadotropic hypogonadism. It has been hypothesized by Bose and co-workers that this occurs because the follicular cells are relatively quiescent in utero and before puberty; hence, they are undamaged. At the beginning of each cycle, they are recruited, and a small amount of estradiol can be produced as a result of StAR-independent mechanisms. This can occur until the follicular cells are engorged and rendered nonfunctional. (From Bose HS, Sujiwara T, Strauss JF III, Miller WL. The pathophysiology and genetics of congenital lipoid adrenal hyperplasia. N Engl J Med 1996;335:1870-1878. Copyright 1996, Massachusetts Medical Society. All rights reserved.) (See text.)
**Figure 22-65** Diagram of the selected mutations identified in the StAR gene associated with congenital lipoid adrenal hyperplasia. Nucleotide (nt) and amino acid numbers are given according to the cDNA sequence. Missense mutations causing amino acid substitutions are indicated by the three-letter abbreviation for the wild-type amino acid, followed by the amino acid number in the protein and the three-letter abbreviation for the substituted amino acid. X indicates a nonsense (stop) mutation. 247/InsG/nt248 is an insertion of a guanine causing a frameshift between nucleotides 247 and 248. T nt261 is a deletion of a thymidine at nucleotide 261. 548/InsTT/nt549 is an insertion of two thymidines in exon 4 between nucleotides 548 and 549, causing a frameshift. T nt593 is a deletion of two thymidines at nucleotide 593, causing a frameshift. Cnt650 is a deletion of a cytosine at nucleotide 650, causing a frameshift. TA @ 11 is a thymidine-to-adenine transition minus 11 nucleotides (5') from the intron 4/exon 5 junction. 947/InsA/nt948 is the insertion of an adenine between nucleotides 947 and 948, which results in a frameshift. (Data from Bose H, Sugawara T, Stauss JF, Miller WL. The pathophysiology and genetics of congenital lipoid adrenal hyperplasia. N Engl J Med 1996; 335:1870-1878.)
Figure 22-66 The biosynthetic defects in converting C19-steroids (androgens, androgen precursors) to C18-steroids (estrogens) in the CYP19 (P450arom)-deficient fetal placental unit. 3-HSD, 3-hydroxysteroid dehydrogenase/4,5-isomerase; 17-HSD, 17-hydroxysteroid dehydrogenase; DHEA, dehydroepiandrosterone; DHEAS, DHEA sulfate; DHT, dihydrotestosterone; T, testosterone; A, androstenedione; E1, estrone; E2, estradiol; E3, estriol. (Modified and redrawn from Conte FA, Grumbach MM, Ito Y, et al. A syndrome of female pseudohermaphroditism, hypergonadotropic hypogonadism, and multicystic ovaries associated with missense mutations in the gene encoding aromatase (P450arom). J Clin Endocrinol Metab 1994; 78:1287-1292. © 1994, The Endocrine Society.)
Figure 22-67 Diagram of the CYP19 (P450arom) gene and selected mutations causing aromatase deficiency. The numbered black boxes represent translated exons. The septum in the open box in exon II represents the 3' acceptor splice junction for the untranslated exons. The multiple alternate promoters and the untranslated exons (open boxes) are indicated. Missense mutations causing amino acid substitutions in the enzyme are indicated by the three-letter abbreviation for the wild-type amino acid, followed by the amino acid number in the enzyme and the three-letter abbreviation for the substituted amino acid. X indicates a nonsense (stop) mutation.

GTAT 3nt X is a guanine-to-adenine transition at the splice junction between exon 3 and intron 3, resulting in a stop codon (X) three nucleotides downstream (3').

GTGC + 29 aa is a thymidine-to-cytosine transition at the splice junction between exon 6 and intron 6, giving rise to a 29-amino-acid insert in the protein. CPro408X, a deletion of a cytosine occurring in codon 408 (proline), results in a frameshift and a stop codon 3 nucleotides downstream (3N).

**Figure 22-68** hCG/LH resistance. Diagram of male sex determination and differentiation showing a defect in the hCG/LH receptor that causes Leydig cell unresponsiveness to hCG (LH) and results in male pseudohermaphroditism. Solid bar delineates defect, and stippled area designates general site of defect. Dashed lines indicate that subsequent processes may be completely or partially affected.
Figure 22-69 Diagram of the LH/hCG receptor with its seven transmembrane alpha helices and selected mutations that cause male pseudohermaphroditism. The open circles represent the amino acid residues on the LH/hCG receptor protein. The solid circles indicate the amino acid substitutions in patients with male pseudohermaphroditism. The mutations are indicated by the three-letter abbreviation of the wild-type amino acid, followed by the position number of the amino acid in the protein and the three-letter abbreviation for the substituted amino acid or the letter X, which indicates a nonsense (stop) mutation leading to an inactive, truncated receptor.
Figure 22-70 Diagram of male sex determination and differentiation showing the consequences of an enzymatic block in biosynthesis of testosterone that results in male pseudohermaphroditism. Solid bar indicates defect (see legend for Figure 22-63).
Figure 22-71 Defects in the biosynthetic pathway for testosterone. Five defects cause male pseudohermaphroditism in affected males (1 to 5). Even though all blocks affect both gonadal and adrenocortical steroidogenesis, only those at steps 1, 2, and 3 are associated with major abnormalities in biosynthesis of glucocorticoids and mineralocorticoids. StAR mutations impair transport of cholesterol to the inner mitochondrial membrane, the site of CYP11A1. In the human, null mutations in CYP11A1 are probably lethal in utero (see text). CYP17 (17-hydroxylase/17,20-lyase) primarily catalyzes the scission of 17-hydroxypregnenolone to dehydroepiandrosterone. Only minimal conversion of 17-hydroxyprogesterone to androstenedione occurs normally. Therefore, synthesis of gonadal steroids is mainly through the 5α-pathway.
Figure 22-72  A and B, Genitalia of male infant with congenital adrenal hyperplasia resulting from 3-HSD 2 deficiency. The boy was admitted at 9 days of age in a salt-losing crisis and died at 3 months of unexplained muscular paralysis. Paresis, resembling that of Werdnig-Hoffmann syndrome, became progressively more severe even though adrenal replacement therapy was adequate and blood electrolytes were normal. Biochemical findings revealed a severe block in the conversion of 5-3-hydroxysteroids to 4-3-ketosteroids.
Figure 22-73 The topology of 5-sterol reductase. Dark circles designate amino acid substitutions due to missense mutations in \textit{DHCR7}. The two arrows indicate site of stop codons. (From Witsch-Baumgartner M, Loffler J, Utermann G. Mutations in the human \textit{DHCR7} gene. Hum Mutat 2001; 17:172182.)
Figure 22-74 Diagram of the gene encoding 17-hydroxysteroid dehydrogenase type 3 with selected mutations reported to cause 17-HSD deficiency. The exons are the numbered black boxes. Missense mutations causing amino acid substitutions in the enzyme are indicated by the three-letter abbreviation for the wild-type amino acid, followed by the amino acid number in the enzyme and the three-letter abbreviation for the substituted amino acid. nt325 + 4 is a splice junction mutation, a transition of adenine (A) to thymidine (T), located four nucleotides downstream (3') of the boundary between exon 3 and intron 3; nucleotide 325 is the closest base pair in the exon to the mutation. nt326 is a splice junction mutation, a transition of guanine (G) to cytosine (C), located one nucleotide upstream (5') of the boundary between intron 3 and exon 4. nt665 is a splice junction mutation, a transition of guanine (G) to adenine (A), located one nucleotide upstream (5') of the splice junction between intron 8 and exon 9. nt777-783 indicates a deletion of nucleotides 777 to 783 in the gene. (Redrawn from Andersson S, Geissler WM, Wu L, et al. Molecular genetics and pathophysiology of 17-hydroxysteroid dehydrogenase 3 deficiency. J Clin Endocrinol Metab 1996; 81:130136. © 1996, The Endocrine Society.)
Figure 22-75 The syndrome of complete androgen resistance and its "variant" form.  

A, A 17-year-old patient with the complete syndrome. This phenotypic female was chromatin negative, had a 46,XY karyotype, and had total absence of sexual hair with female secondary sexual characteristics. A small vagina ended blindly.  

B, The testes exhibited Leydig cell hyperplasia and seminiferous tubules that lacked germinal elements. C, At laparotomy, abdominal testes, rudimentary wolffian structures, and no müllerian structures were found.  

D, The variant form of syndrome in a 25-year-old female. Sexual hair was present, although sparse.  

E, The testes exhibited Leydig cell hyperplasia.  

F, The clitoris was hypertrophied, but there was no labial fusion. A shallow vagina ended blindly. At laparotomy, hypoplastic wolffian structures and absent müllerian structures were noted.
Figure 22-76 Diagram of male pseudohermaphroditism caused by complete or partial androgen resistance illustrating defects in the androgen receptor that result in absent or reduced binding of DHT or impaired function of the ligand-bound receptor.
Figure 22-77. A, Diagram of the androgen receptor gene divided into its eight exons. Exon 1 codes for the NH₂-terminal domain and regulates transcription. Exons 2 and 3 code for two zinc fingers. Exons 4 through 8 code for the androgen-binding domain of the receptor. B, The organization of a steroid-responsive gene. Ligand binding activates the receptor, and it binds to the steroid response elements of the gene (as a dimer along with co-activators; not shown). Enhancers as well as a CAAT and a TATA box are present. Gene transcription begins 19 to 27 base pairs downstream of the TATA box.
Figure 22-78 Diagram of the androgen receptor gene (AR) with missense mutations that cause complete androgen resistance (CAR) and partial androgen resistance (PAR). Asterisks indicate mutations that have been found to cause both complete and partial androgen resistance. Each mutation is indicated by the three-letter abbreviation for the wild-type amino acid, followed by the position number of the amino acid in the protein and the three-letter abbreviation for the substituted amino acid. Exon 1 (open box) regulates transcription. Exons 2 and 3 (heavy diagonal lines) encode the DNA-binding region of the androgen receptor. Part of exon 4 (thin diagonal lines) encodes the "hinge region" of the androgen receptor, which contains a nuclear localization signal. The 3' end of exon 4 through exon 8 (stippled) encodes the ligand (androgen)-binding region. A mutational hot spot is located in exon 5 and can cause both CAR and PAR. Not shown are nonsense, frameshift, and splice junction mutations and deletions that can cause either CAR or PAR. (Redrawn from Quigley CA, De Bellis A, Marschke KB, et al. Androgen receptor defects: historical, clinical and molecular perspectives. Endocr Rev 1995; 16:271321. © 1995, The Endocrine Society.)
Figure 22-79 A patient with partial androgen resistance (Reifenstein's syndrome). Both the patient and his brother had hypospadias, poor masculinization, and marked gynecomastia. Both had a normal 46,XY karyotype, normal wolffian duct derivatives, and no müllerian structures. (Reproduced, with permission, from Bowen P, Lee CSN, Migeon CJ, et al. Hereditary male pseudohermaphroditism with hypogonadism, hypospadias, and gynecomastia [Reifenstein's syndrome]. Ann Intern Med 1965; 62:252270. Courtesy of Dr. E. C. Reifenstein, Jr.)
Figure 22-80 Diagram of male pseudohermaphroditism resulting from 5-reductase deficiency.
Figure 22-81  A. A prepubertal 46,XY child with 5-reductase-2 deficiency who was raised as a female.  B. A postpubertal male with 5-reductase-2 deficiency who has virilized and changed gender role behavior.  (From Peterson RE, Imperato-McGinley J, Gautier T, et al. Male pseudohermaphroditism due to 5-steroid deficiency. Am J Med 1977; 62:170191.)
Figure 22-82 Diagram of the gene encoding 5-reductase-2 with representative mutations causing 5-reductase-2 deficiency. The exons are the numbered black boxes. Missense mutations causing amino acid substitutions in the enzyme are indicated by the three-letter abbreviation for the wild-type amino acid, followed by the position number of the amino acid in the enzyme and the three-letter abbreviation for the substituted amino acid. X indicates a nonsense (stop) mutation. GT nt725 + 1 is a splice junction mutation of guanine (G) to thymidine (T) one nucleotide downstream of the boundary between exon 4 and intron 4. Nucleotide 725 is the closest nucleotide in the exon to the mutation. A cohort of patients from New Guinea has been described with complete deletion of the 5-reductase-2 gene. (Redrawn from Wilson JD, Griffin JE, Russell DW. Steroid 5-reductase deficiency. Endocr Rev 1993; 14:577-593. © 1993, The Endocrine Society.)
Figure 22-83 Diagram of the pathogenesis of dysgenetic male pseudohermaphroditism. This condition can result from a sex chromosome anomaly or from a mutant gene in the male sex determination or differentiation cascade. The degree of masculinization is dependent on the functional ability of the dysgenetic gonads to produce antimüllerian hormone and testosterone.
Figure 22-84 Diagram of the pathogenesis of the persistent müllerian duct syndrome.
Figure 22-85 Diagram of the mutations in the antimüllerian hormone gene that cause the persistent müllerian duct syndrome. Exons are the black numbered boxes. Mutations are indicated by the three-letter abbreviation for the wild-type amino acid, followed by the position number of the amino acid in the protein and the three-letter abbreviation for the substituted amino acid or X for a nonsense (stop) mutation. nt2526 is a deletion of nucleotides 25 and 26; nt353-356, nt1074-1087, and nt2277-2292 indicate similar deletions of the respective nucleotides. (Redrawn from Imbeaud S, Carré Eusebe D, Rey R, et al. Molecular genetics of the persistent müllerian duct syndrome: a study of 19 families. Hum Mol Genet 1994; 3:125131. By permission of Oxford University Press.)
Figure 22-86 Diagram of selected mutations in the gene for the AMH receptor type II. Black numbered boxes are exons. Exons 1 to 3 encode the extracellular domain of the receptor. Exon 4 (diagonal lines) encodes the transmembrane domain, and exons 5 through 11 encode the intracytoplasmic domain. Mutations are indicated by the three-letter abbreviation for the wild-type amino acid, followed by the position number of the amino acid in the receptor protein and the three-letter abbreviation for the substituted amino acid or X for a nonsense (stop) mutation. n8487 designates a deletion/insertion at nucleotides 8487. GA nt615 is a guanine-to-adenine transition at nucleotide 615, which is at the splice site between exon 2 and intron 2. 27nt (open box) is a 27-nucleotide deletion, the most common mutation causing the AMH-positive form of the persistent müllerian duct syndrome, a mutation present in 25% of the patients studied with this form. (Redrawn from Imbeaud S, Belville C, Messike-Zeitoun L, et al. A 27 base-pair deletion of the antimüllerian type II receptor gene is the most common cause of the persistent müllerian duct syndrome. Hum Mol Genet 1996; 5:12691277. By permission of Oxford University Press.)
Figure 22-87 Steps in the diagnosis of intersexuality in infancy and childhood. Step 1 involves initial work-up and provisional diagnosis. Step 2 is used in selected cases.
Figure 22-88 Steps in the differential diagnosis of male pseudohermaphroditism.
**Figure 23-1** Rate of linear growth and weight gain in utero and during the first 40 weeks after birth. Length velocity is expressed in centimeters per week. The solid line depicts actual linear growth rate; the dashed line connecting the prenatal and postnatal length velocity lines depicts the theoretical curve for no uterine restriction late in gestation. The lighter dashed line depicts weight velocity. (Data from Tanner JM. Fetus into Man. Cambridge, Mass, Harvard University Press, 1978.)
Figure 23-2 Height velocity chart for boys constructed from longitudinal observations of British children. The 97th, 50th, and 3rd percentile curves define the general pattern of growth during puberty. Shaded areas define velocities of children who have peak velocities at ages up to 2 standard deviations (SDs) before or after the average age depicted by the percentile lines. Arrows and diamonds mark the 97th, 50th, and 3rd percentiles of peak velocity when the peak occurs at these early or late limits. (Modified from charts prepared by J. M. Tanner and R. H. Whitehouse from data published in Tanner JM, et al. Arch Dis Child 1966; 41:613635; Iranmanesh A, et al. J Clin Endocrinol Metab 1991; 73:10811088; and Tanner JM, et al. Arch Dis Child 1976; 51:170179; reproduced with permission of J. M. Tanner and Castlemead Publications, Ward's Publishing Services, Herts, UK.)
Figure 23-3 Height velocity chart for girls (see legend for Figure 23-2). (Modified and reproduced with permission of J. M. Tanner and Castlemead Publications, Ward’s Publishing Services, Herts, UK.)
Figure 23-4 Technique for measuring recumbent length. A device suitable for measurement of length of infants is available from Raven Equipment Ltd., UK. (Courtesy of Noel Cameron.)
Figure 23-5 Technique for measuring erect height using the Harpenden stadiometer with direct digital display of height. Devices of this type are available from Seritex, Inc., Carlstadt, N.J., and from Holtain Ltd., Wales, UK.
Figure 23-6 Length-for-age and weight-for-age percentiles for boys (birth to 36 months). Developed by the National Center for Health Statistics in collaboration with the National Center for Chronic Disease Prevention and Health Promotion (2000). [http://www.cdc.gov/growthcharts](http://www.cdc.gov/growthcharts)
Figure 23-7  Head circumference-for-age and weight-for-length percentiles for boys (birth to 36 months). Developed by the National Center for Health Statistics in collaboration with the National Center for Chronic Disease Prevention and Health Promotion (2000).  http://www.cdc.gov/growthcharts
Figure 23-8 Length-for-age and weight-for-age percentiles for girls (birth to 36 months). Developed by the National Center for Health Statistics in collaboration with the National Center for Chronic Disease Prevention and Health Promotion (2000). [http://www.cdc.gov/growthcharts](http://www.cdc.gov/growthcharts)
Figure 23-9 Head circumference for age and weight-for-length percentiles for girls (birth to 36 months). Developed by the National Center for Health Statistics in collaboration with the National Center for Chronic Disease Prevention and Health Promotion (2000). http://www.cdc.gov/growthcharts
Figure 23-10 Stature-for-age and weight-for-age percentiles for boys (2 to 20 years). Developed by the National Center for Health Statistics in collaboration with the National Center for Chronic Disease Prevention and Health Promotion (2000). http://www.cdc.gov/growthcharts
Figure 23-11 Stature-for-age and weight-for-age percentiles for girls (2 to 20 years). Developed by the National Center for Health Statistics in collaboration with the National Center for Chronic Disease Prevention and Health Promotion (2000). http://www.cdc.gov/growthcharts
Figure 23-12 Body mass index-for-age percentiles for boys (2 to 20 years). Developed by the National Center for Health Statistics in collaboration with the National Center for Chronic Disease Prevention and Health Promotion (2000). [http://www.cdc.gov/growthcharts](http://www.cdc.gov/growthcharts)
Figure 23-13 Body mass index-for-age percentiles for girls (2 to 20 years). Developed by the National Center for Health Statistics in collaboration with the National Center for Chronic Disease Prevention and Health Promotion (2000). http://www.cdc.gov/growthcharts
**Figure 23-14** Development of pituitary cell lineages.  

A, Schematic representation of pituitary cell precursors showing the expression of prevalent transcription factors at each stage of development. Terminally differentiated cells are shown as larger and shaded circles together with the hormones produced (lineage-specific transcription factors are highlighted in bold in these cells). The interaction with transcription factors and signaling molecules in the hypothalamus is also noted. Transcription factors are represented in lower case (except for SF1 and GATA2), whereas signaling molecules appear in upper case.  

B, Schema showing the timing of appearance and disappearance of pituitary transcription factors during mouse embryogenesis. BMP4, bone morphogenic protein 4; e, embryonic day; ER, estrogen receptor; FGF8, fibroblast growth factor 8; FSH, follicle-stimulating hormone; GHRH, growth hormone-releasing hormone; GnRH, gonadotropin-releasing hormone; GSU, -glycoprotein subunit; LH, luteinizing hormone; POMC, pro-opiomelanocortin; PRL, prolactin; SF1, steroidogenic factor 1; TRH, thyrotropin-releasing hormone; TSH, thyrotropin (thyroid-stimulating hormone). (From Lopez-Bermejo A, Buckway CK, Rosenfeld RG. Genetic defects of the growth hormone-insulin-like growth factor axis. Trends Endocrinol 2000; 11:43.)
Figure 23-15 Main components of the hypothalamic-pituitary portal system. (From Guyton AC, Hall JE. Human Physiology and Mechanisms of Disease, 6th ed. Philadelphia, WB Saunders, 1997, p 600, with permission.)
**Figure 23-16** Covalent structure of human growth hormone. (From Chawla RK, Parks JS, Rudman D. Structural variants of human growth hormone: biochemical, genetic, and clinical aspects. *Annu Rev Med* 1983; 34:519547.)
Figure 23-17 Relation between 24-hour (24h) mean growth hormone (GH) levels and age in boys and men. Bars represent values for the 24-hour mean (± SE) levels of GH (left axis) from 60 24-hour GH profiles of healthy boys and men subdivided according to chronologic age. An idealized growth velocity curve reproduced from the 50th percentile values for whole-year height velocity of North American boys (9) is superimposed. (From Martha PM Jr, Rogol AD, Veldhuis JD, et al. Alterations in the pulsatile properties of circulating growth hormone concentrations during puberty in boys. J Clin Endocrinol Metab 1989; 69:563570.)
Figure 23-18. A, The mean (± SE) 24-hour (24h) levels of growth hormone (GH) for groups of normal boys at varied stages of pubertal maturation. B, The mean (± SE) area under the GH concentration versus time curve for individual GH pulses, as identified by the Cluster pulse detection algorithm. C, The number of GH pulses (± SE), as detected by the Cluster algorithm, in the 24-hour GH concentration profiles for boys in each of the pubertal study groups. Note: The mean 24-hour GH concentration changes are largely mediated by changes in the amount of GH secreted per pulse rather than the frequency of pulses. In each panel, bars bearing the same letter are statistically indistinguishable. (From Martha PM Jr, Rogol AD, Veldhuis JD, et al. Alterations in the pulsatile properties of circulating growth hormone concentrations during puberty in boys. J Clin Endocrinol Metab 1989; 69:563570.)
Figure 23-19 Levels of growth hormone (GH) and growth hormone-binding protein (GHBP) measured in normal pubertal boys throughout adolescence. GHBP levels do not significantly change during puberty, but there is a significant increment of GH production and, therefore, of GH levels during this same time. These data suggest that there may be greater amounts of "free GH" during this period, leading to greater production of insulin-like growth factor I. (Data from Martha PM Jr, et al. J Clin Endocrinol Metab 1989; 69:563570; and Martha PM Jr, et al. J Clin Endocrinol Metab 1991; 73:175181.)
Figure 23-21 Structure of the IGF-I gene. A, The organization of the genes encoding human, rat and chicken insulin-like growth factor I (IGF-I) is depicted. Exons are represented by boxes (coding regions are in black, noncoding regions are in white), polyadenylation sites by arrows, and promoter regions by a bracket and the letter P. The full extent of the second and last human exons and the second rat exon has not been determined, as indicated by the dotted lines. B, Structure and expression of the human IGF-I gene. The structure of the different human IGF-I messenger RNAs (mRNAs) is displayed below the map of the gene. Sites of pre-mRNA processing are indicated by the thin lines. Sites of differential polyadenylation are marked at the 3' end of the gene by vertical arrows and in the mRNAs by horizontal boxes of varying length. (A and B, From Rotwein P. Structure, evolution, expression, and regulation of insulin-like growth factors I and II. Growth Factors 1991; 5:318.)
**Figure 23-22** Structure of the *IGF-II* gene. A. The organization of human, rat, and mouse *IGF-II* genes is shown. Exons are represented by boxes (coding regions are in black, noncoding in white), polyadenylation sites by thick vertical arrows, and promoter regions by a bracket and the letter P. The multiple transcription initiation sites for mouse and rat exon 1 are marked by the thin vertical arrows. The locations of mouse pseudoexons 1 and 2 are indicated. B. Structure and expression of the human *IGF-II* gene. The structure of different human insulin-like growth factor II (IGF-II) messenger RNAs (mRNAs) is displayed below the map of the gene. The patterns of mRNA processing are indicated by the thin lines. Sites of differential polyadenylation are marked at the 3' end of the gene by vertical arrows and in the mRNAs by horizontal boxes of varying length. (A and B, From Rotwein P. Structure, evolution, expression, and regulation of insulin-like growth factors I and II. Growth Factors 1991; 5:318.)
Figure 23-23 Normal serum levels (micrograms per liter) of insulin-like growth factor I (IGF-I) for males (A) and females (B). Lines represent the mean ± 3 SD. (Data courtesy of Diagnostic Systems Laboratories, Inc., Webster, Texas.)
Figure 23-24 Serum insulin-like growth factor I (IGF-I) levels in normal subjects (a and d), normal short stature subjects (b and e), and subjects with growth disorders (c and f). (From Rosenfeld RG, Wilson DM, Lee PDK, Hintz RL. Insulin-like growth factors I and II in the evaluation of growth retardation. J Pediatr 1986; 109:428-433.)
Figure 23-25 Serum insulin-like growth factor II (IGF-II) levels in normal subjects (A), normal short stature subjects (B), and subjects with growth disorders (C). (From Rosenfeld RG, Wilson DM, Lee PDK, Hintz RL. Insulin-like growth factors I and II in the evaluation of growth retardation. J Pediatr 1986; 109:428-433.)
Figure 23-26 Structure of the insulin-like growth factor (IGF) receptors. The insulin and IGF-I receptors are both heterotetrameric complexes composed of extracellular subunits that bind the ligands and subunits that anchor the receptor in the membrane and that contain tyrosine kinase activity in their cytoplasmic domains. The tyrosine kinase domain of the insulin receptor-related receptor (IRR) is homologous to the tyrosine kinase domains of the insulin and IGF-I receptors. The C-terminal domain is deleted in the IRR. Hybrids consist of a hemireceptor from both insulin and IGF-I receptors. The IGF-II/mannose-6-phosphate (M6P) receptor is not structurally related to the IGF-I and insulin receptors or the IRR, having a short cytoplasmic tail and no tyrosine kinase activity. (From LeRoith D, Werner H, Geitner-Johnson D, Roberts CT Jr. Molecular and cellular aspects of the insulin-like growth factor I receptor. Endocr Rev 1995; 16:143-163. © The Endocrine Society.)
Figure 23-27 Structure of the human insulin-like growth factor I (IGF-I) receptor precursor. Molecular cloning of human IGF-I receptor complementary DNAs (cDNAs) isolated from a placental library revealed the presence of an open reading frame of 4101 nucleotides. The 1367-amino acid polypeptide contains, at its N-terminus, a 30-amino acid hydrophobic signal peptide, which is responsible for the transfer of the nascent protein chain into the endoplasmic reticulum. After digestion by endopeptidases at a proteolytic cleavage site (Arg-Lys-Arg-Arg) located at residues 707 to 710, and subunits are released and linked by disulfide bonds to give the configuration of the mature heterotetrameric receptor. Shown in this diagrammatic representation are, in addition, the cysteine-rich domain of the subunit and the transmembrane and tyrosine kinase domains of the subunit. (From LeRoith D, Werner H, Geitner-Johnson D, Roberts CT Jr. Molecular and cellular aspects of the insulin-like growth factor I receptor. Endocr Rev 1995: 16:143163. © The Endocrine Society.)
Figure 23-30 Amino acid sequences of human insulin-like growth factor binding proteins (IGFBPs) 1 to 6, deduced from nucleotide sequences. Sequences in the amino-terminal and carboxyl-terminal residues are aligned to show maximal homologies. Dashes indicate gaps. Residues that are identical in five or six of the six IGFBPs are shaded. (From Rechler MM. Insulin-like growth factor binding proteins. Vitam Horm 1993; 47:114.)
Figure 23-31 Schematic representation of the effect of insulin-like growth factor binding protein (IGFBP) proteases on IGF action. In this model, proteolysis of IGFBPs results in a reduction in their affinity for IGF ligands, resulting in enhanced binding of IGF peptides by IGF receptor. (From Cohen P, Rosenfeld RG. The IGF axis. In Rosenbloom AL [ed]. Human Growth Hormone: Basic and Scientific Aspects. Boca Raton, Fla, CRC Press, 1991, pp 4358.)
Figure 23-32 The effect of insulin-like growth factor binding protein 3 (IGFBP-3) proteolysis by prostate-specific antigen (PSA) on IGFBP-3 affinity for IGF-I (A) and IGF-II (B). (From Cohen P, Peehl DM, Graves HC, Rosenfeld RG. Biological effects of prostate-specific antigen as an insulin-like growth factor binding protein-3 protease. J Endocrinol 1994; 142:407-415.)
Figure 23-33 The effect of insulin-like growth factorbinding protein 3 (IGFBP-3) proteolysis by prostate-specific antigen (PSA) on the ability of IGFBP-3 to inhibit IGF-I (A) and IGF-II (B) action. (From Cohen P, Peehl DM, Graves HC, Rosenfeld RG. Biological effects of prostate-specific antigen as an insulin-like growth factorbinding protein-3 protease. J Endocrinol 1994; 142:407415.)
Figure 23-34  Affinity cross-linking of $[^{125}I]$IGF-I (A) and $[^{125}I]$IGF-II (B) to membranes from Hs578T breast cancer cells. In the absence of unlabeled insulin-like growth factor (IGF) peptide (lane 1), IGF was predominantly bound to 40- to 45-kd IGFBP-3; no type I or type II IGF receptors were observed. Iodinated IGF was readily displaceable by unlabeled IGF-I or IGF-II (lanes 2 to 5) but not by unlabeled IGF-I-insulin hybrid molecule (lanes 6 and 7) or by an IGF analogue with decreased affinity for IGF-binding proteins (QAYL, lanes 9 and 10 in A). However, addition of [Leu27] IGF-II, which has decreased affinity for the type I IGF receptor (lanes 11 and 12 in A and lanes 7 and 8 in B), resulted in "unmasking" of the 130-kd subunit of the type I IGF receptor (A) and the 250-kd type II IGF receptor. (From Oh Y, Muller HL, Lamson G, Rosenfeld RG. Insulin-like growth factor (IGF)-independent action of IGF-binding protein 3 in hs578T human breast cancer cells: cell surface binding and growth inhibition. J Biol Chem 1993; 268:1496414971.)
Figure 23-35 Effect of transfection of Balb/c fibroblasts with a human insulin-like growth factor binding protein 3 (IGFBP-3) complementary DNA (cDNA) (Tx-BP-3) or with the control plasmid (Tx-P) on cell growth. Transfection with the IGFBP-3 cDNA resulted in a decreased cell proliferation (A) and increased cell doubling time (B). The latter effect could not be overcome with insulin, supporting the concept that the inhibitory effects of IGFBP-3 are IGF independent. (From Cohen P, Lamson G, Okajima T, Rosenfeld RG. Transfection of the human insulin-like growth factor binding protein 3 gene into Balb/c fibroblasts inhibits cellular growth. Mol Endocrinol 1993; 7:380386. © The Endocrine Society.)
Figure 23-36 Theoretical mechanisms of cellular insulin-like growth factor binding protein (IGFBP) actions.
Figure 23-37 Inhibition of Hs578T breast cancer cell growth by insulin-like growth factor binding protein 3 (IGFBP-3) is IGF independent. Recombinant IGFBP-3 from *Escherichia coli* results in decreased cell number and cannot be overcome by the addition of an IGF analogue with normal affinity for IGF receptors but decreased affinity for IGFBP-3 (QAYL-Leu-IGF-II). On the other hand, IGF-II, which itself does not stimulate cell proliferation in Hs578T cells, partially releases cells from the growth-inhibitory effects of IGFBP-3, presumably by causing dissociation of IGFBP-3 from the cell membrane. (From Oh Y, Muller HL, Lamson G, Rosenfeld RG. Insulin-like growth factor [IGF]-independent action of IGF-binding protein 3 in Hs578T human breast cancer cells. J Biol Chem 1993; 268:14964-14971.)
Figure 23-38 Transforming growth factor-2 (TGF-2) inhibits Hs578T cell growth by transcriptional regulation of insulin-like growth factor binding protein 3 (IGFBP-3). Reduction in IGFBP-3 messenger RNA (mRNA) and protein levels through the use of an IGFBP-3 antisense oligodeoxynucleotide resulted in significant reduction in the growth inhibitory actions of TGF-2. (From Oh Y, Muller HL, Ng L, Rosenfeld RG. TGF-2-induced cell growth inhibition in human breast cancer cells is mediated through IGFBP-3 action. J Biol Chem 1995; 270:13589-13592.)
Figure 23-39 Schematic diagram of insulin-like growth factor (IGF)-independent and IGF-dependent actions of IGF-binding protein 3 (IGFBP-3), the latter being mediated through a putative membrane-associated IGFBP-3 receptor.
Figure 23-40 Normal serum levels of insulin-like growth factor binding protein 3 (IGFBP-3) (micrograms per milliliter) by age for males (A) and females (B). The lines represent the mean ± 3 SD. (Data courtesy of Diagnostic Systems Laboratories, Inc., Webster, Texas.)
Figure 23-41 Curves of weight and height of a child with growth failure resulting from prolonged self-imposed caloric restriction because of a fear of becoming obese. The crossing of percentiles on the weight curve preceded that for the height curve; when the caloric intake was normalized (arrow), the gain in weight occurred before the improvement in linear growth. At the end of the prolonged period of caloric restriction, weight age (10.2 years) was less than height age (12 years). (From Pugliese MT, Lifshitz F, Grad G, et al. Fear of obesity: A cause of short stature and delayed puberty. N Engl J Med 1983; 309:513518.)
Figure 23-42 Catch-up growth in a girl with gluten-induced enteropathy (celiac disease). After 8 years of growth impairment, the patient was placed on a gluten-free diet and demonstrated substantial catch-up growth. Note the return to the previous growth percentiles. (Courtesy of J. M. Tanner.)
Figure 23-43 Growth curves of two boys with obesity. The boy depicted by the circles had cortisol excess related to Cushing's disease. An onset of rapid weight gain was associated with a decrease in linear growth velocity at 7 years of age. The diagnosis was established, and an adrenalectomy (arrow) was performed at the age of 9½ years, with an almost immediate increase in growth rate and striking catch-up. The boy whose growth is depicted by triangles had exogenous obesity. At the age of 9½ years, his weight was approximately the same as that of the patient with Cushing's disease, but his height was at the 97th percentile, reflecting the enhancement of linear growth in this individual with exogenous obesity.
Figure 23-44 The hypothalamic-pituitary-IGF axis: sites of established and hypothetical defects. The established defects are shown as Roman numerals in the gray-shaded circles or ovals, and the hypothetical defects are shown in the white circles or ovals. ALS, acid-labile subunit; GH, growth hormone; GHBP, GH-binding protein; GHRH, GH-releasing hormone; IGF, insulin-like growth factor; IGFBP, IGF-binding protein; JAK, Janus kinase; MAPK, mitogen-activated protein kinase; STAT, signal transducer and activator of transcription. (From Lopez-Bermejo A, Buckway CK, Rosenfeld RG: Genetic defects of the growth hormone-insulin-like growth factor axis. Trends Endocrinol 2000; 11:43.)
Figure 23-45 Decision tree for the investigation of genetic defects in patients with IGF deficiency. Hypothetical genetic defects are presented in parentheses. In those defects shown with a # sign, abnormalities in other organs and structures besides the hypothalamus-pituitary-IGF axis are expected to occur as a result of these genetic defects. ACTH, adrenocorticotropic hormone; CPHD, combined pituitary hormone deficiencies; FSH, follicle-stimulating hormone; GH, growth hormone; GHR, GH receptor; GHRHR, GH-releasing hormone receptor; IGF, insulin-like growth factor; IGFR, IGF-I receptor; LH, luteinizing hormone; PRL, prolactin; STAT5, signal transducer and activator of transcription 5; TSH, thyrotropin (thyroid-stimulating hormone). (From Lopez-Bermejo A, Buckway CK, Rosenfeld RG: Genetic defects of the growth hormone-insulin-like growth factor axis. Trends Endocrinol 2000; 11:43.)
**Figure 23-46** Magnetic resonance image of infundibular dysgenesis. A, T1-weighted sagittal and coronal images of the hypothalamic-pituitary area in a normal 8-year-old girl. The anterior pituitary (AP) and posterior pituitary (PP) lobes and the pituitary stalk (PS) are marked. B, T1-weighted sagittal and coronal images of the hypothalamic-pituitary area of a 17-year-old boy with isolated human growth hormone deficiency. The anterior pituitary (AP) lobe is hypoplastic, the posterior pituitary (PP) lobe is ectopic, and the pituitary stalk is absent. (From Root AW, Martinez CR. Magnetic resonance imaging in patients with hypopituitarism. Trends Endocrinol Metab 1992; 3:283-287.)
Figure 23-47 Height measurements for Ecuadorian children with insulin-like growth factor deficiency resulting from growth hormone insensitivity.  (From Rosenfeld RG, Rosenbloom AL, Guevara-Aguirre J. Growth hormone [GH] resistance due to primary GH receptor deficiency. Endocr Rev 1994; 15:369390. © The Endocrine Society.)
Figure 23-48 Facial appearance of Ecuadorian patients with insulin-like growth factor deficiency due to growth hormone insensitivity. (From Rosenfeld RG, Rosenbloom AL, Guevara-Aguirre J. Growth hormone [GH] resistance due to primary GH receptor deficiency. Endocr Rev 1994; 15:369-390. © The Endocrine Society; photography by A. L. Rosenbloom, M.D.)
Figure 23-49 The normal GH-IGF axis (A) and the GH-IGF axis showing four potential biochemical defects capable of causing GH insensitivity (B): (1) abnormalities of the GH receptor, binding protein, or both; (2) abnormal signal transduction, resulting from a defect in the intracellular domain of the GH receptor or postreceptor; (3) defect of IGF synthesis; and (4) defect of IGF secretion. GH, growth hormone; GH-BP, GH-binding protein; GHRH, GH-releasing hormone; IGF, insulin-like growth factor; IGFBP, IGF-binding protein; SMS, somatostatin. (From Rosenfeld RG, Rosenbloom AL, Guevara-Aguirre J. Growth hormone [GH] resistance due to primary GH receptor deficiency. Endocr Rev 1994; 15:369-390. © The Endocrine Society.)
Figure 23-50  Serum growth hormone (GH) and GH-binding protein levels in sera of patients with GH receptor deficiency from Ecuador. (From Rosenfeld RG, Rosenbloom AL, Guevara-Aguirre J. Growth hormone [GH] resistance due to primary GH receptor deficiency. Endocr Rev 1994; 15:369-390. © The Endocrine Society.)
**Figure 23-51** Serum levels of insulin-like growth factor I (IGF-I), IGF-II, IGF-binding protein 2 (IGFBP-2) and IGFBP-3 in patients with GH receptor deficiency from Ecuador. (From Rosenfeld RG, Rosenbloom AL, Guevara-Aguirre J. Growth hormone [GH] resistance due to GH receptor deficiency. Endocr Rev 1994; 15:369-390. © The Endocrine Society.)
Figure 23-52 Clinical and biochemical evaluation of growth failure: seeking the diagnosis of IGF deficiency syndrome. The goal of Step 1 is to define the patient to be assessed. Step 2 seeks a wide array of "diseases" associated with poor growth. Step 3 relates to the possibility of GH provocative testing. Step 4 defines assessment of the hypothalamic-pituitary morphology and the consideration of adrenocorticotrophic hormone deficiency. Nonetheless, the primary evaluation is for "IGF deficiency," with studies designed to delineate hypothalamic or pituitary abnormalities or GH insensitivity. CRH, corticotropin-releasing hormone; GH, growth hormone; GHBP, growth hormone-binding protein; GHD, growth hormone deficiency; GHIS, growth hormone insensitivity syndrome; GHR, growth hormone receptor; GHRHR, growth hormone-releasing hormone receptor; HPA, hypothalamic-pituitary-adrenal; IGF, insulin-like growth factor; IGFBP-3, IGF-binding protein 3; ITT, insulin tolerance test; SDS, standard deviation score; T₄, thyroxine; TSH, thyrotropin (thyroid-stimulating hormone).
Figure 23-53 Annual growth velocity (mean ± SD) for prepubertal patients with growth hormone (GH) deficiency prior to and during 4 years of GH treatment, contrasting results with daily (QD) and thrice-weekly (TIW) injections. The mean annual growth velocity in the QD group was significantly greater than that of the thrice-weekly group during each year, although significance diminished from year 1 to year 4. (From MacGillivray MH, Baptista J, Johanson A, and Genentech Study Group. Outcome of a four-year randomized study of daily versus three times weekly somatropin treatment in prepubertal naive growth hormone-deficient children. J Clin Endocrinol Metab 1996; 81:18061809; reproduced by permission of M. H. MacGillivray.)
Figure 23-54 Height standard deviation score (SDS) (mean ± SD) for prepubertal patients with growth hormone deficiency (GHD) before and during 4 years of growth hormone (GH) treatment, contrasting results with daily (QD) and thrice-weekly injections (TIW). The mean SDS in the QD group was significantly greater throughout the treatment period. Younger patients had the greatest increase in height SDS, and the effect of age was more marked in the QD group. (From MacGillivray MH, Baptista J, Johanson A, and Genentech Study Group. Outcome of a four-year randomized study of daily versus three times weekly somatropin treatment in prepubertal naive growth hormonedeficient children. J Clin Endocrinol Metab 1996; 81:18061809 (1328). Reproduced by permission of M. H. MacGillivray.)
**Figure 23-55** Algorithm for transition to adult treatment of growth hormone deficiency (GHD). MPHD, multiple pituitary hormone deficiencies.
Figure 23-57 Height standard deviation score (SDS) (mean ± SD) in 20 growth-retarded prepubertal patients with chronic renal insufficiency. Note that the basal height is outside the normal range (at -2.6 SD), enters the normal range within 1 year of treatment, and does not differ from the mean by the 5th year of growth hormone therapy. (From Fine RN, Kohaut E, Brown D, et al. Long-term treatment of growth-retarded children with chronic renal insufficiency with recombinant human growth hormone. Kidney Int 1996; 49:781785; reproduced by permission of R. N. Fine.)
Figure 23-58 Adult height of patients with Turner's syndrome who were treated with growth hormone (GH) (n = 17) or combination GH plus oxandrolone (n = 45), compared with historical Turner's controls (n = 25), relative to each subject's projected adult height (indicated by dotted zero line). The mean increments in adult height relative to the projected adult height are indicated. Diamond symbols in the combination group indicate two subjects with poor compliance who terminated treatment early. (From Rosenfeld RG, Attie KM, Frane J, et al. Growth hormone treatment of Turner syndrome: beneficial effect on adult height. J Pediatr 1998; 132:319-324.)
Figure 23-59 Growth during insulin-like growth factor I treatment of growth hormone insensitivity, relative to normal standards (Tanner) and median for untreated patients (Laron) in boys (A) and girls (B). Closed circles indicate growth and open circles indicate the onset of puberty. (From Ranke MB, Savage MO, Chatelain PG, et al. Long-term treatment of growth hormone insensitivity syndrome with IGF-I. Horm Res 1999; 51:128134.)
Figure 24-1  A summary and proposed scheme of the evolution of the human pattern of postnatal growth and development during the first 20 years of life. *A. afar*, *Australopithecus afarensis*, a "bipedal chimpanzee"; *A. africana*, *Australopithecus africanus*; *H. habilis*, *Homo habilis* (the toolmaker); *H. erect 1*, early *Homo erectus*; *H. erect 2*, late *Homo erectus*; *H. sapiens*, *Homo sapiens*. The early hominid *australopithecine* specimens from South Africa date to about 3.0 to 1.5 million years ago. *H. afarensis*, while a hominid (the family of all human species), retained many anatomic features of nonhominid species, for example, an adult brain size of about 400 mL compared with *H. habilis* (650 to 800 mL), early *H. erectus* (850 to 900 mL), late *H. erectus* (up to 1100 mL), and modern *H. sapiens* (about 1400 mL). Infancy is defined as the period when the mother's breast milk is the sole or most important source of nutrition and in preindustrialized societies ends at about 36 months. Childhood is the period after weaning, when the child is dependent on others for food and protection; this period ends when the growth of the brain in weight is almost complete, at about age 7 years. The juvenile stage is defined as prepubertal individuals who are no longer dependent on their parents for survival. The adolescent stage begins with the onset of puberty, when adult height is attained. The pattern in *A. afarensis* is no different from that in the chimpanzee (*Pan troglodytes*). Note the first appearance of the childhood stage, *H. habilis* (arising about 2 million years ago) and the first appearance of the adolescent stage in *H. erectus 2* (about 500,000 years ago); *H. sapiens* arose about 120,000 to 150,000 years ago. (Modified from Bogin B. Growth and development: recent evolutionary and biocultural research. In Boaz NT, Wolfe LD [eds]. Biological Anthropology: The State of the Science. Bend, Ore, International Institute for Human Evolutionary Research, 1995, pp 4970.)
Figure 24-2 Changes in age at menarche, 1840 to 1978, illustrating the advance in the age at menarche in Western Europe and the United States since 1840 and the slowing of this trend since about 1965. (Modified from Tanner M, Eveleth PB. Variability between populations in growth and development at puberty. In Berenberg SR [ed]. Puberty, Biologic and Psychosocial Components. Leiden, HE Stenfert Kroese, 1975, pp 256-273. Reprinted by permission of Kluwer Academic Publishers.)
Figure 24-3  Stages of breast development according to Marshall and Tanner and Reynolds and Wines. Stage 1: preadolescent; elevation of papilla only. Stage 2: breast bud stage; elevation of breast and papilla as a small mound, enlargement of areolar diameter. Stage 3: further enlargement of breast and areola with no separation of their contours. Stage 4: projection of areola and papilla to form a secondary mound above the level of the breast. Stage 5: mature stage; projection of papilla only, resulting from recession of the areola to the general contour of the breast. (Photographs from Van Wieringen JD, Wafelbakker F, Verbrugge HP, et al. Growth Diagrams 1965 Netherlands: Second National Survey on 024 Year Olds. Netherlands Institute for Preventative Medicine TNO. Groningen, Wolters-Noordhoff, 1971. © Wolters-Noordhoff, Groningen.)
Figure 24-4 Stages of female pubic hair development, according to Marshall and Tanner, Reynolds and Wines, and Dupertuis and colleagues. Stage 1: preadolescent; the vellus over the pubes is not further developed than that over the anterior abdominal wall; that is, there is no pubic hair. Stage 2: sparse growth of long, slightly pigmented, downy hair, straight or only slightly curled, appearing chiefly along the labia. This stage is difficult to see on photographs. Stage 3: hair is considerably darker, coarser, and curlier. The hair spreads sparsely over the junction of the pubic region. Stage 4: hair is now adult in type, but the area covered by it is still considerably smaller than in most adults. There is no spread to the medial surface of the thighs. Stage 5: hair is adult in quantity and type, distributed as an inverse triangle of the classical feminine pattern. The spread is to the medial surface of the thighs but not up the linea alba or elsewhere above the base of the inverse triangle. (Photographs from Van Wieringen JD, Wafelbakker F, Verbrugge HP, et al. Growth Diagrams 1965 Netherlands: Second National Survey on 624 Year Olds. Netherlands Institute for Preventive Medicine TNO. Groningen, Wolters-Noordhoff, 1971. © Wolters-Noordhoff, Groningen.)
Figure 24-5 Stages of male genital development and pubic hair development, according to Marshall and Tanner, Reynolds and Wines, and Dupertuis and colleagues. Genital development: Stage 1: preadolescent. Testes, scrotum, and penis are about the same size and proportion as in early childhood. Stage 2: the scrotum and testes have enlarged; the scrotal skin shows a change in texture and also some reddening. Stage 3: growth of the penis has occurred, at first mainly in length but with some increase in breadth; there is further growth of the testes and scrotum. Stage 4: the penis is further enlarged in length and breadth with development of the glans. The testes and scrotum are further enlarged. The scrotal skin has further darkened. Stage 5: genitalia are adult in size and shape. No further enlargement takes place after stage 5 is reached. Public hair development: Stage 1: preadolescent; the vellus over the pubic region is not further developed than that over the abdominal wall; that is, there is no pubic hair. Stage 2: sparse growth of long, slightly pigmented, downy hair, straight or slightly curled, appearing chiefly at the base of the penis. Stage 3: hair is considerably darker, coarser, and curlier and spreads sparsely over the junction of the pubes. Stage 4: hair is now adult in type, but the area it covers is still considerably smaller than in most adults. There is no spread to the medial surface of the thighs. Stage 5: hair is adult in quantity and type, distributed as an inverse triangle. The spread is to the medial surface of the thighs but not up the linea alba or elsewhere above the base of the inverse triangle. Most men will have further spread of the pubic hair. (Photographs from Van Wieringen JD, Wafelbakker F, Verbrugge HP, et al. Growth Diagrams 1965 Netherlands: Second National Survey on 024 Year Olds. Netherlands Institute for Preventative Medicine TNO. Groningen, Wolters-Noordhoff, 1971. © Wolters-Noordhoff, Groningen.)
Figure 24-6 Diagram of the sequence of events at puberty in males. An average is represented in relation to the scale of ages; the range of ages within which some of the changes occur is indicated by the figures below. (From Marshall WA, Tanner JM. Variations in the pattern of pubertal changes in boys. Arch Dis Child 1970; 45:1323.)
Figure 24-7 The sequence of events at puberty in females. The design of the figure is described in the legend of Figure 24-6. (From Marshall WA, Tanner JM. Variations in pattern of pubertal changes in girls. Arch Dis Child 1969; 44:291303.)
Figure 24-8 Schematic representation of the growth of ovarian follicles during infancy and childhood. Type 1 (primordial follicle) and type 2 (primary follicle) are composed of a small oocyte and a few to a ring of flat granulosa cells. In the diplotene (nestling) stage of prophase, primary follicles are the predominant form of oocyte and constitute the reservoir of cells from which follicular growth occurs. Types 3 to 5 (preantral follicles) are follicles that have entered the growth phase; the oocyte is enlarging and is surrounded by a zona pellucida, and granulosa cells increase in number and differentiate. The growth of the oocyte is complete by the end of the preantral stage, and the increased follicular size is due to follicular growth and fluid accumulation. Types 6 to 8 represent antral follicles (graafian follicles) and contain a fully grown oocyte, a large number of granulosa cells, a fluid-filled cavity, and a well-developed theca external to the basement membrane. Large preovulatory follicles are absent (10,000 to 15,000 µm). Follicular growth and atresia take place throughout childhood; all follicles that enter the growth phase become atretic, and this can occur at any stage in their development but mainly involves large antral follicles. (From Peters H, Byskov AG, Grinsted J. Follicular growth in fetal and prepubertal ovaries of humans and other primates. Clin Endocrinol Metab 1978; 7:469-485.)
Figure 24-9 High-resolution pelvic ultrasonography. Top left, Prepubertal uterus. Top right, Prepubertal ovary demonstrating four small follicular cysts (arrows). Bottom left, Pubertal postmenarchal uterus. Bottom right, Ovarian cyst in a girl with true precocious puberty.
Figure 24-10  **Left**, Diagram illustrating developmental stages of testicular germ cells based on electron microscopic findings in the rabbit. Note differences between prespermatogonium and spermatogonium.  **Right**, Diagram showing maturation of testicular cell types in the rabbit from prepubertal appearance at left to onset of spermatogenesis at right. Interstitial cells undergo changes in shape, size, and arrangement in the process of Leydig cell differentiation.  *(From Gondos B. Testicular development. In Johnson AD, Gomes WR [eds]. The Testis, vol 4. New York, Academic Press, 1977, pp 137.)*
Figure 24-11 The adolescent growth spurt in girls and boys (growth velocity curves). Note the later onset of the pubertal growth spurt in boys and the approximately 2-year difference in peak height velocity and the greater magnitude of peak height velocity compared with girls. The timing of the effects of estradiol is indicated. Progressive epiphyseal fusion terminates the growth spurt and leads to final or adult height. (From Grumbach MM. Estrogen, bone, growth, and sex: a sea change in conventional wisdom. J Pediatr Endocrinol Metab 2000; 13[suppl 6]:14391455.)
**Figure 24-12** The infancy, childhood, and puberty (ICP) model of Karlberg for mean attained height (left) and height velocity (right) for boys. The mean value for each component (infancy, childhood, puberty) and their sums (combined growth, right; combined velocity, left) are plotted. The growth curve for an individual represents the additive effect of the three biologic phases of the growth process (ICP). Karlberg has provided mathematical functions for each component of his model. **Infancy:** This component starts before birth and falls off by age 3 to 4 years. It can be described by the exponential function \( y = a + b[1 \exp(ct)] \). Average total gain in height for Swedish boys is 79.0 cm (44.0% of final height) and for girls is 76.8 cm (46.2%). **Childhood:** This phase begins at the end of the first year of life and continues to mature height. A second-degree polynomial function describes this component: \( y = a + bt + ct^2 \). Average total gain in height for boys is 85.2 cm (47.4%) and for girls is 78.4 cm (47.3%). **Puberty:** The model for the pubertal growth spurt is a logistic function: \( y = a/[1 + \exp(b(t - t_v))] \). Average total gain in height for boys is 15.4 cm (8.6%) and for girls is 10.9 cm (6.5%); \( v \) designates attained height at time \( t \) in years from birth; \( a, b, \) and \( c \) are constants; \( t_v \) is the age at peak height velocity. (Adapted from Karlberg J. On the construction of the infancy-childhood-puberty growth standard. Acta Paediatr Scand Suppl 1989; 356:2637.)
Figure 24-13 A schematic male growth chart with the features of the ICP (infancy, childhood, puberty) pattern overlaid and illustrating the predominant endocrine mechanisms controlling each phase of growth. The first shaded area emphasizes the decreasing velocity of infantile growth as the individual leaves the rapid growth phase of fetal life. The clear area is the childhood phase, which continues and magnifies the decreased velocity of growth into a plateau of rather constant growth during childhood. These two phases depend, in large part, on the effects of growth hormone (GH) and thyroid hormone with no or little effect derived from gonadal steroids. Finally, there is the period of the pubertal growth spurt in which gonadal steroids exert their direct and indirect effects. Gonadal steroids exert direct effects on the bone by stimulating the generation of insulin-like growth factor I (IGF-I) and other growth factors locally, and exert indirect effects by stimulating increased GH secretion which, in turn, exerts its own effects on bone and stimulates the production of IGF-I. In the female, the major gonadal steroid involved in the pubertal growth spurt is estradiol, whereas in the male, testosterone and estradiol (arising mainly from the aromatization of testosterone) are the major gonadal steroids.
**Figure 24-14** Interactions of the major growth-promoting hormones during puberty. Plus (+) indicates stimulatory action, minus (-) inhibitory action. Circulating insulin-like growth factor I (IGF-I) arises mainly from liver, but other tissues also contribute (endocrine action). Growth hormone and gonadal steroids have a direct stimulatory effect on the generation of IGF-I (paracrine action) locally in bone and cartilage cells. For simplification, the feedback loops for IGF-I and gonadal steroids on the hypothalamic-pituitary unit are omitted.
Figure 24-15  Serum insulin-like growth factor I (IGF-I; also called somatomedin C, SMC) in females and males stratified by age (left) and by pubertal stage (right). Males attain peak IGF-I levels at 15 years (2.5 ± 0.2 U/mL) at pubertal (genital) stage 3 (2.3 ± 0.2 U/mL). IGF-I concentrations reach a plateau between ages 12 and 15 in females (about 2 U/mL) and peak at pubertal (breast) stage 3 (2.5 ± 0.2 U/mL). The mean concentrations during puberty are higher than both adult and prepubertal values.
Figure 24-16a. Spine bone mineral density (BMD) for females by age. Mixed effects and semiparametric models were used to create the mean curves, and robust nonparametric smoothing techniques were employed for estimations of standard deviation (SD) and by age. Mean BMD was significantly greater in black (right) than in nonblack (Asian, Hispanic, and white) subjects (left). The solid line represents the mean level for age, and the dashed lines indicate the SD as indicated.
Figure 24-16b B. Spine BMD for males by age. The curve for black males was significantly greater than the mean levels of all nonblacks; Asian and white males had greater mean spine BMD than Hispanics. Mean and SD curves are shown. (A and B, From Bachrach LK, Hastie T, Wang M, et al. Bone mineral acquisition in healthy Asian, Hispanic, black and caucasian youth: A longitudinal study. J Clin Endocrinol Metab 1999; 84:4707.)
Figure 24-17 Mean plasma estradiol, follicle-stimulating hormone (FSH), and luteinizing hormone (LH) concentrations in prepubertal and pubertal females by pubertal stage of maturation (1 = prepubertal; 5 = menstruating adolescents) and the mean bone age for each stage. Single daytime values of gonadotropins have limited usefulness because of pulsatility of gonadotropin release and the increased amplitude of LH pulses during sleep through puberty. The gonadal steroid values, however, are useful in determining the stage of pubertal development. To convert FSH values (LER-869) to international units per liter, multiply by 8.4. To convert LH values (LER-960) to international units per liter, multiply by 3.8. To convert estradiol values to picomoles per liter, multiply by 3.671. (From Grumbach MM. Onset of puberty. In Berenberg SR [ed]. Puberty, Biologic and Social Components. Leiden, HE Stenfert Kroese, 1975, pp 121. Reprinted by permission of Kluwer Academic Publishers.)
Figure 24-18 Mean plasma testosterone and gonadotropin levels in normal boys by stage of maturation (1 = prepubertal) and mean bone age for each stage. (See legend for Fig 24-17.) To convert testosterone values to nanomoles per liter, multiply by 0.03467. (From Grumbach MM. Onset of puberty. In Berenberg SR [ed]. Puberty, Biologic and Social Components. Leiden, HE Stenfert Kroese, 1975, pp 121. Reprinted by permission of Kluwer Academic Publishers.)
Figure 24-19 Organization and characteristics of the hypothalamic-pituitary-gonadotroph-gonadal system. The medial basal hypothalamus (MBH) contains the transducer luteinizing hormone-releasing hormone (LHRH) neurosecretory neurons. These neurons translate neural signals into a periodic, oscillatory chemical signal, LHRH. This MBH complex functions as an LHRH pulse generator (oscillator), which is frequency coded and releases LHRH from its axon terminals at the median eminence as a largely synchronous intermittent discharge into the primary capillary plexus of the hypothalamo-hypophysial portal circulation. The LHRH pulse generator is influenced by biogenic amine neurotransmitters, peptidergic neuromodulators, neuroexcitatory amino acids, and neural pathways. During the follicular phase in the adult female and the adult male, an LHRH pulse (estimated indirectly by monitoring LH pulses in peripheral blood) occurs approximately every 90 to 120 minutes throughout the day. Changes in the frequency and probably in the amplitude of the LHRH secretory episodes modulate the pattern of LH and follicle-stimulating hormone (FSH). The major site of action of testosterone and progesterone is on the LHRH pulse generator, as these two classes of steroids decrease LH pulse frequency, but a pituitary site of action has also been described. Estrogens have major direct inhibitory and stimulatory effects on the LHRH-primed pituitary gonadotroph: the inhibitory, or negative, feedback action is associated with a decrease in both the frequency and the amplitude of pituitary LH secretion. On the other hand, evidence also supports a negative and positive feedback action of estrogen on the LHRH pulse generator. Inhibin has a direct inhibitory effect on the pituitary gland and the secretion of FSH. The secretion of gonadal steroids by the gonads is controlled mainly by the amplitude of the gonadotropin signal. (Adapted from Grumbach MM, Kaplan SL. The neuroendocrinology of human puberty: an ontogenetic perspective. In Grumbach MM, Sizonenko PC, Aubert ML [eds]. Control of the Onset of Puberty. Baltimore, Williams & Wilkins, 1990, pp 168. © 1990, The Williams & Wilkins Co., Baltimore.)
Figure 24-20 The immortalized hypothalamic luteinizing hormone-releasing hormone (LHRH) (LHRH) neuronal cell line. Left, Phase-contrast micrograph illustrating the neuronal phenotype (GT13 cell line) including the extension of multiple long neurites, cell-cell contacts, and growth cones. The neuroendocrine function of GT cells is limited to expression of LHRH and GAP. Magnification × 175. Right, Demonstration of autonomous LHRH (gonadotropin-releasing hormone, GnRH) pulses at about 20-minute intervals by the LHRH neurons in culture. This is the same frequency as that for LH pulses in vivo, in castrated adult mice and rats. To convert LHRH values to picomoles per liter, multiply by 0.8460. (Micrograph from Mellon PL, Windle II, Goldsmith PC, et al. Immortalization of hypothalamic GnRH neurons by genetically targeted tumorigenesis. Neuron 1990; 5:110. Copyright by Cell Press. Graph courtesy of G. Martinez de la Escalera and R. I. Weiner.)
Figure 24-21 Effect of pulsatile administration of luteinizing hormone-releasing hormone (LHRH) in contrast to continuous infusion of LHRH in adult oophorectomized rhesus monkeys in which gonadotropin secretion has been abolished by lesions that ablated the medial basal hypothalamic LHRH pulse generator. Note the high concentrations of plasma LH and follicle-stimulating hormone (FSH) in monkeys given one LHRH pulse per hour, the suppression of gonadotropin secretion by continuous infusion of LHRH even though the total dose of LHRH was the same, and the restoration of FSH and LH secretion when the pulsatile mode of LHRH administration was reinitiated. (From Belchetz PE, Plant TM, Nakai Y, et al. Hypophysial responses to continuous and intermittent delivery of hypothalamic gonadotropin releasing hormone. Science 1978; 202:631633. Copyright 1978 by the AAAS.)
Figure 24-22 Ontogeny of luteinizing hormone-releasing hormone (LHRH) neurons in the mouse. The route of migration of the LHRH neurosecretory neurons (black dots) in the mouse embryo is shown from their origin in the medial olfactory placode (a plate-like thickening of embryonic ectoderm) in the nasal region through the forebrain into the hypothalamus and preoptic areas. At embryonic (E) day 11 to 11.5 LHRH cells are in the anlage of the vomeronasal organ and medial wall of the olfactory placode. By E day 13 the number of LHRH neurons has increased, and most are in the nasal septum with the nervus terminalis and the vomeronasal nerves; only a few cells are in the brain. By E day 14 the majority of LHRH cells are in the ganglion terminale and the central root of the nervus terminalis and arch through the forebrain to the hypothalamus. By E day 16 most of the LHRH neurons are in the hypothalamus and preoptic areas, and the migration is almost complete. GT, ganglion terminale; OB, olfactory bulb; POA, preoptic area, VNO, vomeronasal organ. (Adapted from Schwanzel-Fukuda M, Pfaff DW. Origin of luteinizing hormone-releasing hormone neurons. Nature 1989; 338:161-164. Reprinted by permission from Nature, Vol. 338, pp. 161-164. Copyright © 1989 Macmillan Magazines Ltd.)
Figure 24-23 Ontogeny of the luteinizing hormonereleasing hormone (LHRH) neurons in the rhesus monkey. In the 36-day embryo the LHRH cells (black dots) are located deep in the nasal septum along the path of the nervus terminalis but not within the brain. By day 38 LHRH cells are clustered along the dorsal region of the olfactory bulbs and nervus terminalis with a few cells arching back along the ventral surface of the forebrain. By 55 days the LHRH neurons are in the process of migration, but clusters of LHRH cells have entered the central nervous system and reached the basal hypothalamus. BH, basal hypothalamus; LT, lamina terminalis; LV, lateral ventricle; NA, nasal area; NE, nasal epithelium; NT, nervus terminalis; OB, olfactory bulb; OC, optic chiasm; Tu, olfactory tubercle. (Adapted from Ronnekleiv OK, Resko JA. Ontogeny of gonadotropin-releasing hormonecontaining neurons in early fetal development of rhesus macaques, Endocrinology 1990; 126:498511. © by The Endocrine Society.)
**Figure 24-24** Comparison of the pattern of change of serum testosterone, human chorionic gonadotropin (hCG), and serum and pituitary luteinizing hormone (LH) (LER-960) and follicle-stimulating hormone (FSH) (LER-869) levels in the human male fetus during gestation in relation to the morphologic changes in fetal testes. The top graph illustrates the regression curve for the increment (Δ) between a baseline plasma LH and FSH level and the 15-minute response to administration of LHRH to the male fetus plotted as a function of gestational age. The scale masks the slight increase in plasma FSH. Data were recalculated from Takagi and colleagues. The evidence supports the hypothesis that the hypothalamic LHRH pulse generator is functional early in gestation and mediates the rise in serum concentration of fetal pituitary gonadotrophs. To convert plasma hCG values to international units per liter, multiply by 1.0. Other conversions are in the legends of Figure 24-17 and Figure 24-18. (Modified from Kaplan SL, Grumbach MM. Pituitary and placental gonadotropins and sex steroids in the human and subhuman primate fetus. Clin Endocrinol Metab 1978; 7:487511; and Gluckman PD, Grumbach MM, Kaplan SL. The human fetal hypothalamus and pituitary gland. In Tulchinsky D, Ryan KJ [eds]. Maternal-Fetal Endocrinology. Philadelphia, WB Saunders, 1980, pp 196232.)
Figure 24-25 Pattern of change of serum follicle-stimulating hormone (FSH), luteinizing hormone (LH), and human chorionic gonadotropin (hCG) levels; concentration of pituitary FSH and LH; and increment (Δ) between baseline FSH and LH and the 15-minute response to administration of LHRH in the human female fetus during gestation with the development of the fetal ovary. See legends of Figure 24-17, Figure 24-18, and Figure 24-24 for conversions to SI units. (Modified from Kaplan SL, Grumbach MM. Pituitary and placental gonadotropins and sex steroids in the human and subhuman primate fetus. Clin Endocrinol Metab 1978; 7:487-511.)
Figure 24-26  Pulsatile luteinizing hormone (LH) secretion in the ovine fetus. GA, gestational age. The length of gestation is 145 days in the sheep.  (From Clark SJ, Ellis N, Styne DM, et al. Hormone ontogeny in the ovine fetus. XVII. Demonstration of pulsatile luteinizing hormone secretion by the fetal pituitary gland. Endocrinology 1984; 115:17741779. © by The Endocrine Society.)
Figure 24-27 Left, The effect in the ovine fetus of administration for 7 days of luteinizing hormone-releasing hormone (LHRH) agonist (10 µg intravenously daily) on the acute LH response to LHRH agonist. Right, Recovery of the LH response was impaired 8 days after discontinuing LHRH agonist administration to the ovine fetus. (From Grumbach MM, Kaplan SL. The neuroendocrinology of human puberty: an ontogenetic perspective. In Grumbach MM, Sizonenko PC, Aubert ML [eds]. Control of the Onset of Puberty. Baltimore, Williams & Wilkins, 1990, pp 168. © 1990, the Williams & Wilkins Co., Baltimore.)
Figure 24-28 Change in the pattern of pulsatile follicle-stimulating hormone (FSH) and luteinizing hormone (LH) secretion in early infancy, childhood, and puberty. The data for early infancy are derived from Waldhauser and colleagues. Note the pulsatile secretion in the infant and the striking difference in the amplitude of FSH and LH pulses between male and female infants. After infancy, the amplitude and frequency of gonadotropin pulses decrease greatly for almost a decade (juvenile pause) until the onset of puberty. (From Grumbach MM, Kaplan SL. The neuroendocrinology of human puberty: an ontogenetic perspective. In Grumbach MM, Sizonenko PC, Aubert ML [eds]. Control of the Onset of Puberty. Baltimore, Williams & Wilkins, 1990, pp 168. © 1990, the Williams & Wilkins Co., Baltimore.)
Figure 24-29 Mean leptin ±95% CI (log scale) by puberty stage in boys ( and ---) and girls ( and ), with significant sex differences marked. *, $P < 0.05$; **, $P < 0.005$; ***$, P < 0.0005$. (From Ahmed ML, Ong KK, Morrell NJ, et al. Longitudinal study of leptin concentrations during puberty. J Clin Endocrinol Metab 1999; 84:902.)
Figure 24-30 The postulated action of leptin secreted by adipocytes on the hypothalamic luteinizing hormone-releasing hormone (LHRH) pulse generator. Its indirect action through hypothalamic neural networks is illustrated as well as direct action. Leptin appears to function as a permissive factor, not a trigger, in the onset of human puberty. Although leptin is reported to advance puberty in rodents, its role in "triggering" puberty in humans has not been established and is speculative. FSH, follicle-stimulating hormone; mRNA, messenger ribonucleic acid. (See text.)
Figure 24-31 Postulated dual mechanism of restraint of puberty involves both gonadal steroid-dependent and gonadal steroid-independent (intrinsic central nervous system inhibitory mechanism) processes. (Modified from Grumbach MM, Kaplan SL. The neuroendocrinology of human puberty: an ontogenetic perspective. In Grumbach MM, Sizonenko PC, Aubert ML [eds]. Control of the Onset of Puberty. Baltimore, Williams & Wilkins, 1990, pp 168. © 1990, the Williams & Wilkins Co., Baltimore.)
Figure 24-32 Change in pattern of the plasma concentration of follicle-stimulating hormone (FSH) with age in 58 patients with the syndrome of gonadal dysgenesis. Mixed longitudinal (n = 23) and cross-sectional (n = 35) data. Triangles designate patients with 45,X karyotype. Circles indicate Turner's syndrome patients with X chromosome mosaicism or structural abnormalities of the X chromosome, or both. Note the values in the 2- and 3-day-old infants. The solid line represents a regression line of best fit. The hatched area indicates the mean plasma values in normal females. To convert FSH values to international units per liter, multiply by 8.4. (From Conte FA, Grumbach MM, Kaplan SL. A diphasic pattern of gonadotropin secretion in patients with the syndrome of gonadal dysgenesis. J Clin Endocrinol Metab 1975; 40:670674. © by The Endocrine Society.)
Figure 24-33 Effect of administration of ethinylestradiol (2 µg/day) on the urinary excretion of luteinizing hormone (LH) and follicle-stimulating hormone (FSH) in a prepubertal normal male aged 11 years, 2 months. Note the rapid and significant decrease in LH and FSH levels by day 3 after treatment with estradiol; by day 4 the excretion of FSH and LH is less than 0.01 IU. (From Kelch RP, Kaplan SL, Grumbach MM. Suppression of urinary and plasma follicle-stimulating hormone by exogenous estrogens in prepubertal and pubertal children. Reproduced from the Journal of Clinical Investigation, 1973, vol. 52, pp. 11221128 by copyright permission of the American Society for Clinical Investigation.)
Figure 24-34 Interaction of the negative feedback mechanism and the putative intrinsic central nervous system (CNS) inhibitory mechanism in restraining puberty as extrapolated from the pattern of change in the concentrations of follicle-stimulating hormone (FSH) and luteinizing hormone (LH) in agonadal infants, children, and adolescents. (See Figure 24-33 for key to symbols; the solid line is the regression curve of best fit; the solid bars connote the mean normal concentrations + 1 SD of FSH and LH.) For about the first 3 years of life the sensitive gonadal steroid, negative feedback mechanism has a dominant role in restraining gonadotropin secretion, as exemplified by the high gonadotropin concentrations in this age group in the absence of gonads (and gonadal steroid feedback). A major role of the intrinsic CNS inhibitory mechanism in this age group is unlikely in light of the rise in gonadotropins to castrate levels in the absence of functional gonads. From 4 to 6 years of age the postulated intrinsic CNS inhibitory mechanism is dominant, as indicated by the fall in FSH and LH concentrations in the absence of gonads. Even in this age group the augmented gonadotropin response evoked by LHRH and the slightly higher mean basal gonadotropin concentrations in agonadal individuals support a role, although a subsidiary one, for gonadal steroid negative feedback in the suppression of gonadotropin secretion during this period of the juvenile pause. The authors suggest that the intrinsic CNS inhibitory mechanism suppresses the functional LHRH pulse generator. Finally, after about 10 years of age the CNS inhibition gradually wanes, resulting in disinhibition of the LHRH pulse generator. The gonadal steroid negative feedback mechanism with an adult-type set point and inhibin play a dominant role in regulating the LHRH pulse generator-pituitary gonadotropin system. For conversion to SI units, see the legend of Fig. 24-17 (Modified from Grumbach MM, Kaplan SL. The neuroendocrinology of human puberty: an ontogenetic perspective. In Grumbach MM, Sizonenko PC, Aubert ML [eds]. Control of the Onset of Puberty. Baltimore, Williams & Wilkins, 1990, pp 168. © 1990, the Williams & Wilkins Co., Baltimore.)
Figure 24-35 Postulated ontogeny of the dual mechanism for the inhibition of puberty. Interrupted arrows indicate inhibition. Note the action of both components during the juvenile pause (prepuberty). See Figure 24-34 for the relative role of these two mechanisms during development. LHRH is given as LRF in the figure; MBH, medial basal hypothalamus. (Modified from Grumbach MM, Kaplan SL. The neuroendocrinology of human puberty: an ontogenetic perspective. In Grumbach MM, Sizonenko PC, Aubert ML [eds]. Control of the Onset of Puberty. Baltimore, Williams & Wilkins, 1990, pp 168. © 1990, the Williams & Wilkins Co., Baltimore.)
Figure 24-36. A. True precocious puberty in a 2 9/12-year-old girl (SM) secondary to a large bilateral congenital suprasellar arachnoid cyst. Signs of sexual precocity were noted during the preceding year. The head circumference was +5 SD above the mean value for age, and frontal bossing was present. Breasts were Tanner stage 3. Serum estradiol, 26 pg/mL; estrone, 38 pg/mL; dehydroepiandrosterone sulfate (DHEAS), less than 3 µg/dL. The serum luteinizing hormone (LH) concentration rose from 1.4 to 8.7 ng/mL (LER-960) after IV administration of LH-releasing hormone (LHRH), a pubertal response. Bone age, 3 6/12 years. Pelvic ultrasonography showed pubertas-like uterus and ovaries. To convert estrone values to picomoles per liter, multiply by 3.899. To convert DHEAS values to micromoles per liter, multiply by 0.02714. For other conversions see legend of Figure 24-17. B. Cranial CT scans for SM showing low-density fluid collection in the middle cranial fossa, thinning of the cortex, and striking compression of the lateral and third ventricles. C. Cranial CT scans 8 months later, after decompression of the arachnoid cyst and creation of a communication between the cyst and the basal cerebrospinal fluid cisterns and a cystoperitoneal shunt. Note the striking decrease in size of the fluid collections and expansion of the cerebral cortex. D. Basal and peak LH and follicle-stimulating hormone (FSH) concentrations after LHRH administration in SM and serum estradiol values before and 2 weeks and 9 months after surgical decompression of the arachnoid cyst. Note prepubertal LH response to LHRH and fall in serum estradiol level by 9 months after surgery. The bone age had increased by 3 years over an 11-month period, but the velocity has now returned to normal. The patient remained prepubertal during follow-up. (From Grumbach MM, Kaplan SL. The neuroendocrinology of human puberty: an ontogenetic perspective. In Grumbach MM, Sizonenko PC, Aubert ML [eds]. Control of the Onset of Puberty. Baltimore, Williams & Wilkins, 1990, pp 168. © 1990, the Williams & Wilkins Co., Baltimore.)
Figure 24-37 Hypothalamic hamartoma as an ectopic luteinizing hormone-releasing hormone (LHRH) pulse generator that escapes the intrinsic central nervous system inhibitory mechanism and results in true precocious puberty. Two possible mechanisms are proposed. **Left**, The LHRH neurosecretory neurons in the hamartoma functioning as an LHRH pulse generator without activation of the suppressed normally located LHRH pulse generator. **Right**, The hamartoma acting as an ectopic LHRH pulse generator but communicating with and activating (possibly through axonic connections or by LHRH itself) the normally located hypothalamic LHRH pulse generator, which then functions synchronously with the hamartoma.
Figure 24-38 The striking developmental changes in GABA and gonadotropin-releasing hormone (GnRH; luteinizing hormonereleasing hormone) release between the prepubertal and the pubertal rhesus monkey as measured in 10-minute perfusate samples from the stalk median eminence. In each animal, multiple samples were obtained. Mean ± SEM. ** $P < .01$; * $P < .05$ versus prepubertal monkeys. *(From Mitsushima D, Hei DL, Terasawa E. *Aminobutyric acid is an inhibitory neurotransmitter restricting the release of luteinizing hormonereleasing hormone before the onset of puberty. Proc Natl Acad Sci USA 1994; 91:395399.*)
Figure 24-39 The yin and the yang of the neuroendocrinology of the prepubertal juvenile pause and its intrinsic central inhibition of the luteinizing hormonereleasing hormone (LHRH) pulse generator and the reversal of this inhibition and termination of the juvenile pause, which leads to the onset of puberty. The GABAergic neuronal network and its neurotransmitter -aminobutyric acid (GABA) constitute the most ubiquitous inhibitory transmitter in the hypothalamus as well as the brain. During the prepubertal juvenile pause, this neurotransmitter system appears to play the major neural role in inhibiting the LHRH pulse generator. (Suppression of GABA inhibition during this period promptly results in reactivation of the suppressed LHRH pulse generator in the rhesus monkey.) With the approach of puberty GABA inhibition of the LHRH pulse generator wanes, and its reactivation gradually occurs. This reactivation is quite likely augmented by stimulatory neurotransmitters (e.g., excitatory amino acids), some of which are dependent on increased gonadal steroids for their activation, and by neurotrophic factors and growth peptides. As a consequence, the amplitude and, to a lesser extent, the frequency of LHRH pulses increase, which, in turn, leads to increased pulsatile secretion of follicle-stimulating hormone (FSH) and LH and the activation of the ovary and testis. As shown experimentally in the monkey, the LHRH pulse generator can function in the absence of hypothalamic stimulatory factors. The nature of and factor or factors responsible for this transition from central inhibition and the postulated dominance of GABA to the release of inhibition and reactivation of the LHRH pulse generator are unknown.
Figure 24-40 Plasma luteinizing hormone (LH) and testosterone sampled every 20 minutes in a 14-year-old boy in pubertal stage 2. The histogram displaying sleep stage sequence is depicted above the period of nocturnal sleep. Sleep stages are rapid eye movement (REM) with stages I to IV shown by depth of line graph. Plasma LH is expressed as mIU/mL. Plasma testosterone is expressed as nanograms per 100 mL. To convert LH values to international units per liter, multiply by 1.0. To convert testosterone values to nanomoles per liter, multiply by 0.03467. (From Boyar RM, Rosenfeld RS, Kapen S, et al. Human puberty: simultaneous augmented secretion of luteinizing hormone and testosterone during sleep. Reproduced from the Journal of Clinical Investigation, 1974, vol. 54, pp. 609618 by copyright permission of the American Society for Clinical Investigation.)
Figure 24-41  Changes in plasma luteinizing hormone (LH) (top) and follicle-stimulating hormone (FSH) (bottom) levels in prepubertal, pubertal, and adult individuals. Note the limited LH response in prepubertal children compared with that of pubertal and adult subjects. The FSH response to LH-releasing hormone (LHRH) is similar in prepubertal, pubertal, or adult males. In females, the FSH response is significantly greater than that of prepubertal, pubertal, or adult males. For conversion to SI units, see the legend of Figure 24-17. (Modified from Grumbach MM, Roth JC, Kaplan SL, et al. Hypothalamic pituitary regulation of puberty in man: evidence and concepts derived from clinical research. In Grumbach MM, Grave GD, Mayer FE [eds]. Control of the Onset of Puberty. New York, John Wiley & Sons, 1974, pp 115166.)
**Figure 24-42** Relation of plasma dehydroepiandrosterone sulfate (DHEAS [DHAS]) to growth of the zona reticularis and increase in adrenal volume with age. **Top,** The close correlation between the development of the zona reticularis and the increase in plasma DHEAS level. **Middle,** The age at which either focal islands of reticular tissue or a continuous reticular zone was found in a series of patients with sudden death who had not had an antecedent illness. **Bottom,** The increase in adrenal volume at the time of puberty. For conversion to SI units, see the legend of **Figure 23-37.** (From Grumbach MM, Richards HE, Conte FA, et al. Clinical disorders of adrenal function and puberty: assessment of the role of the adrenal cortex and abnormal puberty in man and evidence for an ACTH-like pituitary adrenal androgen stimulating hormone. In James VHT, Serio M, Giusti G, et al [eds]. The Endocrine Function of the Human Adrenal Cortex. New York, Academic Press, 1978, pp 583612.)
Figure 24-43 Adrenarche and the zona reticularis. The rise in circulating dehydroepiandrosterone sulfate (DHEAS) is the biochemical hallmark of adrenarche. The diagram compares and contrasts the major steroidogenic pathway in the zona fasciculata with that in the zona reticularis. In contrast to the zona fasciculata, the expression of 3-hydroxysteroid, 4,5-isomerase type 2 messenger ribonucleic acid (mRNA) and its activity (the enzyme that irreversibly traps 5-precursors into 4-steroids) is very low in the zona reticularis, whereas the expression of and activity of steroid sulfotransferase is high. A single gene, CYP17, encodes a single enzyme that has both 17-hydroxylase and 17,20-lyase activity, but the ratio of 17,20-lyase to 17-hydroxylase activity is relatively high in the zona reticularis compared with that in the zona fasciculata. Some of the factors that seem to augment the increased 17,20-lyase activity of CYP17 are the augmented serine phosphorylation of the enzyme and the apparent increased abundance of the electron-donating redox partner, including P450 reductase and of cytochrome b5. (See text.)
Figure 24-44 Hypothesis of the control of pituitary adrenal androgen secretion by a putative separate adrenal androgenstimulating hormone acting on a corticotropin (ACTH)-primed adrenal cortex. Although this diagram suggests that "AASH" arises from the pituitary gland, a distinct pituitary factor with AASH activity has not been isolated; an extrapituitary factor is not excluded. The lower part of the diagram shows the relationship of adrenarche to gonadarche, including dissociation in various clinical disorders of sexual development (+, present; -, absent). (Modified from Sklar CA, Kaplan SL, Grumbach MM. Evidence for dissociation between adrenarche and gonadarche: studies in patients with idiopathic precocious puberty, gonadal dysgenesis, isolated gonadotropin deficiency, and constitutionally delayed puberty. J Clin Endocrinol Metab 1980; 51:548556. © by The Endocrine Society.)
Figure 24-45 A boy 16 years, 2 months of age, with constitutional delay in growth and puberty. Height, 149.5 cm (4 SD below the mean value for age); upper/lower body ratio, 1.1 (retarded for age); phallus, 6.0 × 1.6 cm; testes, 2.5 × 1.4 cm; the scrotum showed early thinning. At a chronologic age of 15 years, 4 months the bone age was 11 years and the sella turcica was normal. The plasma concentration of luteinizing hormone (LH) was 0.7 ng/mL (LER-960); follicle-stimulating hormone (FSH), 0.5 ng/mL (LER-869). On LH-releasing hormone (LHRH) testing the plasma concentration of LH increased to 2.2 ng/mL (an increment of 1.5 ng/mL), and the testosterone level rose from 52 to 77 ng/dL. The testes subsequently spontaneously enlarged, and the patient progressed through puberty. For conversion to SI units, see the legends of Figure 24-17 and Figure 24-18. (From Styne DM, Grumbach MM. Puberty in the male and female: its physiology and disorders. In Yen SCC, Jaffe RB [eds]. Reproductive Endocrinology, 2nd ed. Philadelphia, WB Saunders, 1986, pp 313384.)
Figure 24-46 The various patterns of pulsatile luteinizing hormone (LH) secretion that can occur in isolated hypogonadotropic hypogonadism (B to D) compared with LH secretion in a normal man (A). A, The discrete LH pulses occurring about every 2 hours in a normal 36-year-old man. B, Typical apulsatile LH pattern associated with a low testosterone concentration usually found in isolated hypogonadotropic hypogonadism. C, Pattern of developmental arrest with low-amplitude nocturnal LH pulses apparent only during sleep. D, Low-amplitude LH pulse pattern during sleep and wake periods. To convert LH values to international units per liter, multiply by 1.0. (From Spratt DI, Crowley WF. Hypogonadotropic hypogonadism: GnRH therapy. In Krieger DT, Bardin CW [eds]. Current Therapy in Endocrinology and Metabolism, 19851986. Toronto, BC Decker, 1985, pp 155159.)
Figure 24-47  Craniopharyngioma in a short 5-year-old girl with a history of frontal headaches, impaired vision, and poor growth.  Left, Midline sagittal T1-weighted image that shows a hyperintense region superiorly and an inferior hypointense region. The combination of hyper- and hypointense areas in a noncontrast-enhanced examination is the most characteristic finding in craniopharyngioma. Note erosion of dorsum sellae (solid white arrow) and posterior pituitary bright spot. Right, Coronal-weighted T1 image shows tumor extending upward to the inferior frontal horns, narrowing the foramen of Monro and causing mild hydrocephalus. The open white arrows indicate the upper border of the hyperintense area of the tumor.
Figure 24-48 A girl 18 years, 8 months of age, with isolated gonadotropin deficiency (sexual infantilism and primary amenorrhea). Height was 173 cm (+1 SD), weight was 66.5 kg (+1 SD), and skeletal age was 13 years. Adrenarche with pubic hair development occurred at age 13 ½ years. At the time of the photograph, pubic hair was in stage 3 and there was slight breast and nipple development resulting from a previous short course of estrogen therapy. Immature labia minora and majora were noted, and no estrogen effect was present on the vaginal mucosa. Olfactory testing was normal. The plasma luteinizing hormone (LH) (LER-960) level after LH-releasing hormone (LHRH) administration rose from 0.5 to 1.8 ng/mL (a prepubertal response). Serum estradiol was undetectable. The dehydroepiandrosterone sulfate (DHEAS) level was 92 µg/dL (appropriate for pubic hair stage 2). Note the discrepancy between adrenarche and gonadarche. For conversion to SI units, see the legends of Figure 24-17 and Figure 24-36. (From Styne DM, Grumbach MM. Puberty in the male and female: its physiology and disorders. In Yen SCC, Jaffe RB [eds]. Reproductive Endocrinology, 2nd ed. Philadelphia, WB Saunders, 1986, pp 313384.)
Figure 24-49 A boy of 15 years, 10 months, with isolated gonadotropin deficiency and anosmia (Kallmann's syndrome). He had undescended testes, but after administration of 10,000 U of human chorionic gonadotropin (hCG) the testes descended and were palpable in the scrotum. Height, 163.9 cm (-1.5 SD); the upper/lower body ratio was 0.86, which is eunuchoid. The phallus measured 6.3 × 1.8 cm, and the testes were 1.2 × 0.8 cm. The concentration of plasma luteinizing hormone (LH) was less than 0.3 ng/mL; of follicle-stimulating hormone (FSH), 1.2 ng/mL; of testosterone, 16 ng/dL. After 100 µg of LH-releasing hormone (LHRH) the plasma LH (LER-960) was 0.7 ng/mL and FSH (LER-869) 2.4 ng/mL. For conversion to SI units, see the legends of Figure 24-17 and Figure 24-18. (From Styne DM, Grumbach MM. Puberty in the male and female: its physiology and disorders. In Yen SCC, Jaffe RB [eds]. Reproductive Endocrinology; 2nd ed. Philadelphia, WB Saunders, 1986, pp 313384.)
Figure 24-50  Serum luteinizing hormone (LH) and follicle-stimulating hormone (FSH) responses to the administration of LH-releasing hormone (LHRH) in 25 males with an isolated gonadotropin deficiency with or without anosmia, segregated according to whether the volume of the testes was prepubertal or greater than 2.5 cm$^3$; testicular volume in those with testes greater than 2.5 cm$^3$ were as large as 4 cm$^3$. Basal and LHRH-stimulated gonadotropin levels after the intravenous injection of 100 µg LHRH (peak value) are shown. *$P < .05$. For conversion to SI units, see the legend of Figure 24-17. (From Van Dop C, Burstein S, Conte FA, et al. Isolated gonadotropin deficiency in boys: clinical characteristics and growth. J Pediatr 1987; 111:684-692.)
Figure 24-51 Comparison of the brain and nasal cavities of a normal 19-week-old male fetus (upper left) and those of a male fetus of similar age with Kallmann's syndrome caused by an X chromosome deletion at Xp22.3 (upper right). In the normal fetal brain the luteinizing hormone-releasing hormone (LHRH) neurosecretory neurons (black dots) are located in the hypothalamic area including the medial basal hypothalamus; the anterior hypothalamic area; and, of interest regarding hypothalamic hamartoma as an ectopic LHRH pulse generator, the pre- and retro-mammillary areas. A small cluster of LHRH neurons is present among the fibers of the terminalis nerve on the floor of the nasal septum. In the male fetus with Kallmann's syndrome, no LHRH neurons were detected in the hypothalamic region including the basal hypothalamus, median eminence, and preoptic area. The LHRH cells fail to migrate to and enter the brain from their origin in the nose; these cells end in a tangle beneath the forebrain on the dorsal surface of the cribriform plate and in the nasal cavity. AC, anterior commissure; CG, crista galli; IN, infundibular nucleus; NT, terminalis nerve; OC, optic chiasm; POA, preoptic area. (Adapted from Schwanzel-Fukuda M, Bick D, Pfaff DW. Luteinizing hormone-releasing hormone (LHRH)-expressing cells do not migrate normally in an inherited hypogonadal (Kallmann) mouse. Mol Brain Res 1989; 6:311-326.) Lower panels show magnetic resonance imaging scans of brain (coronal section, TI-weighted image). Lower left, Normal olfactory sulci (open white arrows) and bulbs (small solid white arrows) in a 15-year-old boy. Lower right, Absent olfactory sulci (open white arrows) and bulbs in a 17-year-old anosmic, sexually infantile boy with Kallmann's syndrome.
A 20-year-old male with idiopathic hypopituitary dwarfism and deficiencies of gonadotropins, thyrotropin, corticotropin, and growth hormone, who had a history of arrested hydrocephalus. Height, 129 cm (-8 SD); the phallus was 2 cm in length, and the testes measured 1.5 × 1 cm. He had received thyroid and glucocorticoid replacement. Basal luteinizing hormone (LH) was less than 0.2 ng/mL (LER-960), follicle-stimulating hormone (FSH) was 0.5 ng/mL (LER-869), and testosterone was less than 0.1 ng/mL. In response to 100 µg of LH-releasing hormone (LHRH), the plasma LH concentration increased slightly to 0.6 ng/mL, and there was no increase in plasma testosterone. The excretion of urinary 17-ketosteroids was 1.1 mg/24 hours. The bone age was 10 years, and the volume of the sella turcica was small on skull radiographs. For conversion to SI units, see the legends of Figure 24-17 and Figure 24-18. (From Styne DM, Grumbach MM. Puberty in the male and female: its physiology and disorders. In Yen SCC, Jaffe RB [eds]. Reproductive Endocrinology, 2nd ed. Philadelphia, WB Saunders, 1986, pp 313384.)
Figure 24-53 47,XXY Klinefelter's syndrome in 17-year-old identical twins. At age 15 gynecomastia was noted. The twins had a eunuchoid habitus and poorly developed male secondary sexual characteristics. Both were 187 cm in height; arm spans were 187 cm and 189.5 cm; the voices were high-pitched; the testes measured 1.8 × 1.5 cm; penis length was 7.5 cm. Gynecomastia and signs of androgen deficiency were more evident in the twin on the left. Urinary gonadotropins, greater than 50 mU/24 hours. The testes exhibited extensive tubular fibrosis, small dysgenetic tubules, and clumping or pseudoadenomatous formation of Leydig cells; germ cells were rare. The microscopic appearance was typical of seminiferous tubule dygenesis. (Patients are described in Grumbach MM, Barr ML. Cytologic tests of chromosome sex in relation to sexual anomalies in man. Recent Prog Horm Res 1958; 14:255324.)
Figure 24-54 Left, A 14 10/12-year-old patient with the typical form of the syndrome of gonadal dysgenesis (Turner's syndrome). The X chromatin pattern was negative, and the karyotype was 45,X. She was short (height 134.5 cm; height age 9 5/12 years) and sexually infantile except for the appearance of sparse pubic hair, and exhibited characteristic stigmata of the syndrome: a short webbed neck, shield-like chest with widely separated nipples, bilateral metacarpal signs, puffiness over the dorsum of the fingers, cubitus valgus, increased number of pigmented nevi, characteristic facies, and low-set ears. The bone age was 13 6/12 years; urinary 17-ketosteroids 5.1 mg/day; urinary gonadotropin greater than 100 mU/day. Vaginal smears and the urocytogram showed an immature pattern in which cornified squamous cells were absent. With estrogen therapy, female secondary sexual characteristics were induced; the cyclic administration resulted in periodic estrogen withdrawal bleeding. Right, A 45,X, 9 11/12-year-old patient with Turner's syndrome. Apart from short stature (height 118 cm; age 6 10/12 years), increased pigmented nevi, and subtle changes in the fingers and toes, she had few somatic anomalies. In contrast to the patient at the left, the main clinical feature was short stature.
Figure 24-55 The evaluation of delayed puberty in boys.
Figure 24-56 The evaluation of delayed puberty in girls.
Figure 24-57 Age at onset of idiopathic true precocious puberty in 106 children. Open bars, female; hatched bars, male. At all ages, the frequency is greater in females than in males. The peak prevalence in girls is between ages 6 and 8 years. (From Kaplan SL, Grumbach MM. The neuroendocrinology of human puberty: an ontogenetic perspective. In Grumbach MM, Sizonenko PC, Aubert ML [eds]. Control of the Onset of Puberty. Baltimore, Williams & Wilkins, 1990, pp 168. © 1990, the Williams & Wilkins Co., Baltimore.)
Figure 24-58 Left, A boy 2 years, 5 months of age with idiopathic precocious puberty. He had pubic hair and phallic and testicular enlargement by 10 months of age. At 1 year of age, his height was 86 cm (+4 SD); the phallus measured 10 × 3.5 cm, and the testes measured 2.5 × 1.5 cm. Plasma luteinizing hormone (LH) was 1.9 ng/mL (LER-960); follicle-stimulating hormone (FSH) 1.2 ng/mL (LER-869); and testosterone 416 ng/dL. After 100 µg of LH-releasing hormone (LHRH), the plasma LH increased to 8.4 ng/mL and FSH to 1.8 ng/mL, a pubertal response. When photographed, the patient had been treated with medroxyprogesterone acetate for 1.5 years. His height was 95.2 cm (+1 SD), the phallus was 6 × 3 cm, and the testes were 2.4 × 1.3 cm. Basal concentrations of LH (LER-960) were 0.9 ng/mL; FSH (LER-869) 0.8 ng/mL; and testosterone 7 ng/dL. After 100 µg of LHRH, LH concentrations rose to 2.3 ng/mL, whereas FSH concentrations did not change when he was on treatment with medroxyprogesterone acetate. For conversion to SI units, see the legends of Figure 24-17 and Figure 24-18. (Left, From Styne DM, Grumbach MM. Puberty in the male and female: its physiology and disorders. In Yen SCC, Jaffe RB, [eds]. Reproductive Endocrinology, 2nd ed. Philadelphia, WB Saunders, 1986, pp 313384.) Right, A 3 3/12-year-old girl with idiopathic true precocious puberty who had recurrent vaginal bleeding since 9 months of age. Height age, 4 5/12 years; bone age, 8 10/12 years.
Figure 24-59  **Left,** Mean basal plasma luteinizing hormone (LH) level (LER-960) and mean peak and increment after intravenous LH-releasing hormone (LHRH) (100 µg) in normal prepubertal and pubertal females and in females with idiopathic true precocious puberty. The mean peak and increments of plasma LH are higher in true precocious puberty than in normal puberty.  **Right,** Basal follicle-stimulating hormone (FSH) level (LER-1364) and mean peak and increment after intravenous LHRH (100 µg) in normal prepubertal and pubertal females with true precocious puberty. The concentration of FSH and the response to LHRH were greater in females with true precocious puberty and normal puberty than in prepubertal females.  (From Kaplan SL, Grumbach MM. Pathogenesis of sexual precocity. In Grumbach MM, Sizonenko PC, Aubert ML [eds]. Control of the Onset of Puberty. Baltimore, Williams & Wilkins, 1990, pp 620-660. © 1990, the Williams & Wilkins Co., Baltimore.)
Figure 24-60  Left, Serial determinations of plasma estradiol in three girls with idiopathic true precocious puberty. Note the striking fluctuations in values.  Right, Serial determinations of plasma testosterone in three boys with true precocious puberty (B.L. and J.C. have a hypothalamic hamartoma; M.D. has the idiopathic form). For conversion to SI units, see the legends of Figure 24-17 and Figure 24-18.  (From Kaplan SL, Grumbach MM. Pathogenesis of sexual precocity. In Grumbach MM, Sizonenko PC, Aubert ML [eds]. Control of the Onset of Puberty. Baltimore, Williams & Wilkins, 1990, pp 620660. © 1990, the Williams & Wilkins Co., Baltimore.)
Figure 24-61 A, A 17-month-old male infant with hamartoma of the tuber cinereum and true precocious puberty. At 8 months of age, secondary sexual development was noted, and the patient was misdiagnosed as having congenital virilizing adrenal hyperplasia. He was treated with glucocorticoids, which slowed his growth but did not affect his sexual development and bone age advancement. When he was first seen at 17 months, height was 84.2 cm; weight was 14.8 kg; the pubic hair stage was stage II; the penis was 10.4 × 2.2 cm; the testes were 1.5 × 2.8 cm; and the scrotum was thinned and rugated. The bone age was 4 3/12 years. After luteinizing hormone-releasing hormone (LHRH) administration, the LH level rose from 0.5 to 3.1 ng/dL (LER-960), the follicle-stimulating hormone (FSH) level from 0.5 to 1.2 ng/mL (LER-869), and the testosterone level from 409 to 450 ng/dL. Dehydroepiandrosterone sulfate (DHEAS) was 17 µg/dL (preadrenarchal value). The patient was treated with a potent long-acting LHRH agonist deslorelin (D-Trp⁶ Pro⁹ NEt-LHRH), which resulted in arrest of his pubertal advancement and a striking decrease in the plasma concentration of testosterone, LH pulses, and the response to exogenous LHRH. B, Computed tomographic scan of the patient, demonstrating a 1.5-cm mass posterior and rostral to the dorsum sella, which depresses the flow of the third ventricle. For conversion to SI units, see the legends of Figure 24-17, Figure 24-18, and Figure 24-36. (From Styne DM, Grumbach MM. Puberty in the male and female: its physiology and disorders. In Yen SSC, Jaffe RB [eds]. Reproductive Endocrinology, 2nd ed. Philadelphia, WB Saunders, 1986, pp 31384.)
Figure 24-62  **Left,** Magnetic resonance imaging scan demonstrating a hypothalamic hamartoma (solid white arrow) in a 4-year-old boy with true precocious puberty; sagittal T1-weighted image. The posterior pituitary hot spot is designated by the solid black arrow.  **Right,** Computed tomographic brain scan (coronal section) showing an isodense, pedunculated, collar button-shaped hypothalamic hamartoma (arrow) in a 2-year-old girl with true precocious puberty.
Figure 24-63 Pulsatile luteinizing hormone (LH) secretion before and during LH-releasing hormone (LHRH) agonist therapy in a boy (right) and a girl (left) with true precocious puberty secondary to a hypothalamic hamartoma. For conversion to SI units, see the legend of Figure 24-17.
Figure 24-64 A boy of 8 years, 8 months with neurofibromatosis and precocious puberty, secondary to a hypothalamic glioma. He had tonic-clonic seizures at 2½ years and rapid growth starting at 4 years; an enlarged penis and testes and the presence of public hair were first noted at 7½ years. At this time, his height was 139.9 cm (+ 1.4 SD); the phallus was 9 × 3 cm; the right testis measured 5.5 × 3.2 cm and the left measured 5.4 × 2.9 cm. He had stage 3 pubic hair and 24 large café-au-lait spots. Computed tomographic scans and pneumoencephalography revealed a 1.5 × 2.5 cm hypothalamic mass, which was treated with radiation. The plasma concentration of luteinizing hormone (LH) was 0.5 ng/mL (LER-960); follicle-stimulating hormone (FSH) 0.4 ng/mL (LER-869); testosterone 221 ng/dL. After 100 µg of intravenous LH-releasing hormone (LHRH) the peak concentration of LH was 4.9 ng/mL, and that of FSH 1.4 ng/mL, a pubertal response. For conversion to SI units, see the legends of Figure 24-17 and Figure 24-18. (From Styne DM, Grumbach MM. Puberty in the male and female: its physiology and disorders. In Yen SCC, Jaffe RB [eds]. Reproductive Endocrinology, 2nd ed. Philadelphia, WB Saunders, 1986, pp 313384.)
Figure 24-65 Effect of administration of the luteinizing hormonereleasing hormone (LHRH) agonist deslorelin (4 µg/kg/day subcutaneously) on pulsatile secretion of LH (top), LH response to LHRH (middle), and plasma concentration of estradiol (bottom) in a 5 1/12-year-old girl with idiopathic true precocious puberty. This patient, who had a bone age of 13 years when treatment was begun, has been administered deslorelin for 7 years. During this period, the estimated predicted final height increased by 15 cm. Surprisingly, the bone age advanced by only about 6 months on serial examinations for several years. For conversion to SI units, see the legend of Figure 24-17. (Modified from Grumbach MM, Kaplan SL. Recent advances in the diagnosis and management of sexual precocity. Acta Paediatr Jpn 1988; 30[supp]:155175.)
Figure 24-66 Deslorelin treatment (4 µg/kg/day subcutaneously) of girls and boys with true precocious puberty: effect during the first 12 weeks of treatment on the luteinizing hormone (LH) and follicle-stimulating hormone (FSH) response to a challenge with LH-releasing hormone (LHRH) (mean peak response and maximum increment) and on the maximal unstimulated concentration of plasma estradiol in the girls and of plasma testosterone in the boys. Note the relatively rapid change from pubertal values to prepubertal values. For conversion to SI units, see the legends of Figure 24-17 and Figure 24-18. (From Styne DM, Harris DA, Egli CA, et al. Treatment of true precocious puberty with a potent luteinizing hormone releasing factor agonist: effect on growth, sexual maturation, pelvic sonography, and the hypothalamic pituitary gonadal axis. J Clin Endocrinol Metab 1985; 61:142181. © by The Endocrine Society.)
Figure 24-67 A 2 5/12-year-old girl with true precocious puberty after 6 weeks of deslorelin therapy (4 µg/day subcutaneously). Note the regression in the size of the breasts; however, the rapid rate of growth had not decreased. At the end of 1 year of therapy, growth rate was suppressed to 4 cm/year, and bone age advanced only 1 year. CA, chronologic age; HT, height; WT, weight; BA, bone age. (From Styne DM, Grumbach MM. Puberty in the male and female: Its physiology and disorders. In Yen SCC, Jaffe RB [eds]. Reproductive Endocrinology, 2nd ed. Philadelphia, WB Saunders, 1986, pp 313384.)
Figure 24-68 Effect of luteinizing hormonereleasing hormone (LHRH) agonist therapy in true precocious puberty on growth. Left, Changes in mean height velocity (cm/year ± 1 SE) after the initiation of LHRH agonist therapy with DTrp⁶ Pro⁹ Net (LHRH) (filled bars) or with nafarelin (hatched bars). A sharp decrease in height velocity occurred within 1 year. Right, Mean (±1 SE) height for bone age before and during LHRH agonist treatment. The discrepancy between height and the more advanced bone age decreases (reverts to normal) with chronic LHRH agonist treatment. (From Kaplan SL, Grumbach MM. True precocious puberty: treatment with GnRH agonists. In Delemarre-Van de Waal H, Plant TM, van Rees GP, et al [eds]. Control of the Onset of Puberty III. Amsterdam, Elsevier, 1989, pp 357-373.)
**Figure 24-69** A 1 5/12-year-old boy with a human chorionic gonadotropin (hCG)secreting hepatoblastoma. Note the outline of the large liver (left) and the penile enlargement (right). The testes were 2 × 1 cm, and public hair was stage 2. The plasma hCG level was 50 mlU/mL, plasma testosterone 168 ng/dL, and plasma -fetoprotein 160,000 ng/mL. Metastatic lesions in both lungs were seen on the radiograph of the chest. To convert testosterone values to SI units, see the legend of Figure 24-17. To convert hCG values to international units per liter, multiply by 1.0. To convert -fetoprotein values to micrograms per liter, multiply by 1.0. (From Kaplan SL, Grumbach MM. Pathogenesis of sexual precocity. In Grumbach MM, Sizonenko PC, Aubert ML [eds]. Control of the Onset of Puberty. Baltimore, Williams & Wilkins, 1990, pp 620660. © 1990, the Williams & Wilkins Co., Baltimore.)
Figure 24-70 Familial testotoxicosis. **Left,** A 5½-year-old boy and his 28-year-old father with the disorder. The boy exhibited signs of sexual precocity by 3 years of age. Height was 130.6 cm (+4.8 SD); bone age 12½ years. The plasma testosterone level was 267 ng/dL; dihydrotestosterone 46 ng/dL; dehydroepiandrostosterone sulfate (DHEAS) 23 µg/dL. The plasma luteinizing hormone (LH) and follicle-stimulating hormone (FSH) levels were low, and neither rose after treatment. Pulsatile LH secretion was not demonstrable. Treatment with deslorelin, an LHRH agonist, was without effect. The father had begun sexual maturation by 3 years of age and had reached a final height of 162.6 cm in his early teens. The plasma testosterone level was 294 ng/dL; LH 0.5 ng/mL (LER-960); and FSH 0.5 ng/mL (LER-869). The father had an adult-type LH and FSH response to LHRH; the LH level increased to 7.5 ng/mL, and the FSH level to 2 ng/mL. At least 28 male family members over nine generations are affected. To convert dihydrotestosterone values to nanomoles per liter, multiply by 0.03467. For other conversions to SI units, see the legends of Figure 24-17 and Figure 24-18. **Center,** External genitalia of the 5½-year-old boy. The penis measured 12 × 2.8 cm; the right testis was 4 × 2 cm, and the left testis 3.5 × 2.5 cm. **Right,** Testis of the boy showed Leydig cell maturation without Reinke crystalloids and spermatogenesis (Mallory trichome).
Figure 24-71. A. The serpentine seven transmembrane Gₛ protein coupled hLH/hCG receptor with its large extracellular domain and the intracellular domain. The seven helical transmembrane domains are indicated by Roman numerals. B. The two-dimensional seven-transmembrane topology of the hLH/hCG receptor with positions of constitutively activating mutations causing testotoxicosis (male-limited autosomal dominant sexual precocity). The mutations are indicated by solid circles and the residue number. Note the cluster of mutations in the VI transmembrane helix and third cytoplasmic loop. The aspartine 578→glycine mutation is the most common. (Redrawn from Yano K, Kohn LD, Saji M, et al. A case of male limited precocious puberty caused by a point mutation in the second transmembrane domain of the luteinizing hormone choriogonadotropin receptor gene. Biochem Biophys Res Commun 1996; 220:1036-1042.)
Figure 24-73 A 10/12-year-old girl with recurrent "autonomous" follicular cysts of the ovary. MPA, medroxyprogesterone acetate (oral). For conversion to SI units, see the legend of Figure 24-17. (From Kaplan SL, Grumbach MM. Pathogenesis of sexual precocity. In Grumbach MM, Sizonenko PC, Aubert ML [eds]. Control of the Onset of Puberty. Baltimore, Williams & Wilkins, 1990, pp 620-660. © 1990, the Williams & Wilkins Co., Baltimore.)
Figure 24-74  A 7 4/12-year-old girl with luteinizing hormone-releasing hormone (LHRH) independent sexual precocity associated with McCune-Albright syndrome. She had breast development since infancy, and it increased noticeably at about 3 years of age; 6 months later episodes of recurrent vaginal bleeding began. Growth of pubic hair was noted at about 4 to 5 years of age. At age 5 1/12 years the bone age was 6 11/12 years; height was +1 SD above the mean value for age. By 6½ years of age, when she was seen at the University of California, San Francisco, the bone age had advanced to 9 years, and height was at +1 SD. Breasts were at Tanner stage 4; pubic hair at stage 3. Extensive irregular café-au-lait macules cover the right side of the face, left lower abdomen and thigh, and both buttocks. A bone survey showed widespread involvement of the long bones with typical polyostotic fibrous dysplasia, and the floor of the anterior fossa of the skull was sclerotic and the diploetic space widened. She has had two pathologic fractures through bone cysts in the right upper femur. Note the osseous deformities. Plasma estradiol concentrations were consistently in the pubertal range; LH response to LHRH was prepubertal. Results of thyroid function studies were normal, including the thyrotropin response to thyrotropin-releasing hormone administration and antithyroid antibodies were not detected. Treatment with oral medroxyprogesterone acetate suppressed menses and arrested pubertal development but did not slow skeletal maturation. Her final height is 142 cm (-2.5 SD). Menstrual cycles are regular.
Figure 24-75a Bone lesions in McCune-Albright syndrome. A, The skull with severe thickening primarily at the base due to fibrous dysplasia. The auditory and optic nerves could be caught in narrowed foramina but that is not the case in these patients. B and C, distortions of the long bones, which can develop into a "shepherd's crook" appearance; note the multiple bone cysts.
Figure 24-75b D. Bone scan showing the areas of remodeling that "light up" depending upon the area affected in individual patients. There are examples of patients primarily affected in the craniofacial area, in the appendicular area, and in both areas as well as the axial skeleton.  (Courtesy of Michael T. Collins, M.D., National Institutes of Health, Bethesda, Maryland and Sandra Gorges, M.D., University of California, Davis.)
Figure 24-76 Serial pelvic ultrasonograms at 2-week intervals in a 6-year-old girl with McCune-Albright syndrome. Breast development and vaginal bleeding coincided with the enlargement of the ovarian cyst. With the spontaneous regression of the large ovarian cyst, the breasts regressed in size and vaginal bleeding ceased. (From Kaplan SL, Grumbach MM. Pathogenesis of sexual precocity. In Grumbach MM, Sizonenko PC, Aubert ML [eds]. Control of the Onset of Puberty. Baltimore, Williams & Wilkins, 1990, pp 620660. © 1990, the Williams & Wilkins Co., Baltimore.)
The G protein guanosine triphosphatase (GTPase) cycle. The heterotrimeric guanine nucleotide-binding proteins (G proteins) composed of three subunits (, , ) couple cell-surface receptors consisting of a single serpentine polypeptide having seven helical membrane-spanning domains with an effector, in this instance, adenylate cyclase (AC) that catalyzes the transformation of adenosine triphosphate (ATP) to cyclic adenosine monophosphate (cAMP). The G protein stimulation subunit , , mediates the stimulation of cAMP generation. In the inactive, unstimulated state, the G protein is a heterotrimer and GDP is tightly bound to the subunit. When the cell-surface receptor is activated by its cognate agonist, the receptor catalyzes the release of the tightly bound GDP, which enables GTP to bind to the subunit. The GTP-bound subunit (-GTP) dissociates from the tightly bound dimer, and both play a role in the G protein activation of the effector, adenylate cyclase. The intrinsic GTPase activity of the subunit ends the stimulation of the effector by converting the bound -GTP to -GDP; as a consequence, the subunit again returns to its inactive state and reassociates with high affinity with the subunit, yielding the , , heterotrimer. Disorders of signal transduction can arise from germ cell or somatic mutations at any of the five stages of the cycle. The gain-of-function, activating somatic mutations in the GNAS1 gene that encodes the G subunit and leads to McCune-Albright syndrome (shown in the bracket), involves the highly conserved arginine 201 residue. These mutations inhibit the intrinsic GTPase activity of the subunit and hence the conversion of the bound GTP to GDP. (See text.) Alanine 366 to serine mutation (shown in the bracket) was detected in two boys, both of whom had pseudohypothyroidism Ia (PHP) and testotoxicosis. The mutant protein was constitutively activated in the Leydig cells at the scrotal temperature (32 to 33°C), leading to testotoxicosis, but was rapidly degraded at body temperature, 37°C, which led to PHP1a. (See text.) (Modified from Spiegel AM. Mutations in G proteins and G protein-coupled receptors in endocrine disease. J Clin Endocrinol Metab 1996; 81:2432–2442.)
Figure 24-78 Left and center, Severe, chronic hypothyroidism of Hashimoto’s thyroiditis in a 7 1/12-year-old girl with sexual precocity (without pubic or axillary hair), episodic vaginal bleeding, and galactorrhea. She had symptoms of hypothyroidism and a sharply decreased rate of growth over the previous 2 years (height, -1 SD; bone age, 5 3/12 years). Breast development was Tanner stage 3; the labia minora were enlarged, and the vaginal mucosa was dull pink, thickened, and rugated with evidence of an estrogenic effect. No acne, seborrhea, or hirsutism was present. The uterus was of adolescent size, and the endometrial mucosa was in a proliferative phase. Urinary gonadotropins were barely detectable by bioassay. Right, Striking change in appearance after 8 months of thyroid hormone treatment. She had grown 7 cm in height and lost 8.1 kg in weight; the breasts had decreased in size, galactorrhea was no longer demonstrable, the labia minora had regressed, and the vaginal mucosa was pink and glistening (no estrogen effect). Ten weeks after the initiation of thyroid hormone replacement therapy, she developed a right slipped capital femoral epiphysis that was repaired surgically; recovery was uneventful.
Figure 24-79 Left. Radiograph of the skull of a patient with hypothyroidism illustrating an enlarged pituitary fossa in the lateral view. The dorsum sellae was thin and demineralized, and the floor had a double contour line. The area of the sella turcica was 150 mm$^2$. Pneumoencephalography showed a suprasellar mass impinging on the cisterna chiasmatica. After thyroid hormone treatment for 8 months, the area of the sella had decreased 30% in volume to 100 mm$^2$, the dorsum sellae had remineralized, and the double floor was no longer evident. Right. Growth curve illustrating the decrease in growth rate despite the sexual precocity and the catch-up growth induced by thyroid hormone therapy. (From Van Wyk JJ, Grumbach MM. Syndrome of precocious menstruation and galactorrhea in juvenile hypothyroidism: an example of hormonal overlap in pituitary feedback. J Pediatr 1960; 57:416435.)
Figure 24-80 The diagnosis of sexual precocity in girls.
Figure 24-81 The evaluation of pubic hair in normal phenotypic girls before 7 years.
Figure 24-82 The diagnosis of sexual precocity in a phenotypic male.
Figure 25-1 During aging, declines in the activities of a number of hormonal systems occur. PRL, prolactin; T<sub>4</sub>, thyroxine; TSH, thyrotropin. Left, A decrease in growth hormone (GH) release by the pituitary gland causes a decrease in the production of insulin-like growth factor I (IGF-I) by the liver and other organs (somatopause). Middle, A decrease in release of gonadotropin luteinizing hormone (LH) and follicle-stimulating hormone (FSH) and decreased secretion at the gonadal level (from the ovaries, decreased estradiol [E<sub>2</sub>] from the testicle, decreased testosterone [T]) cause menopause and andropause, respectively. (Immediately after the initiation of menopause, serum LH and FSH levels increase sharply.) Right, The adrenocortical cells responsible for the production of dehydroepiandrosterone (DHEA) decrease in activity (adrenopause) without clinically evident changes in corticotropin (adrenocorticotropic hormone, ACTH) and cortisol secretion. A central pacemaker in the hypothalamus or higher brain areas (or both) is hypothesized, which together with changes in the peripheral organs (the ovaries, testicles, and adrenal cortex) regulates the aging process of these endocrine axes.
**Figure 25-2** Changes in the hormone levels of normal women (left) and men (right) during the aging process. A and B, Estrogen secretion throughout an individual normal woman’s life (expressed as urinary estrogen excretion) (A) and mean free testosterone (T) index (the ratio of serum total T to sex hormone-binding globulin levels) during the life span of healthy men (B). (From Guyton AC. In Guyton AC [ed]. Textbook of Medical Physiology, 8th ed. Philadelphia, WB Saunders, 1991, pp 899914.) C and D, Serum dehydroepiandrosterone sulfate (DHEAS) concentrations in 114 healthy women (C) and 163 healthy men (D). (Adapted from Ravaglia G, et al. J Clin Endocrinol Metab 1996; 81:11731178.) E and F, The course of serum insulin-like growth factor I (IGF-I) concentrations in 131 healthy women (E) and 223 healthy men (F) during aging. Note the difference in the distribution of ages in the different panels. (Adapted from Corpas E, et al. Endocr Rev 1993; 14:2039.)
Figure 25-3 Effect of raloxifene administration (60 to 120 mg/day) on the cumulative incidence of breast cancer in 7705 postmenopausal women (mean age, 66.5 years) with osteoporosis. Statistical significance of the difference between the groups was $P < .001$. (From Cummings SR, Eckert S, Krueger KA, et al. The effect of raloxifene on risk of breast cancer in postmenopausal women: results from the MORE randomized trial. Multiple Outcomes of Raloxifene Evaluation. JAMA 1999; 281:21892197.)
Figure 25-4 AC. Mean (± standard error) change from baseline in fat mass, lean mass, and bone mineral density of the lumbar spine (L2 to L4) as determined by dual-energy x-ray absorptiometry in 106 men older than 65 years who were treated with either testosterone or placebo (54 men each). The decrease in fat mass (P < .005) and the increase in lean mass (P < .01) in the testosterone-treated subjects were significantly different from those in placebo-treated subjects at 36 months. Bone mineral density increased significantly in both groups. (A and B, from Snyder PJ, et al. J Clin Endocrinol Metab 1999; 84:19661972; C, from Snyder PJ, et al. J Clin Endocrinol Metab 1999; 84:26472653.)
Figure 25-5 Serum total testosterone levels in a group of healthy young (n = 58; aged 21 to 35 years) and older (n = 96; aged 60 to 80 years) men. Mean ± standard deviation. Serum total testosterone is 18.2 ± 4.2 nmol/L (525 ± 122 ng/dL) for the young men and 14.5 ± 4.5 nmol/L (420 ± 12.9 ng/dL) for the older men. (From Tenover JS. Androgen administration to aging men. Endocrinol Metab Clin North Am 1994; 23:877-892.)
Figure 25-6 Human steroidogenic enzymes in peripheral intracrine tissues. DHEA, dehydroepiandrosterone; DHEAS, DHEA sulfate; DHT, dihydrotestosterone; 5-DIOL, androsten-5-ene-3,17-diol; 4-Dione, androstenedione; E₁, estrone; E₂, estradiol; 3-HSD, 3-hydroxysteroid dehydrogenase; 17-HSD, 17-hydroxysteroid dehydrogenase; Testo, testosterone. (Modified and adapted from Labrie F, Luu-The V, Lin SX, et al. Intracrinology: role of the family of 17 beta-hydroxysteroid dehydrogenases in human physiology and disease. J Mol Endocrinol 2000; 25:116.)
Figure 25-7  Percentage of healthy adult men and women who reported an improved sense of well-being after 12 weeks of oral administration of 50 mg of DHEA nightly in comparison with placebo: *** P < .005 compared with placebo values. Scored values of libido on a Visual Analogue Scale in men and women did not change during DHEA administration. (From Morales AJ, Nolan JJ, Nelson JC, et al. Effects of replacement dose of dehydroepiandrosterone in men and women of advancing age. J Clin Endocrinol Metab 1994; 78:13601367.)
Figure 25-8 Knee extension-flexion muscle strength at baseline (100%) and percentage change in response to placebo and DHEA (100 mg/day) in aging men ($n = 8$) and women ($n = 8$), expressed as number of feet. * $P < .05$, placebo versus baseline; ** $P < .05$, DHEA versus placebo. (From Yen SS, Morales AJ, Khorram O. Replacement of DHEA in aging men and women. Potential remedial effects. Ann NY Acad Sci 1995; 774:128142.)
Figure 26-1 Distribution and function of calcium and phosphate. Note the dramatic differences between intracellular and extracellular concentrations of calcium ion and the dramatically different functions of calcium and phosphate inside cells.
Figure 26-2 Parathyroid hormone (PTH)-calcium feedback loop that controls calcium homeostasis. Four organs—the parathyroid glands, intestine, kidney, and bone—together determine the parameters of calcium homeostasis. +, positive effect; -, negative effect.
Figure 26-3 Sequences of pre-parathyroid hormone from six species. Completely conserved residues are in boldface. Arrows indicate the sites of signal sequence ("pre") and "pro" sequence cleavage. Numbers start at residue +1 of mature parathyroid hormone (PTH); because of gaps, the numbers correspond only to the mammalian and not to the chicken sequence. Amino acids are indicated by the single letter code: A, Ala; R, Arg; N, Asn; D, Asp; C, Cys; Q, Gln; E, Glu; G, Gly; H, His; I, Ile; L, Leu; K, Lys; M, Met; F, Phe; P, Pro; S, Ser; T, Thr; W, Trp; Y, Tyr; V, Val.
Figure 26-4 Intracellular processing of pre-proparathyroid hormone (pre-pro-PTH). Diagonal arrows indicate sites of cleavage by enzymes that generate pro-PTH in the rough endoplasmic reticulum (ER), PTH in the Golgi, and carboxy-terminal fragments of PTH in the secretory granule.
Figure 26-5 Parathyroid hormone (PTH) secretion. A, Secretory response of bovine parathyroid glands to induced alterations of plasma calcium concentration. Calves were infused with calcium chloride or ethylenediaminetetraacetic acid (EDTA), and PTH secretion was assessed by measuring PTH levels in the parathyroid venous effluent. The symbols and vertical bars indicate the secretory rate (mean ±SE) in calcium concentration ranges of 1.0 or 0.5 mg/100 mL. The number of calves and samples are indicated, respectively, by numbers below and above the bars. (From Mayer GP, Hurst JG. Sigmoidal relationship between parathyroid hormone secretion rate and plasma calcium concentration in calves. Endocrinology 1978; 10:10371042.) B, Sigmoidal curve generated by the equation \( Y = \frac{[A D]}{[1 + (X/C)^B]} + D \). Such a curve can be defined by four parameters: the maximal secretory rate (A), the slope of the curve at its mid-point (B), the level of calcium at the mid-point (often called the set-point) (C), and the minimal secretory rate (D); the significance of A, B, C, and D is described in the text. (Modified from Brown EM. Four-parameter model of the sigmoidal relationship between parathyroid hormone release and extracellular calcium concentration in normal and abnormal parathyroid tissue. J Clin Endocrinol Metab 1983; 56:572581.) C, Relationships between calcium and PTH levels when each in turn is treated as an independent variable. The dashed line represents the sigmoidal relationship between calcium and PTH, when calcium is the independent variable. This curve is the same as that in A and B, but it is turned on its side, because the axes are reversed. The solid line represents the relationship between calcium and PTH when PTH is considered the independent variable; values for this curve result from measurements made during PTH infusion into parathyroidectomized animals. Actual data are limited; thus, the curves should be viewed as illustrative. (From Parfitt AM. Calcium homeostasis. In Mundy GR, Martin TJ [eds]. Physiology and Pharmacology of Bone. Berlin, Springer-Verlag, 1993.)
Figure 26-6 Structural model of parathyroid cell calcium-sensing receptor predicted from its amino acid sequence. The large amino-terminal domain is extracellular. Conserved residues among the metabotropic glutamate receptors and the bovine parathyroid calcium-sensing receptor are indicated by the symbols noted in the box. Also indicated are potential glycosylation and protein kinase C phosphorylation sites. SP, signal peptide; HS, hydrophobic segment. Missense and chain-terminating mutations identified in patients with either familial hypocalciuric hypercalcemia or familial hypoparathyroidism due to activating mutations of the calcium sensing receptor are indicated as well. (From Brown EM, Bai M, Pollak M. Familial benign hypocalciuric hypercalcemia and other syndromes of altered responsiveness to extracellular calcium. In Krane SM, Avioli LV [eds]. Metabolic Bone Diseases, 3rd ed. San Diego, Academic Press, 1998.)
Figure 26-7 Effects of parathyroid hormone (PTH) on distal tubular calcium transport. PTH acts to increase chloride efflux through channels in the basolateral membrane. As indicated, this increase leads to calcium influx through apical calcium channels. PTH also increases basolateral Ca\(^{2+}\)-Na\(^{+}\) exchange. The apical Na\(^{+}\)-Cl\(^{-}\) cotransporter allows chloride to enter the cell and is the target of thiazide diuretics. Cl\(^{-}\), intracellular chloride. (Adapted from Friedman PA, Gesek FA. Calcium transport in renal epithelial cells. Am J Physiol 1993; 264:F181-F198.)
Figure 26-8 Nomogram for determining renal threshold phosphate concentration ($Tm_{\text{PO}_4}/\text{GFR}$) from the plasma phosphate concentration and the fractional reabsorption of filtered phosphate (TRP) or fractional excretion of filtered phosphate (1TRP, or $C_{\text{PO}_4}/C_{\text{creat}}$). Because the blood level of phosphate influences the renal handling of phosphate, the renal threshold phosphate concentration best separates normal from abnormal renal phosphate handling. C, clearance; creat, creatinine; GFR, glomerular filtration rate; TRP, tubular resorption of phosphate. (From Walton RJ, Bijvoet QLM. Nomogram of derivation of renal threshold phosphate concentration. Lancet 1975; 2:309310.)
Figure 26-9 Osteoblast lineage. All precursors of osteoblasts can proliferate; osteoblasts are transformed to osteocytes and lining cells without further proliferation. Some data suggest that lining cells may revert to osteoblast function after parathyroid hormone stimulation. At each stage in the lineage, apoptotic cell death is probably an alternative fate.
Figure 26-10 Stromal cell control of osteoclastogenesis and osteoclast activity. Parathyroid hormone (PTH) acts on PTH/PTH-related protein (PTHrP) receptors on precursors of osteoblasts to increase the production of macrophage colony-stimulating factor (M-CSF) and RANK ligand and to decrease the production of osteoprotegerin (OPG). M-CSF and RANK ligand stimulate the production of osteoclasts and increase the activity of mature osteoclasts by binding to the receptor RANK. OPG blocks the interaction of RANK ligand and RANK.
Figure 26-11 Parathyroid hormone (PTH)/PTH-related protein (PTHrP) receptors act as nucleotide exchangers. PTH binding to the receptor leads to exchange of guanosine triphosphate (GTP) for guanosine diphosphate (GDP) bound to G subunits. G subunits bound to GTP are released from the receptor and from the subunits and then activate effectors. Gs activates adenylate cyclase, leading to the formation of cyclic AMP (cAMP), which then activates protein kinase A (PKA). Gq and related subunits activate phospholipase C (PLC). PLC hydrolyzes phosphatidylinositol (1,4,5)tris-phosphate to generate diacyl glycerol (DAG) and inositol (1,4,5)tris-phosphate (IP3). The DAG then activates protein kinase C (PKC), and the IP3 activates a receptor on microsomal vesicles that directs the movement of calcium from microsomal vesicles into the cytosol.
Figure 26-12  Network of parathyroid hormone (PTH) ligands and receptors (R). PTH and PTH-related protein (PTHrP) closely resemble each other at the amino-terminal region; TIP39 is more distantly related. Although only the PTH/PTHrP receptor and the PTH2 receptors have been cloned, biologic actions suggest receptors specific for the carboxy-terminal portion of PTH, as well as distinct receptors for the mid-region of PTHrP and for a more distal region of PTHrP. Not shown are possible nuclear sites of action of PTHrP.
Figure 26-13 Interactions between parathyroid hormone (PTH) and the PTH/PTH-related protein (PTHrP) receptor. Key residues involved in PTH/PTH receptor function are indicated (amino acid residues and position numbers correspond to the human PTH/PTHrP receptor sequence): circles, residues identified in crosslinking studies; octagon, a site involved in Blomstrand's chondrodysplasia; ovals, hydrophobic residues important for PTH(134) and PTH(114) binding; triangles, residues mutated in patients with Jansen's chondrodysplasia; hexagons, sites at which the corresponding residues in the PTH2 receptor play a role in discriminating between PTH and PTHrP; squares, residues that determine agonist versus antagonist action of Arg2-PTH(134); diamonds, sites at which mutations impair PTH(134) binding but not PTH(334) binding; rectangles, residues that when mutated alter G protein coupling; and dashed curves with arrows, interdomain interactions as determined by paired mutations affecting PTH(134) interaction (R233 and Q451) or by zinc chelation studies (H307 and R408). (From DeGroot LJ (ed). Endocrinology, 4th ed. Philadelphia, WB Saunders, 2001, p 986.)
Figure 26-14  Sequences of parathyroid hormone-related protein (PTHrP) from five species. Completely conserved residues are in **boldface**; note the high level of conservation through residue 111. Arrows indicate sites of internal cleavage after residues 37 and 95, which lead to generation of PTHrP(3894) amide and PTHrP(3895). Another site of cleavage, generating PTHrP(38101) and, perhaps, PTHrP(107139) is not shown. The three human sequences represent proteins synthesized from alternatively spliced mRNAs and differ only after residue 139. Amino acids are indicated by the single letter code; see legend to Figure 26-3 for code.
Figure 26-15 The amino acid sequences of calcitonin, calcitonin generelated peptide (CGRP), and amylin and adrenomedullin (ADM) from selected species. The bold Cs represent the cysteine residues that form the disulfide linkages critical for the secondary structure of these peptides. The other residues conserved among species are indicated by a dashed line. See the legend to Figure 26-3 for the single-letter amino acid codes.
Figure 26-16 Tissue-specific expression of the calcitonin gene. Splicing of alternative exons leads to two different messenger RNAs (mRNAs). The mRNA encoding calcitonin is found predominantly in the thyroid C cell; the mRNA encoding calcitonin gene related-peptide (CGRP) is found predominantly in the hypothalamus and other nervous tissue. (From Amara SG, Jonas V, Rosenfeld MG, et al. Alternative RNA processing in calcitonin gene expression generates mRNAs encoding different polypeptide products. Nature 1982; 298:240-244.)
Figure 26-17  Vitamin D precursors and alternative reaction products. The numbering system for vitamin D carbons and the distinct structures of vitamin D$_2$ (ergocalciferol) and D$_3$ (cholecalciferol) are noted, as is the structure of dihydrotachysterol, a synthetic product not produced in vivo. Note that the 3-hydroxyl group of dihydrotachysterol is in a pseudo 1-hydroxyl configuration. This may explain the relatively high potency of dihydrotachysterol in conditions associated with low 1-hydroxylase activity.
Figure 26-18 Relative potency of analogues of 1,25(OH)₂D₃ (1,25-dihydroxyvitamin D₃) in competitive binding to vitamin D receptors of chick intestinal mucosa. Slopes, are plotted for (left to right): 1,25(OH)₂-D₃, 1,25-dihydroxyvitamin D₃, 3 deoxy-1,25(OH)₂D₃, 3 deoxy-1,25-dihydroxyvitamin D₃, 25-OH-DHT, 25-hydroxydihydrotachysterol, 25-OH-5,6-trans-D₃, 25-hydroxy-5,6 transvitamin D₃, 25-OH-D₃, 25-hydroxyvitamin D₃, 1--OH-D₃, 1--hydroxyvitamin D₃; 24,25-OH₂-D₃, 24,25-dihydroxyvitamin D₃; 3-deoxy-1-OH-D₃, 3-deoxy-1-hydroxyvitamin D₃; D₃, vitamin D₃; DHT, dihydrotachysterol. (From Proscai DA, Okamura WH, Norman AW. Structural requirements for the interaction of 1,25-(OH)₂vitamin D₃ with its chick intestinal system. J Biol Chem 1975; 250:8382-8388.)
Figure 26-19 Transcriptional activation by 1,25-dihydroxyvitamin D₃ (1,25(OH)₂ D₃). A heterodimer of retinoid X receptor (RXR) and vitamin D receptor (VDR) binds to a pair of hexameric sequences separated by three intervening bases (ATG). Arrows indicate that the hexamers found in the up-regulated rat osteocalcin gene are variants of a consensus sequence, repeated here with identical orientations (direct repeats). Upon binding to DNA, the RXR-VDR heterodimer facilitates formation of a transcription initiation complex, which binds to DNA at and near the TATA sequence.
Figure 26-20 Homeostatic responses to variations in dietary calcium content. Major homeostatic responses to dietary calcium deprivation or loading are depicted. Arrow thickness indicates relative activity of transport or secretory mechanisms, whereas amounts of hormones or transported ions are related to the size of their notations. Parentheses indicate an inhibitory regulation. Note that the extracellular calcium concentration is well maintained, although different underlying mechanisms are involved in the two circumstances (see text for details).
Figure 26-21 Intact immunoreactive parathyroid hormone (PTH) determined using a two-site immunoradiometric assay in normal and three different patient groups. Note some overlap between normal people and patients with primary hyperparathyroidism, but no overlap between hypercalcemic patients with primary hyperparathyroidism and those with hypercalcemia of malignancy. (From Segre GV. Advances in techniques for measurement of parathyroid hormone: current applications in clinical medicine and directions for future research. Trends Endocrinol Metab 1990; 1:243-247.)
Figure 26-22 Plasma PTHrP(174) determined by two-site immunoradiometric assay in selected patient groups and normals. Also shown are concentrations of PTHrP in human milk (filled circles) and in bovine milk (open circles). Two normocalcemic patients with cancer (filled triangles) subsequently became hypercalcemic. Hatched area denotes levels too low to detect with this assay. PTHrP, parathyroid hormone-related protein. (Adapted from Burtis WJ, Brady TG, Orloff JJ, et al. Immunochemical characterization of circulating parathyroid hormone-related protein in patients with humoral hypercalcemia of cancer. N Engl J Med 1990; 322:11061112.)
Figure 26-23 Radiograph of hand from a patient with severe primary hyperparathyroidism. Note the dramatic remodeling associated with the intense region of high bone turnover in the third metacarpal in addition to widespread evidence of subperiosteal, endosteal, and trabecular resorption.  (Courtesy of Fuller Albright Collection, Massachusetts General Hospital.)
Figure 26-24 Iliac crest biopsy specimens from a patient with primary hyperparathyroidism (left) and a normal control (right), viewed by scanning electron microscopy. Note the thin cortices and contrasting maintenance of trabecular bone in the patient. (From Parisien M, Silverberg SJ, Shane E, et al. The histomorphometry of bone in primary hyperparathyroidism: preservation of cancellous bone structure. J Clin Endocrinol Metab 1990; 70:330938.)
Figure 26-25 Abnormal patterns of parathyroid hormone (PTH) secretion from cells prepared from adenomatous glands and stimulated with varying levels of calcium in tissue culture. The shaded area shows the pattern of PTH release (±1 SD) from normal human parathyroid cells. Panel A illustrates the pattern from four patients with little suppression of PTH secretion by calcium. Panel B illustrates the pattern from four patients with relatively intact mechanism of suppression of PTH secretion by calcium. Even in this group the set-point for calcium suppression is shifted to the right. (From Brown, EM. Calcium-regulated parathyroid hormone release in primary hyperparathyroidism, studies in vitro with dispersed parathyroid cells. Am J Med 1979; 66:923931.)
**Figure 26-26** Sites of ectopic location of 104 parathyroid glands found at reoperation for primary hyperparathyroidism.  *(From Wang C-A. A clinical and pathological study of 112 cases. Ann Surg 1977; 186:140145.)*
Figure 26-27  Technetium Tc 99m sestamibi $^{123}$I subtraction scanning of a patient with persistent hyperparathyroidism after two previous unsuccessful operations. Arrow points to parathyroid adenoma, shown as increased tracer uptake in the aortopulmonary window. (From Thule P, Thakore K, Vansant J, et al. Preoperative localization of parathyroid tissue with technetium-99m sestamibi $^{123}$I subtraction scanning. J Clin Endocrinol Metab 1994; 78:7782.)
Figure 26-28 Index of urinary excretion rate for calcium as a function of creatinine clearance. Each point represents the mean of multiple determinations for a hypercalcemic patient with familial hypocalcic hypercalcemia (filled circles) or with typical primary hyperparathyroidism (open circles). The data are based on average 24-hour urinary excretion values and average fasting serum samples. (From Marx SJ, Attie MF, Levine M, et al. The hypocalcic or benign variant of familial hypercalcemia: clinical and biochemical features in fifteen kindreds. Medicine 1981; 60:397-412.)
Figure 26-29 A patient with Jansen's metaphyseal chondrodysplasia at ages 5 years and 22 years. Note the short stature, characteristic facies, and misshapen metaphyseal region of long bones. (From Frame B, Poznanski AK. Conditions that may be confused with rickets. In DeLuca HF and Anastas CS (eds). Pediatric Diseases Related to Calcium. New York, Elsevier, 1980, pp 269289.)
Figure 26-30 Approach to the management of the hypercalcemic patient. BUN, blood urea nitrogen; CT, computed tomography; IEP, immunoelectrophoresis; PTH, parathyroid hormone.
Figure 26-31 Approach to the management of the hypercalcemic patient with parathyroid hormone-dependent hypercalcemia. Cl, clearance; FHH, familial hypocalciuric hypercalcemia; Li, lithium; PTH, parathyroid hormone; Fam. Hx., family history.
Figure 26-32 Trouseau's sign. (From Burnside JW, McGlynn TJ. Physical Diagnosis, 17th ed. Baltimore, Williams & Wilkins, 1987, p 63.)
Figure 26-33 Daughter (left) and mother (right) with pseudohypoparathyroidism and Albright's hereditary osteodystrophy.
**Figure 26-34** Radiograph of hand from a patient with pseudohypoparathyroidism and Albright's hereditary osteodystrophy. Note the shortened fourth metacarpal.
**Figure 27-2** Synthesis and assembly of collagen fibrils. A, Intracellular post-translational modifications of pro alpha chains, association of propeptide domains, and folding into triple-helical conformation. Gal, galactose; Glc, glucose; Glc Nac, N-acetylglucosamine; (Man)n, mannose. B, Enzymatic cleavage of procollagen to collagen, self-assembly of collagen monomers into fibrils, and cross-linking of fibrils. (Modified from Prockop DJ, Kivirikko K. Heritable diseases of collagen. N Engl J Med 1984; 311:376386.)
Figure 27-3 The staggered arrangement of individual molecules in collagen fibers results in hole zones between the head of one molecule and the tail of the next. Mineral deposition (bottom) begins within the hole zones. (From Glimcher MJ, Krane SM. Treatise on Collagen 2, part B. New York, Academic Press, 1968, pp 672-5.)
Figure 27-4 Electron micrograph of rat calvarial bone showing mature osteoblasts with their dense, rough endoplasmic reticulum and large Golgi apparatus (a); an osteocyte embedded in the bone (b); and a less differentiated cell that may represent a preosteoblast (c). (Courtesy of Dr. Marijke E. Holtrop.)
**Figure 27-5** A. Cross-section of an osteon. B. Cultured cells from avian bone, showing osteocytes and their cytoplasmic connections. (From Aarden EM, Burger EH, Nijweide PJ. Function of osteocytes and bone. J Cell Biochem 1994; 55:287299. Reprinted by permission of Wiley-Liss, Inc., a subsidiary of John Wiley & Sons, Inc.)
Figure 27-6 Osteoclast formation. Osteoclasts form from osteoclast precursor cells, which are derived from hematopoietic lineage cells. These express C-FMS (the receptor for M-CSF) and RANK and attach to stromal/osteoblastic cells, which express M-CSF (both membrane-bound and soluble), membrane-bound RANKL, and OPG under the influence of stimulators of resorption (PTH, 1,25-Vit D, IL-1, TNF, IL-6, IL-11, or prostaglandins [PG]). If stromal or osteoblastic cells produce more RANKL than OPG, osteoclasts are formed and activated, which increases bone resorption. If stromal or osteoblastic cells produce more OPG than RANKL, OPG binds the available RANKL and new osteoclast formation is prevented. During states of inflammation, T lymphocytes are activated and produce both membrane-bound and soluble RANKL, which can, in turn, stimulate osteoclast-mediated bone resorption. It has also been shown that IL-1 and TNF can augment the effects of RANKL and M-CSF on osteoclast formation and bone resorption by directly stimulating osteoclast precursor cells and mature osteoclasts. IL, interleukin; M-CSF, macrophage colony-stimulating factor; OPG, osteoprotegerin; PTH, parathyroid hormone; RANK, receptor activator of nuclear factor B; RANKL, RANK ligand; TNF, tumor necrosis factor; 1,25 Vit D, 1,25-dihydroxyvitamin D.
Figure 27-7 Functional elements of the fully differentiated osteoclast. (From Suda T, Takahashi N, Martin T.J. Modulation of osteoclast differentiation. Endocr Rev 1992; 13:6680. Copyright © 1992, by The Endocrine Society.)
Figure 27-8 Stages of bone remodeling. The resorptive, reversal, and formative phases of bone remodeling and a completed bone structural unit (BSU) on a trabecular surface are illustrated. The morphologic features of the activation step have not been defined. (Courtesy of Dr. Robert E. Schenk, University of Berne, Switzerland.)
Figure 27-9 Three-dimensional reconstruction of the remodeling sequence in human trabecular bone. 1, Early bone resorption with osteoclasts (OCL); 2, late bone resorption with mononuclear cells (MON); 3, reversal phase with preosteoblasts (POB); 4, early matrix formation by osteoblasts (OB); 5, late bone formation with mineralization; 6, completed remodeling cycle with reversion to lining cells. (From Eriksen EF. Normal and pathological remodeling of human trabecular bone: three-dimensional reconstruction of the remodeling sequence in normals and in metabolic bone disease. Endocr Rev 1986; 7:379408. Copyright © 1986, by The Endocrine Society.)
Figure 27-10 Use of dual-energy x-ray absorptiometry for vertebral body morphometry. Posterior vertebral body heights and the ratio of anterior to posterior (A/P) height are presented in terms of standard deviation scores. Note that minor anterior wedging alone may not indicate an osteoporotic fracture. (Courtesy of Dr. Richard B. Mazess.)
Collagen cross-links. Cross-links are formed between the COOH-terminal and NH₂-terminal nonhelical portions of collagen and adjacent helical molecules. Immunoassays are available for the pyridinoline and deoxypyridinoline molecules themselves and for the adjacent nonhelical peptides. (Redrawn from Eyre DR. The specificity of collagen cross-links as markers of bone and connective tissue degradation. Acta Orthop Scand 1995; 66:16170.)
Figure 27-12  Tetracycline labels sites of active mineralization and is deposited at the calcification front (Cf) (top). A double-label technique can be used to measure the rate of mineralization; label A was administered about 10 days before label B (bottom). Undecalcified iliac crest, ultraviolet light, ×113. (From Aaron J. Histology and microanatomy of bone. In Nordin BEC [ed]. Calcium, Phosphate and Magnesium Metabolism. Edinburgh, Churchill Livingstone, 1976, pp 298356.)
Figure 27-13 Estimation of the current prevalence of osteoporosis in the United States. On the basis of World Health Organization criteria, more than 9 million women in the United States have osteoporosis; more than half of these women have established osteoporosis with fractures. In addition, 17 million postmenopausal women have osteopenia (low bone mass) and are at risk for development of osteoporosis. (From Melton LJ. How many women have osteoporosis now? J Bone Miner Res 1995; 10:175177.)
Figure 27-14 Age-specific incidence rates for hip, vertebral, and Colles’ fractures in Rochester, Minnesota. (From Cooper C, Melton L.J. Epidemiology of osteoporosis. Trends Endocrinol Metab 1992; 3:224. Copyright © 1992, by Elsevier Science Inc.)
Figure 27-15 Cross-section of an osteoporotic vertebra showing extensive loss of trabecular bone architecture. (Courtesy of Dr. Anders Odgaard, Århus, Denmark.)
Figure 27-16  Types of vertebral compression fractures. Changes in vertebral height can be quantitated by measuring percent change or standard deviations from expected normal heights. (From Genant HK, Wu CY, van Kuijk C, Nevitt MC. Vertebral fracture assessment using a semiquantitative technique. J Bone Miner Res 1993; 8:1137-1148.)
Figure 27-17 Diagnosis and management of osteoporosis. The diagram outlines an approach based largely on evidence from studies of postmenopausal white women, with dual-energy x-ray absorptiometry used to measure bone mineral density (BMD). Its application to other populations, including patients with secondary osteoporosis and other methods of assessing BMD, is not established.
Figure 27-18 Rickets. Left, Active rickets in a patient with tissue resistance to 1,25-dihydroxyvitamin D at age 21 months with genu varum, irregular metaphyses, and widened growth plates. Right, Inactive rickets in the same patient at age 27 months after treatment with massive doses of ergocalciferol. (From Marx SJ, Spiegel AM, Brown EM, et al. Familial syndrome of decrease in sensitivity to 1,25-hydroxyvitamin D. J Clin Endocrinol Metab 1978; 47:13031310. Copyright © 1978, by The Endocrine Society.)
Figure 27-19 Active osteomalacia in a patient (a sibling of the patient in Fig. 27-18) with hereditary tissue resistance to 1,25-dihydroxyvitamin D at age 18 with pseudo-fracture of the left tibia. (From Marx SJ, Spiegel AM, Brown EM, et al: Familial syndrome of decrease in sensitivity to 1,25-hydroxyvitamin D. J Clin Endocrinol Metab 1978; 47:13031310. Copyright © 1978, by The Endocrine Society.)
Figure 27-20 Electron micrograph of an osteoclast nucleus from a patient with Paget's disease showing characteristic intranuclear inclusion (consisting of microfilaments 125 nm in diameter). Decalcified bone, ×32,400. (Courtesy of Dr. Barbara G. Milts and Dr. Frederick R. Singer.)
Figure 27-21 Paget's disease of the tibia. Note the bowing, marked irregularity of the anterior cortex and the flame-shaped lytic lesion of the posterior cortex. (Courtesy of Dr. Ethel S. Siris.)
Figure 27-22 Radiograph of the skull of a patient with advanced Paget's disease showing thickening, disordered new bone formation (cotton-wool patches), and basilar impression. (From Singer FR. Paget's Disease of Bone. New York, Plenum, 1977.)
Figure 27-23 Radiographs of the lower limb of a patient with osteopetrosis at age 2 months (A) before bone marrow transplantation and at age 9 months after transplantation (B) showing formation of normal medullary bone. (From Ballet JJ, Griscelli C. Lymphoid cell transplantation in human osteopetrosis. In Horton JE, Tarpley TM, Davis WP [eds]: Mechanisms of Localized Bone Loss. London, IRI, 1978, pp 399-414, by permission of Oxford University Press.)
Figure 28-1 Genetic hypercalciuric stone-forming (GHS) rats. (From Bushinsky DA. Genetic hypercalciuric stone forming rats. Curr Opin Nephrol Hypertens 1999; 8:479488.)
Figure 28-2 Evaluation of stone formers.
Figure 28-3 Treatment of the patient with recurrent calcium oxalate stones. phos, phosphorus; RTA, renal tubular acidosis.
Figure 29-1 American Diabetes Association consensus.
Figure 29-4 Relationship between body mass index (A) or intra-abdominal fat (B) and insulin sensitivity. (A, From Fujimoto WY, Bergstrom RW, Boyko EJ, et al. Obesity Res 1995; Suppl 2:17951863; B, from Kahn SE, Pigeon RL, McCulloch DK, et al. Quantification of the relationship between insulin sensitivity and beta-cell function in human subjects: evidence for a hyperbolic function. Diabetes 1993; 42:16631672.)
Figure 29-5  Tissue uptake of glucose in nondiabetic and insulin-resistant diabetic subjects during a hyperinsulinemic-euglycemic clamp. (From DeFronzo RA. Lilly lecture 1987. The triumvirate: beta-cell, muscle, liver. A collusion responsible for NIDDM. Diabetes 1988; 37:667-687.)
Figure 29-6  Simplified schematic diagram demonstrating fatty acid (FA) uptake, activation (formation of FAcoenzyme A [CoA]), and intracellular transport to different organelles within a muscle cell. IMTG, intramuscular triglyceride; PG, prostaglandin; PL, phospholipid; SL, sphingolipid.
Figure 29-7  Glucose effect on triglyceride metabolism. Increased uptake of glucose results in an increase in the production of acetyl coenzyme A (CoA) as a product of glycolysis. The increased tricarboxylic acid (TCA) cycle activity associated with oxidation of both triglycerides and glucose increases the production of citrate, which is shuttled to the cytoplasm, activates the enzyme acetyl-CoA carboxylase (ACC) by allosteric mechanisms, and increases the susceptibility of ACC to phosphatases. This leads to increased ACC activity converting acetyl CoA to malonyl CoA. Malonyl CoA is a potent inhibitor of carnitine palmitoyltransferase I (CPT-I) on the outer mitochondrial membrane, which leads to accumulation of fatty acyl CoAs in the cytoplasm. This can result in the production of signaling molecules that can increase the activity of kinases and other enzymes and lead to insulin resistance.
Figure 29-9 Insulin resistance and dyslipidemia. The suppression of lipoprotein lipase and very-low-density lipoprotein (VLDL) production by insulin is defective in insulin resistance, leading to increased free fatty acid (FFA) flux to the liver, and increased VLDL production, which results in increased circulating triglyceride concentrations. The triglycerides are transferred to low-density lipoprotein (LDL) and high-density lipoprotein (HDL) and the VLDL particle gains cholesterol esters by the action of the cholesterol ester transfer protein (CETP). This leads to increased catabolism of HDL particles by the liver and loss of apolipoprotein A (ApoA), resulting in low HDL concentrations. The triglyceride-rich LDL particle is stripped of the triglycerides, resulting in the accumulation of atherogenic small, dense LDL particles.
Figure 29-10 Insulin suppresses hepatic glucose production by direct and indirect mechanisms. In insulin resistance, insulin's ability to suppress lipolysis in adipose tissue and glucagon secretion by alpha cells in the islet results in increased gluconeogenesis. In addition, insulin inhibition of glycogenolysis is impaired. Thus, both hepatic and peripheral insulin resistance results in abnormal glucose production by the liver.
Figure 29-11 Relationship between fasting hepatic glucose output and fasting plasma glucose levels. Open squares represent nondiabetic control subjects; closed squares represent diabetic subjects. (From Maggs DG, Buchanan TA, Burant CF, et al. Metabolic effects of troglitazone monotherapy in type 2 diabetes mellitus: a randomized, double-blind, placebo-controlled trial. Ann Intern Med 1998; 128:176185.)
Figure 29-12 Mean 24-hour profiles of plasma concentrations of glucose, C peptide, and insulin in normal and obese subjects. (From Polonsky KS, Given BD, van Cauter E. Twenty-four-hour profiles and pulsatile patterns of insulin secretion in normal and obese subjects. J Clin Invest 1988; 81:442448.)
Figure 29-13 Mean 24-hour profiles of insulin secretion rates in normal and obese subjects (top). The hatched areas represent ± 1 standard error of the mean. The curves in the lower panel were derived by dividing the insulin secretion rate measured in each subject by the basal secretion rate derived in the same subject. Mean data for the normal (dashed line) and obese (solid line) subjects are shown. (From Polonsky KS, Given BD, van Cauter E. Twenty-four-hour profiles and pulsatile patterns of insulin secretion in normal and obese subjects. J Clin Invest 1988; 81:442448.)
Figure 29-14 Patterns of insulin secretion in normal and obese subjects. Four representative 24-hour profiles from two normal-weight subjects (left) and two obese subjects (right). Meals were consumed at 0900, 1300, and 1800 hours. Statistically significant pulses of secretion are shown by the arrows. (From Polonsky KS, Given BD, van Cauter E. Twenty-four-hour profiles and pulsatile patterns of insulin secretion in normal and obese subjects. J Clin Invest 1988; 81:442448.)
Figure 29-15 Plasma insulin concentrations (A) and insulin secretion rates (B) in response to molar increments in the plasma glucose concentration during the graded glucose infusion in the insulin-resistant (dotted line) and insulin-sensitive (solid line) groups. (From Jones CNO, Pei D, Staris P, et al: Alterations in the glucose-stimulated insulin secretory dose-response curve and in insulin clearance in nondiabetic insulin-resistant individuals. J Clin Endocrinol Metab 1997; 82:1834-1838.)
Figure 29-16 Dose-response relationship between glucose and insulin secretory rate (ISR) after an overnight fast in control subjects (CON), normoglycemic subjects with a family history of noninsulin dependent diabetes mellitus (FDR), subjects with a nondiagnostic OGTT (NDX), subjects with impaired glucose tolerance (IGT), and subjects with noninsulin-dependent diabetes mellitus (NIDDM). BME, body mass index. (From Byrne MM, Sturis J, Sobel RJ, Polonsky KS. Elevated plasma glucose 2 h postchallenge predicts defects in beta-cell function. Am J Physiol 1996; 270:E572-E579. The American Physiological Society, copyright 1996.)
Figure 29-17 Oscillatory glucose infusions were administered with a periodicity of 144 minutes in representative subjects with type 2 diabetes, impaired glucose tolerance (IGT), and normal glucose tolerance. In the control subject, the insulin secretion rate (ISR) adjusts and responds to the 144-minute oscillations in glucose, resulting in sharp spectral peak at 144 minutes. In the subjects with IGT and type 2 diabetes, the ISR does not respond to the oscillatory glucose stimulus and although oscillations in insulin secretion are evident, they are irregular, resulting in markedly reduced spectral peaks at 144 minutes and small-amplitude high-frequency spectral peaks. (Adapted from O’Meara NM, Sturis J, Van Cauter E, Polonsky KS. Lack of control by glucose of ultradian insulin secretory oscillations in impaired glucose tolerance and in noninsulin-dependent diabetes mellitus. J Clin Invest 1993; 92:262271.)
**Figure 29-18** Mean (± standard error of the mean [SEM]) rates of insulin secretion in type 2 diabetic patients compared with control subjects. The *shaded area* corresponds to 1 SEM above and below the mean in control subjects. The curves in the lower panel were derived by dividing, for each subject, the insulin secretion rate at each sampling time by the average fasting secretion rate measured between 6 AM and 9 AM in the same subject.
Figure 29-19 Temporal variations in postbreakfast, postlunch, and postdinner rates of insulin secretion in control and diabetic subjects. In each subject, the secretion rates during the 30 minutes before the meal and the 4 hours after breakfast or the 5 hours after lunch or dinner were expressed as a percentage of the mean rate of insulin secretion during that interval. The curves were obtained by concatenating the resulting postmeal profiles in eight representative subjects. The times at which the meals were served to successive subjects in the series are indicated by arrows. (From Polonsky KS, Given BD, Hirsch LJ, et al. Abnormal patterns of insulin secretion in noninsulin-dependent diabetes mellitus. N Engl J Med 1988; 318:1231-1239.)
**Figure 29-20** Treatment algorithm for type 2 diabetes. FPG, fasting plasma glucose; PPG, postprandial plasma glucose.
Figure 30-1 Progression to diabetes of initially discordant monozygotic twins of patients with type 1 diabetes subdivided by the age of diabetes onset of the first twin to develop diabetes (proband). Late progression to diabetes is evident, with some twins becoming diabetic more than 20 years after their twin mate. For discordant twins whose twin mate developed diabetes after age 25, the risk of diabetes is less than 10%. (From Redondo MJ, Yu L, Hawa M, et al. Heterogeneity of type 1 diabetes: analysis of monozygotic twins in Great Britain and the United States. Diabetologia 2001; 44:354-362.)
Figure 30-2 Genes within the human leukocyte antigen (HLA) region (major histocompatibility complex) with HLA class I, class II, and class III regions illustrated. Each class II molecule is made up of two chains. DRB alleles are polymorphic; DRA does not vary. DQA and DQB molecules are both polymorphic. DPA and DPB are both also polymorphic. The class III region contains important genes such as complement components as well as the tumor necrosis factor gene. The class I region includes MIC-A and MIC-B genes as well as the classical histocompatibility HLA genes, A, B, and C.
Figure 30-4 Progression to type 1 diabetes of first-degree relatives of patients with diabetes subdivided by the number of anti-islet autoantibodies of insulin, GAD65 (glutamic acid decarboxylase), and ICA512 (IA-2) expressed. (From Verge CF, Gianani R, Kawasaki E, et al. Prediction of type 1 diabetes in first-degree relatives using a combination of insulin, GAD, and ICA512 bdc/IA-2 autoantibodies. Diabetes 45:926933, 1996.)
**Figure 31-1** Relative risks for the development of diabetic complications at different levels of mean hemoglobin A₁c (HbA₁c, glycated hemoglobin), obtained from the Diabetes Control and Complications Trial. *(Adapted from Skyler J: Diabetic complications: the importance of glucose control. Endocrinoel Metab Clin North Am 1996; 25:243254.)*
Figure 31-2 Lack of down-regulation of glucose transport in cells affected by diabetic complications. **Upper**, 2-deoxyglucose (2DG) uptake in vascular smooth muscle cells preexposed to either 1.2, 5.5, or 22 mM glucose. **Lower**, 2DG uptake in bovine endothelial cells preexposed to either 1.2, 5.5, or 22 mM glucose. (From Kaiser N, Feener EP, Boukobza-Vardi N, et al. Differential regulation of glucose transport and transporters by glucose in vascular endothelial and smooth muscle cells. Diabetes 1993; 42:8089.)
Figure 31-3 Overexpression of GLUT1 in mesangial cells cultured in normal glucose mimics the diabetic phenotype. Mesangial cells transfected with either LacZ (MCLacZ)- or GLUT1 (MCGLT)-expressing constructs were cultured in 5-mM glucose, and the amount of the indicated matrix components secreted was determined. (From Heilig CW, Concepcion LA, Riser BL, et al. Overexpression of glucose transporters in rat mesangial cells cultured in a normal glucose milieu mimics the diabetic phenotype. J Clin Invest 1995; 96:1802-1814.)
Figure 31-4 Development of retinopathy during posthyperglycemic normoglycemia ("hyperglycemic memory"). Quantitation of retinal microaneurysms and acellular capillaries in normal dogs, dogs with poor glycemic control for 5 years, dogs with good glycemic control for 5 years, dogs with poor glycemic control for 2.5 years (PGa), and the same dogs after a subsequent 2.5 years of good glycemic control (PGb). (Adapted from Engerman RL, Kern TS. Progression of incipient diabetic retinopathy during good glycemic control. Diabetes 1987; 36:808-812.)
**Figure 31-5** Cumulative incidence of further progression of retinopathy 4 years after the end of the Diabetes Control and Complications Trial. Median glycosylated hemoglobin was 8.2% for the conventional therapy group and 7.9% for the intensive therapy group. EDIC, Epidemiology of Diabetes Interventions and Complications [Research Group]. (From Retinopathy and nephropathy in patients with type 1 diabetes four years after a trial of intensive therapy. The Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Research Group. N Engl J Med 2000; 342:381389.)
Figure 31-7 Adjusted death rates by number of cardiovascular disease (CVD) risk factors for diabetic and nondiabetic men. Subjects are participants from the Multiple Risk Factor Intervention Trial (MRFIT) study; risk factors are hypercho-esterolemia, hypertension, and cigarette smoking. (From Stamler J, Vaccaro O, Neaton JD, Wentworth D. Diabetes, other risk factors, and 12-year cardiovascular mortality for men screened in the Multiple Risk Factor Intervention Trial. Diabetes Care 1993; 2:434444.)
Figure 31-9 Schematic summary relating insulin resistance (IR) to the characteristic dyslipidemia of type 2 diabetes mellitus. IR at the adipocyte results in increased free fatty acid (FFA) release. Increased FFA flux stimulates very-low-density lipoprotein (VLDL) secretion, causing hypertriglyceridemia (TG). VLDL stimulates a reciprocal exchange of TG to cholesteryl ester (CE) from both high-density lipoprotein (HDL) and low-density lipoprotein (LDL), catalyzed by CE transfer protein (CETP). TG-enriched HDL dissociates from ApoA-I, leaving less HDL for reverse cholesterol transport. TG-enriched LDL serves as a substrate for lipases that convert it to atherogenic small, dense LDL particles (SD LDL). (From Ginsberg HN. Insulin resistance and cardiovascular disease. J Clin Invest 2000; 106:453458.)
Figure 31-10 Aldose reductase and the polyol pathway. Aldose reductase reduces reactive oxygen species (ROS)-generated toxic aldehydes to inactive alcohols, and glucose to sorbitol, using triphosphopyridine nucleotide, reduced form of NADP (NADPH) as a cofactor. In cells where aldose reductase activity is sufficient to deplete reduced glutathione (GSH), oxidative stress would be augmented. Sorbitol dehydrogenase (SDH) oxidizes sorbitol to fructose using nicotinamide-adenine dinucleotide (NAD⁺) as a cofactor. GSSG, oxidized glutathione.
Figure 31-11  Potential pathways leading to the formation of advanced glycation end product (AGE) from intracellular dicarbonyl precursors. Glyoxal arises from the auto-oxidation of glucose, 3-deoxyglucosone arises from decomposition of the Amadori product, and methylglyoxal arises from fragmentation of glyceraldehyde 3-phosphate. These reactive dicarbonyls react with amino groups of proteins to form AGEs. Methylglyoxal and glyoxal are detoxified by the glyoxalase system.

Figure 31-12 Potential mechanisms by which intracellular production of advanced glycation end-product (AGE) precursors damages vascular cells. First, intracellular protein modification alters protein function. Second, extracellular matrix modified by AGE precursors has abnormal functional properties. Third, plasma proteins modified by AGE precursors bind to AGE receptors on adjacent cells such as macrophages, thereby inducing receptor-mediated production of deleterious gene products such as cytokines. mRNA, messenger RNA; NFB, neurotropic factor-B; ROS, reactive oxygen species. (Adapted from Brownlee M. Lilly Lecture 1993: Glycation and Diabetic Complications. Diabetes 1994; 43:836841.)
Figure 31-14 Schematic representation of the hexosamine pathway. The glycolytic intermediate fructose-6-phosphate (Fru-6-P) is converted to glucosamine-6-phosphate (Glc-6-P) by the enzyme glutamine:fructose 6-phosphate amidotransferase (GFAT). Increased donation of N-Acetylglucosamine moieties to serine and threonine residues of transcription factors such as Sp1 increases production of such complication-promoting factors as PAI-1 and TGF-β. See text for additional abbreviations. (Adapted from Du XL, Edelstein D, Rossetti L, et al. Hyperglycemia-induced mitochondrial superoxide overproduction activates the hexosamine pathway and induces plasminogen activator inhibitor-1 expression by increasing Sp1 glycosylation. Proc Natl Acad Sci USA 2000; 97:12222-12226.)
Figure 31-15 Production of superoxide by the mitochondrial electron transport chain. Increased hyperglycemia-derived electron donors from the tricarboxylic acid cycle (NADH and FADH$_2$) generate a high mitochondrial membrane potential ($\mu$H$^+$) by pumping protons across the mitochondrial inner membrane. This inhibits electron transport at complex III and increases the half-life of free radical intermediates of coenzyme Q, which reduce O$_2$ to superoxide. See text for abbreviations. (From Boss O, Hagen T, Lowell BB. Uncoupling proteins 2 and 3: potential regulators of mitochondrial energy metabolism. Diabetes 2000; 49:143156.)
Figure 31-16  Effect of agents that alter mitochondrial electron transport chain function on the three main pathways of hyperglycemic damage.  A, Hyperglycemia-induced protein kinase C (PKC) activation.  B, Intracellular advanced glycation end-product (AGE) formation.  C, Sorbitol accumulation. Cells were incubated in 5-mM glucose, 30-mM glucose alone, and 30-mM glucose plus either agents that uncouple oxidative phosphorylation and reduce the high mitochondrial membrane potential (TTFA, CCCP, UCP-1), or dismutate superoxide (Mn-SOD). See text for additional abbreviations. (From Nishikawa T, Edelstein D, Du XL, et al. Normalizing mitochondrial superoxide production blocks three pathways of hyperglycaemic damage. Nature 2000; 404:787-790.)
Figure 31-17 Potential mechanism by which hyperglycemia-induced mitochondrial superoxide overproduction activates four pathways of hyperglycemic damage. Excess superoxide partially inhibits the glycolytic enzyme glyceraldehyde-3-phosphate dehydrogenase, thereby diverting upstream metabolites from glycolysis into pathways of glucose overutilization. This results in increased flux of triose phosphate to diacylglycerol (DAG), an activator of protein kinase C (PKC), and to methylglyoxal, the major intracellular advanced glycation end-product (AGE) precursor. Increased flux of fructose-6-phosphate to UDP- N-acetylglucosamine increases modification of proteins by hexosamine, and increased glucose flux through the polyol pathway consumes NADPH and depletes GSH. See text for additional abbreviations.
Figure 31-18 Diabetic retinopathy pathogenesis flow chart. The schematic flow chart represents the major preclinical and clinical findings associated with the full spectrum of diabetic retinopathy and macular edema. VEGF, vascular endothelial growth factor.
Figure 31-20 Major multicenter clinical trials of diabetic retinopathy. Schematic representation of the major multicenter clinical trials of diabetic retinopathy and the levels of diabetic retinopathy that they primarily addressed. DCCT, Diabetes Control and Complications Trial; DRS, Diabetic Retinopathy Study; DRVS, Diabetic Retinopathy Vitrectomy Study; ETDRS, Early Treatment Diabetic Retinopathy Study; PDR, proliferative diabetic retinopathy.
Figure 31-21 Initial ophthalmic examination flow chart. Schematic flow chart of major principles involved in determining the timing of initial ophthalmic examination following diagnosis of diabetes mellitus. These are maximal recommended guidelines. Ocular symptoms, complaints, or other associated medical issues may necessitate earlier evaluation.
Figure 31-22 Diabetic retinopathy and macular edema examination and treatment flow chart: nonpregnant patients. The schematic flow chart is of major principles involved in determining routine ophthalmic follow-up and indications for treatment in nonpregnant patients with diabetes. These are only general, maximal recommended guidelines. Ocular symptoms, complaints, or other associated ophthalmic or medical issues may necessitate earlier evaluation and/or an altered approach. CSME, clinically significant macular edema; DR, diabetic retinopathy; HRC PDR, high-risk character proliferative diabetic retinopathy; ME, macular edema; NPDR, nonproliferative diabetic retinopathy; PDR, proliferative diabetic retinopathy.
Figure 31-23 Diabetic retinopathy and macular edema examination and treatment flow chart: pregnant patients. The schematic flow chart is of major principles involved in determining routine ophthalmic follow-up and indications for treatment in pregnant patients with diabetes. These are only general, maximal recommended guidelines. Ocular symptoms, complaints, or other associated ophthalmic or medical issues may necessitate earlier evaluation and/or an altered approach. Because retinopathy may progress rapidly in patients with diabetes, careful and more frequent evaluation is often indicated. CSME, clinically significant macular edema; DR, diabetic retinopathy; HRC PDR, high-risk character proliferative diabetic retinopathy; ME, macular edema; NPDR, nonproliferative diabetic retinopathy.
**Figure 31-24** Photocoagulation flow chart. The schematic flow chart details general photocoagulation treatment approaches in patients with diabetic retinopathy and/or diabetic macular edema. These are only general guidelines, and actual treatment choices can be affected by numerous other factors, including findings in the same eye, contralateral eye, systemically, and others. DR, diabetic retinopathy; ME, macular edema; PRP, panretinal (scatter) photocoagulation.
**Figure 31-25** Fluorescein angiogram flow chart. The schematic flow chart details a general algorithm for appropriate use of fluorescein angiography in the ocular evaluation of patients with diabetes mellitus. In unusual cases, confounding factors may alter the appropriate approach.
Figure 31-26  End-stage renal disease (ESRD) among Medicare patients in 1998 with diagnosis specified. USRDS, United States Renal Data System.  *(From The absence of a glycemic threshold for the development of long-term complications: the perspective of the Diabetes Control and Complications Trial. Diabetes 1996; 45:1289-1298.)*
Figure 31-27  A, Periodic acid Schiff (PAS) stain of a normal glomerulus sectioned through the vascular pole. The mesangial cellularity and matrix are normal.  B, Glomerulus showing marked diffuse diabetic glomerulosclerosis. There is extreme expansion of the mesangium with both matrix and cellular material (PAS).  C, Example of nodular diabetic glomerulosclerosis (Kimmelstiel-Wilson lesions). Note the reduction of glomerular capillary luminal space (PAS).  D, Electron photomicrograph showing a normal glomerular basement membrane width (bottom panel) and diffuse thickening of the glomerular basement membrane secondary to diabetes (upper panel).
Figure 31-28 Schematic representation of the progression of diabetic nephropathy. Glomerular hyperfiltration and microalbuminuria are the earliest manifestations of glomerulopathy. Subsequently, urinary protein increases to nephrotic range (>3.5 g/day) and glomerular filtration rate (GFR) declines relentlessly.
Figure 31-29 Flow chart illustrating the management of patients with diabetic nephropathy before the onset of renal failure. ACE, angiotensin-converting enzyme; HIV, human immunodeficiency virus; LDL, low-density lipoprotein.
Figure 31-30 Flow chart illustrating the management of patients with diabetic nephropathy after the onset of clinical proteinuria. ACE, angiotensin-converting enzyme; HIV, human immunodeficiency virus; LDL, low-density lipoprotein.
Figure 31-31 Flow chart illustrating the management of patients after onset of renal failure. ACE, angiotensin-converting enzyme; HIV, human immunodeficiency virus; LDL, low-density lipoprotein.
Figure 31-32 A theoretical framework for the development of diabetic neuropathy. Ab, antibody; AGE, advanced glycation end product; C', complement; DAG, diacylglycerol; EDRF, endothelium-derived relaxing factor; ET, endothelin; GF, growth factor; IGF, insulin-like growth factor; NGF, nerve growth factor; NO, nitric oxide; PKC, protein kinase C; ROS, reactive oxygen species. (Adapted from Vinik AI, Newlon P, Lauterio TJ, et al. Diabetes Rev 1995; 3:139157.)
Figure 31-35 Impaired neurovascular blood flow. (From Stansberry KB, Hill MA, Shapiro SA et al. Impairment of peripheral blood flow responses in diabetes resembles an enhanced aging effect. Diabetes Care 1997; 20:1711-1716.)
**Figure 31-36** Mechanisms of neuropathic pain. **A**, Normal. **B**, C-fiber sensitization. Peripheral nociceptive fibers can be abnormally sensitized and then central nociceptive second-order neurons in the spinal cord dorsal horn can also be sensitized. Then they are hyperexcitable and start responding to non-noxious stimuli. **C**, C-fiber loss or degeneration may trigger anatomic sprouting of low-threshold mechanosensitive terminals to central nociceptive neurons and may subsequently induce synaptic reorganization in the dorsal horn. **D**, Central disinhibition and cold hyperalgesia. A selective damage of cold-sensitive A delta (A) fibers leads to a loss of central inhibition mediated by interneurons (disinhibition), resulting in cold hyperalgesia. DRG, dorsal root ganglion.
Figure 31-37 Diagnostic algorithm for assessment of neurologic deficit and classification of neuropathic syndrome. Ab, antibody; EMG, electromyography; Hx, history; NCV, nerve conduction velocity; NIS, neurologic impairment score; NSS, neurologic symptom score; QAFT, quantitative autonomic function test; QST, quantitative sensory test.
Figure 31-38 Management of painful diabetic neuropathy. (Modified from Vinik AI, Holland MT, LeBeau JM, et al. Diabetic neuropathies. Diabetes Care 1992; 15:19261975.)
Figure 31-39 Diagnosis and treatment algorithm of diabetic autonomic neuropathy. ACE, angiotensin-converting enzyme; ARI, aldose reductase inhibitor; BP, blood pressure; DM, diabetes mellitus; GI, gastrointestinal; HRV, heart rate variability; PGE$_1$, prostaglandin E$_1$; SSR, sympathetic skin response.
Figure 31-40 Marked increase in the risk of coronary artery disease in patients with type 2 diabetes mellitus compared to nondiabetic subjects, in a population-based study in Finland, over a 7-year follow-up period. Patients with diabetes without a history of previous myocardial infarction had an approximately equal risk for first myocardial infarction as nondiabetic subjects who had already sustained a myocardial infarction. These data support recent recommendations from the American Diabetes Association to treat diabetic subjects as if they already have established coronary artery disease. (From Haffner SM, Lehto S, Ronnemaa T, et al. Mortality from coronary heart disease in subjects with type 2 diabetes and in nondiabetic subjects with and without prior myocardial infarction. N Engl J Med 1998; 339:229234.)
Figure 31-41 Age-adjusted cardiovascular disease (CVD) death rates by presence of number of risk factors for men with and without diabetes at baseline screened for the Multiple Risk Factor Intervention Trial. In the presence of diabetes the cardiovascular death rate steeply rises at any level of concomitant risk factors. SBP, systolic blood pressure. (From Stamler J, Vaccaro O, Neaton JD, et al. Diabetes, other risk factors, and 12-year cardiovascular mortality for men screened in the Multiple Risk Factor Intervention Trial. Diabetes Care 1993; 16:434444.)
Figure 31-42 All-cause mortality, cardiovascular mortality, and ischemic heart disease mortality in patients with type 2 diabetes mellitus by quintiles of average fasting blood glucose (FBG). Cardiovascular mortality and all-cause mortality increase throughout the range of fasting plasma glucose in a graded fashion. (From Anderson DK, Svardsudd K. Long-term glycemic control relates to mortality in type II diabetes. Diabetes Care 1995; 18:15341543.)
Figure 31-43 A and B, The high-risk neuropathic foot. Two lateral views of a patient with typical signs of a high-risk neuropathic foot. Note the small muscle wasting, clawing of the toes, and marked prominence of the metatarsal heads. At presentation of type 2 diabetes mellitus, this patient had severe neuropathy with foot ulceration on both right (as noted in these figures) and left feet. (© Copyright University Department of Medical Illustration, Royal Infirmary, Manchester M13 9WJ, England. Reproduced with permission.)
Figure 31-44 Simple algorithm for risk screening in the diabetic foot.
Figure 32-1 Schematic representation of glucose metabolism. TCA, tricarboxylic acid.
Figure 32-4 Summary of studies of the mechanisms of glucose recovery from short-term hypoglycemia in healthy humans. Insulin was injected intravenously at time 0 minutes. Interventions were started at time 0 minutes and stopped at time 90 minutes (i.e., between the vertical lines in each panel). Plasma glucose curves during control studies (solid curves, same in all six panels) and as modified (dashed curves) by the following. A, Somatostatin infusion (glucagon plus growth hormone [GH] deficiency. B, somatostatin infusion plus growth hormone replacement (glucagon deficiency). C, Somatostatin infusion plus glucagon replacement (GH deficiency). D, Phentolamine and propranolol infusion (combined -adrenergic and -adrenergic blockade) or studies performed in bilaterally adrenalectomized individuals (epinephrine deficiency). E, Somatostatin, phentolamine, and propranolol infusion (glucagon deficiency + -adrenergic and -adrenergic blockade). F, Somatostatin infusion in bilaterally adrenalectomized individuals (glucagon + epinephrine deficiency). (Curves derived from data in Clarke WL et al. Am J Physiol 1979; 236:E147E152; Gerich JE et al. Am J Physiol 1979; 236:E370-E385; and Rizza RA et al. J Clin Invest 1979; 64:6271. From Cryer PE. Glucose counter-regulation in man. Diabetes 1981; 30:261264. Copyright 1981, American Diabetes Association, Alexandria, Va.)
Figure 32-6 Neurogenic (autonomic) and neuroglycopenic symptoms of hypoglycemia in normal humans. Among the neurogenic symptoms, "sweaty," "hungry," and "tingling" are cholinergic and "shaky/tremulous," "heart pounding," and "nervous/anxious" are adrenergic. See text for discussion. Mean (±SE) subject scores for awareness of hypoglycemia (blood sugar low) during clamped euglycemia (EU) and during hypoglycemia (Hypo) alone (closed column), with combined α-adrenergic and β-adrenergic blockade with infused phentolamine and propranolol (ADB, crosshatched column), and with combined α-adrenergic and β-adrenergic blockade plus muscarinic cholinergic blockade with atropine, panautonomic blockade (PAB, open column), are also shown. Data from Towler DA et al. Diabetes 1993; 42:1791-1798. (From Cryer PE. Hypoglycemia: the limiting factor in the management of IDDM. Diabetes 1994; 43:1378-1389. Copyright, 1994, American Diabetes Association, Alexandria, Va.)
Figure 32-7 Proportion of patients affected and event rates for severe hypoglycemia (left) and severe hypoglycemia with coma or seizure (right) in the Diabetes Control and Complications Trial. (Data from The Diabetes Control and Complications Trial Research Group. Diabetes 1997; 46:271286. From Cryer PE. Hypoglycemia: the limiting factor in the management of IDDM. Diabetes 1994; 43:13761389. Copyright 1994, American Diabetes Association, Alexandria, Va.)
**Figure 32-8** Mean (±SE) plasma glucose, insulin, epinephrine, and glucagon concentrations during hyperinsulinemic stepped hypoglycemic glucose clamps in nondiabetic subjects (open squares and columns), people with type 1 diabetes mellitus (IDDM, insulin-dependent diabetes mellitus) with classic diabetic autonomic neuropathy (CDAN, open triangles and crosshatched columns), and people with type 1 diabetes mellitus without CDAN (closed circles and columns). (From Dagogo-Jack SE, Craft S, Cryer PE. Hypoglycemia-associated autonomic failure in insulin dependent diabetes mellitus. J Clin Invest 1993; 91:819-828. Copyright, 1994, American Society for Clinical Investigation, New York.)
Figure 32-9 Mean (±SE) plasma glucose, insulin, epinephrine, and glucagon concentrations during hyperinsulinemic stepped hypoglycemic glucose clamps in patients with type 1 diabetes mellitus (IDDM, insulin-dependent diabetes mellitus) without classical diabetic autonomic neuropathy on mornings following afternoon hyperglycemia (Hyper., closed circles and columns) and on mornings following afternoon hypoglycemia (Hypo., open circles and columns). (From Dagogo-Jack SE, Craft S, Cryer PE. Hypoglycemia-associated autonomic failure in insulin dependent diabetes mellitus. J Clin Invest 1993; 91:819-828. Copyright, 1993, American Society for Clinical Investigation, New York.)
Figure 32-10 Mean (±SE) total, neurogenic, and neuroglycopenic symptom scores during hyperinsulinemic, stepped hypoglycemic glucose clamps in patients with type 1 diabetes mellitus (IDDM, insulin-dependent diabetes mellitus) without classic diabetic autonomic neuropathy on mornings following afternoon hyperglycemia (hyper., closed columns) and on mornings following afternoon hypoglycemia (hypo., open columns). (From Dagogo-Jack SE, Craft S, Cryer PE. Hypoglycemia-associated autonomic failure in insulin dependent diabetes mellitus. J Clin Invest 1993; 91:819828. Copyright, 1993, American Society for Clinical Investigation, New York.)
**Figure 32-11** Mean (±SE) baseline and nadir plasma glucose concentrations during morning insulin infusion tests following afternoon hyperglycemia (hyper., closed columns) and following afternoon hypoglycemia (hypo., open columns) in people with type 1 diabetes without classic diabetic autonomic neuropathy. (Data from Dagogo-Jack SE, Craft S, Cryer PE. Hypoglycemia-associated autonomic failure in insulin dependent diabetes mellitus. J Clin Invest 1993; 91:619-628. Copyright, 1993, American Society for Clinical Investigation, New York.)
Figure 32-12 Schematic representation of the pathophysiology of glucose counterregulation in people with type 1 diabetes mellitus (T1DM). See text for discussion.
Figure 32-14 Mean (±SE) neurogenic (autonomic) and neuroglycopenic symptom scores during hyperinsulinemic stepped hypoglycemic glucose clamps in nondiabetic subjects (rectangles) and people with type 1 diabetes (IDDM, insulin-dependent diabetes mellitus) selected for hypoglycemia unawareness studied at baseline before (0 days, open columns), and after 3 days (first set of cross-hatched columns), 3 to 4 weeks (closed columns), and 3 months (second set of cross-hatched columns) of scrupulous avoidance of iatrogenic hypoglycemia. (From Dagogo-Jack S, Rattarasam C, Cryer PE. Reversal of hypoglycemia unawareness, but not defective glucose counterregulation, in IDDM. Diabetes 1994; 43:14261434. Copyright, 1994, American Diabetes Association, Alexandria, Va.)
Figure 32-15 Mean (±SE) plasma glucose concentrations during hypoglycemia produced by subcutaneous insulin injection in people with type 1 diabetes in response to 10 g (circles) and 20 g (squares) of oral (p.o.) glucose and 1.0 mg of subcutaneous (S.C.) glucagon (triangles) compared with placebo (shaded area). (From Wiethop BV, Cryer PE. Alanine and terbutaline in the treatment of hypoglycemia in IDDM. Diabetes Care 1993; 16:1131-1136. Copyright, 1993, American Diabetes Association, Alexandria, Va.)
Figure 32-16 Diagnostic algorithm for suspected hypoglycemia. GI, gastrointestinal.
Figure 33-1 Relationship between body mass index and cardiovascular mortality in adult men and women in the United States who never smoked and had no preexisting illness. The vertical line separates underweight and lean subjects (left side) from overweight and obese subjects (right side). (Adapted from Calle EE, Thun MJ, Petrelli JM, et al. Body-mass index and mortality in a prospective cohort of U.S. adults. N Engl J Med 1999; 341:1097.)
Figure 33-2 Relationship between body mass index and type 2 diabetes in adult men and women in the United States. The vertical line separates underweight and lean subjects (left side) from overweight and obese subjects (right side). The data demonstrate that the risk of diabetes begins to increase at the upper end of the lean body mass index category. (Adapted from Colditz GA, Willett WC, Rotnitzky A, Manson JE. Weight gain as a risk factor for clinical diabetes mellitus in women. Ann Intern Med 1995; 122:481-486; Chan JM, Rimm EB, Colditz GA, et al. Obesity, fat distribution, and weight gain as risk factors for clinical diabetes in men. Diabetes Care 1994; 17:961969.)
Figure 33-3 Progression of 3T3-L1 preadipocyte differentiation with subsequent changes in cellular characteristics. The distinct stages of differentiation (very early, early, intermediate, and late) are shown. C/EBP, CCAAT/enhancer binding protein; DEX, dexamethasone; LPL, lipoprotein lipase; MIX, methylisobutylxanthine; PPAR, peroxisome proliferator-activated receptor. (Modified from Ntambi JM, Kim Y-C. Adipocyte differentiation and gene expression. J Nutr 2000; 130:3122S-3126S.)
Figure 33-4 Medical complications associated with obesity.
Figure 33-5 Weight loss in obese subjects treated with anorexiant medication (sibutramine) alone, medication plus group behavior modification therapy, or medication plus group behavior modification therapy and meal replacements. These data demonstrate that greater weight loss is achieved when antiobesity medications are used in conjunction with lifestyle modification than when they are used alone. (Adapted from Wadden TA, Berkowitz RI, Sarwer DB, et al. Benefits of lifestyle modification in the pharmacologic treatment of obesity. Arch Intern Med 2001; 161:218227.)
Figure 33-7 Percentage of excess weight (mean ± standard deviation) lost over 36 months after the gastric bypass procedure (GBP) and vertical banded gastroplasty (VBGP). (Adapted from Sugerman HJ, Starkey JV, Birkenhauer R. A randomized prospective trial of gastric bypass versus vertical banded gastroplasty for morbid obesity and their effects on sweets versus non-sweets eaters. Ann Surg 1987; 205:613624.)
Figure 34-1 Structures of the common lipids.
Figure 34-2  A, Cholesterol biosynthesis. 3-Hydroxy-3-methyl glutaryl coenzyme A (HMG-CoA) reductase is a rate-limiting enzyme regulating cholesterol biosynthesis. The enzyme is down-regulated by excess cholesterol in the cell.  B, Enterohepatic circulation of cholesterol and bile acids. About 50% of cholesterol and 97% of bile acids are reabsorbed from the intestine and recirculated to the liver.  (A, modified from Brown MS, Goldstein JL. A receptor-mediated pathway for cholesterol homeostasis. Science 1986; 232:3447.)
Figure 34-3 Triglyceride synthesis and the DGAT reaction. 
A. Two major pathways of triglyceride synthesis have been described: the glycerol-phosphate pathway and the monoacylglycerol pathway, which is prominent in the small intestine. 
B. DGAT catalyzes a reaction in which 1,2-diacylglycerol and fatty acyl CoA react to form triacylglycerol at the surface of the endoplasmic reticulum. (From Farese RV Jr, Cases S, Smith SJ. Triglyceride synthesis: insights from the cloning of diacylglycerol acyltransferase. Curr Opin Lipidol 2000; 11:229234.)
Figure 34-4 Cascade of reactions involved in activation of hormone-sensitive lipase. ATP, adenosine triphosphate; cAMP, cyclic adenosine monophosphate; DG, diglyceride; FFA, free fatty acid; MG, monoglyceride. (From Steinberg D, Huttunen JK. The role of cyclic AMP in activation of hormone-sensitive lipase of adipose tissue. Adv Cyclic Nucleotide Res 1972; 1:4762.)
Figure 34-5 General structure of lipoproteins (a schematic representation of very-low-density lipoprotein, VLDL).
Figure 34-6  Polyacrylamide gel showing the various apolipoproteins characteristic of each type of plasma lipoprotein particle. HDL, high-density lipoprotein; LDL, low-density lipoprotein; VLDL, very-low-density lipoprotein. (Modified from Mahley RW, Innerarity TL. Lipoprotein receptors and cholesterol homeostasis. Biochim Biophys Acta 1983; 737:197-222. With permission from Elsevier Science-NL, Sara Burgerhartstraat 25, 1055 KV Amsterdam, The Netherlands.)
Figure 34-7 Synthesis of apolipoprotein B100 (apo-B100) and apo-B48 by a unique messenger ribonucleic acid (mRNA) editing mechanism. In the human intestine, a specific cytosine (C) is changed to a uracil (U) in the apo-B mRNA. This change results in a stop codon and the formation of apo-B48, which contains only the first 2152 amino acids of the full-length apo-B100 (4536 amino acids).
Figure 34-8 Schematic representation of apolipoprotein B100 (apo-B100) on the surface of a low-density lipoprotein (LDL) particle. The receptor-binding domain may form a cluster of positively charged arginine and lysine residues (a basic patch) capable of interacting with critical negatively charged glutamic and aspartic acid residues in the ligand-binding domain of the LDL receptor. (Adapted from Yang C-Y; Gu Z-W; Weng S-A, et al. Structure of apolipoprotein B-100 of human low density lipoproteins. Arteriosclerosis 1989; 9:96108.)
Figure 34-9: Isoelectric focusing gels of very-low-density lipoprotein (VLDL) apolipoproteins from three individuals homozygous for the common apo-E phenotypes. The relative charge differences among the different apo-E isoforms are accounted for by the specific amino acid substitutions that are responsible for the three isoforms. The minor, more acidic apo-E isoforms represent sialylated forms of the protein. (From Mahley RW, Rall SC Jr. Type III hyperlipoproteinemia [dysbeta lipoproteinemia]: the role of apolipoprotein E in normal and abnormal lipoprotein metabolism. In Scriver CR, Beaudet AL, Sly WS, et al [eds]. The Metabolic and Molecular Bases of Inherited Disease, 7th ed. New York, McGraw-Hill, 1995, pp 1953-1980.)
**Figure 34-10** Predicted secondary structure of apolipoprotein E (apo-E). The majority of the structure is composed of alpha helices, beta-sheet structures, and beta turns. A region of random structure encompassing residues 165 to 200 appears to form a boundary or hinge region between the two functional domains. HDL, high-density lipoprotein; VLDL, very-low-density lipoprotein.
Figure 34-11 Schematic representation of the receptor-binding domain of apolipoprotein E, indicating the location and identity of naturally occurring amino acid substitutions that lead to type III hyperlipoproteinemia. In each substitution, the bottom amino acid represents the mutant.
Figure 34-12 Three-dimensional structure of the amino-terminal region (residues 1 to 191) of apolipoprotein E, which forms a four-helix bundle. The receptor-binding domain resides in helix 4. (Modified from Wilson C, Wardell MR, Weisgraber KH, et al. Three-dimensional structure of the LDL receptor-binding domain of human apolipoprotein E. Science 1991; 252:1817-1822.)
Figure 34-13 Low-density lipoprotein (LDL) receptor pathway. The LDLs interact with their receptors on the cell surface. The complex enters the coated pit and is internalized. The coated vesicle loses its clathrin coat and becomes an endosome, the site of lipoprotein and receptor dissociation. The receptors recycle to the cell surface, and the lipoproteins are degraded. Alternatively, new receptors are synthesized in the rough endoplasmic reticulum and transported to the cell surface. (Modified from Brown MS, Goldstein JL. A receptor-mediated pathway for cholesterol homeostasis. Science 1986; 232:3447; and Myant NB. Cholesterol Metabolism, LDL, and the LDL Receptor. San Diego, Academic Press, 1990.)
Figure 34-14 Functional domains of the low-density lipoprotein receptor. See text for complete description. EGF, epidermal growth factor.
Figure 34-15 Low-density lipoprotein (LDL) receptor gene regulation. SREBP, sterol regulatory element binding protein.
**Figure 34-16** Lipoprotein lipase (LPL), attached by interaction with glycosaminoglycans on the endothelial cells, interacts with chylomicrons to catalyze the hydrolysis of the chylomicron triglycerides (Tg) to form free fatty acids (FFA). Apolipoprotein CII on the lipoprotein serves as a cofactor for LPL.
Figure 34-17 General scheme summarizing the major pathways involved in the metabolism of chylomicrons synthesized by the intestine and very-low-density lipoprotein (VLDL) synthesized by the liver. Apo-E, apolipoprotein E; FFA, free fatty acid; HL, hepatic lipase; IDL, intermediate-density lipoprotein. (Modified from Mahley RW. Biochemistry and physiology of lipid and lipoprotein metabolism. In Becker KL [ed]. Principles and Practice of Endocrinology and Metabolism, 2nd ed. Philadelphia, JB Lippincott, 1995, pp 13691378.)
Figure 34-18 Pathways involved in chylomicron remnant metabolism. In sequestration, chylomicron remnants are trapped in the space of Disse, possibly through apolipoprotein E-mediated proteoglycan binding. In processing, enzymes, including lipases, may continue processing the remnants to smaller particles. In uptake, receptors involved in the uptake of the remnants appear to include the low-density lipoprotein (LDL) receptor and the LDL receptor-related protein (LRP).
Figure 34-19 Heparan sulfate proteoglycan (HSPG)/low-density lipoprotein (LDL) receptorrelated protein (LRP) pathway. Apo-E, apolipoprotein E; HL, hepatic lipase; LPL, lipoprotein lipase.
Figure 34-20 Very-low-density lipoprotein (VLDL) biosynthesis by hepatocytes. The nascent apolipoprotein B (apo-B) containing apolipoproteins synthesized by the rough endoplasmic reticulum (RER) apparently combine with the lipids in the smooth endoplasmic reticulum (SER). The VLDLs are processed in the Golgi apparatus and accumulate in large secretory vesicles. They are then released into the space of Disse, from which they enter the plasma. (Modified from Alexander CA, Hamilton RL, Havel RJ. Subcellular localization of B apoprotein of plasma lipoproteins in rat liver. J Cell Biol 1976; 69:241263; by copyright permission of The Rockefeller University Press)
Figure 34.21 Origin of high-density lipoprotein (HDL) from liver, intestine, and surface material from chylomicrons and very-low-density lipoprotein (VLDL). Apo-AI, free cholesterol; HDL-E, HDL with apo-E; LCAT, lecithin:cholesterol acyltransferase; PL, phospholipid; Tg, triglyceride.
Figure 34-22 Role of high-density lipoprotein (HDL) in the redistribution of lipids from cells with excess cholesterol to cells requiring cholesterol or to the liver for excretion. The reverse cholesterol transport pathway is indicated by bold arrows (net transfer of cholesterol from cells HDL LDL liver). CETP, cholesteryl ester transfer protein; FC, free cholesterol; HDL-E, HDL with apolipoprotein E; IDL, intermediate-density lipoprotein; LCAT, lecithin:cholesterol acyltransferase; LDL, low-density lipoprotein; PL, phospholipid; Tg, triglyceride; VLDL, very-low-density lipoprotein.
Figure 34-23 Relation between plasma cholesterol levels and coronary heart disease (CHD) mortality in the Multiple Risk Factor Intervention Trial. (Modified from Stamler J, Wentworth D, Neaton JD. Is relationship between serum cholesterol and risk of premature death from coronary heart disease continuous and graded? Findings in 356,222 primary screenees of the Multiple Risk Factor Intervention Trial (MRFIT). JAMA 1986; 256:28232828; Copyright © 1986, by the American Medical Association.)
Figure 34-24 Coronary heart disease mortality (10-year death rates) versus median serum cholesterol (mg/dL). All men were free of heart disease at the beginning of the study. B = Belgrade, Serbia; C = Crevalcore, Italy; D = Dalmatia, Croatia; E = East Finland; G = Corfu, Greece; I = Italian railroad workers (Rome division); K = Crete, Greece; M = Montegiorgio, Italy; N = Zutphen, The Netherlands; R = American railroad workers (Green Bay, WI; San Francisco, CA; Seattle, WA); S = Slavonia, Croatia; T = Tanushimaru, Japan; U = Ushibuka, Japan; W = West Finland; Z = Zrenjanin, Serbia. (From Keys A. Seven Countries: A Multivariate Analysis of Death and Coronary Heart Disease. Cambridge, Mass, Harvard University Press, 1980.)
Figure 34-25 Schematic representation of the progression of atherogenesis.  
A. Normal artery wall showing the three major regions of the vessel wall: intima, media, and adventitia. The thickness of the intima beneath the endothelial cell layer is exaggerated relative to the media to allow illustration of the changes that occur within the subendothelial intima.  
B. Initial events in lesion formation include recruitment of monocyte-macrophages to the subendothelial space and the infiltration of plasma low-density lipoproteins (LDLs) (small circles), which are oxidized by unknown mechanisms that may include reactive oxygen species. Oxidized LDLs are taken up by macrophages, leading to the formation of foam cells. MCP-1, monocyte chemoattractant protein 1.  
C. Fatty streak lesion. Further recruitment of monocyte-macrophages from the plasma takes place along with smooth muscle cell proliferation and collagen synthesis (rows of vertical lines). Elastin fibers (thin curved lines) begin to accumulate.  
D. Proliferative or fibrous lesion. Atherogenesis continues as the lesion begins to extend into the vessel lumen. Necrosis of foam cells begins, and smooth muscle cells start to migrate from the media through the disrupted internal elastic lamina. Some smooth muscle cells accumulate lipid droplets.  
E. Complicated lesion. The endothelial cell layer covering the lesion is lost. As a result, the surface of the lesion becomes thrombogenic, inducing thrombus formation. Cellular debris increases. Calcification and appearance of cholesterol crystals can occur.
Figure 35-1  X chromosome inactivation analysis of tumor clonality with the use of the M27 polymorphism.  

A. Diagrammatic illustration of general principles. Left, Lightly shaded and dark rectangles represent the maternally and paternally inherited copies of the two X chromosomes of somatic cells early in the development of a female embryo. As embryogenesis proceeds, one of the X chromosomes in each somatic cell is randomly chosen for inactivation (lyonization); the inactivated chromosome is represented as a small oval with shading corresponding to its origin. Subsequently, daughter cells (third column) faithfully maintain the same selection of inactivated X chromosome as found in their parent cells. Accordingly, an adult tissue typically contains a mixture of approximately 50% cells with the maternally inherited X chromosome inactive and 50% with the paternally inherited X chromosome inactive. Lower right, A polyclonal tumor arising from a large number of cells in a tissue, maintains this relatively even mixture of cells with different X inactivation patterns. Upper right, A monoclonal tumor, derived from a single somatic cell, has a uniform pattern of X chromosome inactivation in all cells. B, Partial restriction endonuclease map of the M27 locus (DXS255) and an example of two distinguishable M27 alleles. The variable number of tandem repeat (VNTR, minisatellite) region, which is highly variable in its length from person to person, is shown in stripes. A 2.5-kb DNA fragment used as the hybridization probe is shown as a solid rectangle. Cleavage sites for restriction enzyme  PstI  flank the locus. The enzyme  MspI  cleaves the sequence CCGG whether or not the internal cytosine is methylated. In contrast, the enzyme  HpaII  cleaves this sequence only if the internal cytosine is unmethylated. The diagrammed  MspI-HpaII  site actually represents a 270-base-pair region containing three such sites, two of which vary in their methylation status in accord with location on the active versus the inactive X chromosome. In the example of two distinguishable alleles, variation in size of the minisatellite repeat region (VNTR) causes a difference in the size of the  PstI  restriction fragment detectable by hybridization to the labeled M27 probe. If an individual with these two alleles had a monoclonal tumor, in which the larger allele was uniformly associated with the active X chromosome, the  MspI-HpaII  site of this allele would be consistently methylated and, therefore, resistant to cleavage by  HpaII  (asterisks). The resulting Southern blot pattern would correspond to monoclonal pattern 1 in C. C, Schematic diagram of prototypical Southern blot hybridization patterns for X inactivation analysis using M27. A monoclonal tumor can exhibit only one of the two monoclonal patterns shown. The  PstI  +  MspI  control digestion is useful for marking the sizes of fully cleaved alleles. (A to C, From Arnold A, Brown MF, Urena P, et al. Monoclonality of parathyroid tumors in chronic renal failure and in primary parathyroid hyperplasia. J Clin Invest 1995; 95:2047-2053. Copyright, The American Society for Clinical Investigation.)
Figure 35-2 Diagram of the molecular structure of the parathyroid hormone-cyclin D1 DNA rearrangement in a subset of parathyroid adenomas and its functional consequences. The dark X represents the chromosome breakpoint between the PTH gene regulatory region, plus PTH noncoding exon 1 (solid light vertical bar) and part of its first intron, from 11p15 (left), and the intact promoter and five exons of the cyclin D1 gene from 11q13. Cyclin D1 gene transcription proceeds in a left-to-right direction, as drawn. (Modified from Arnold A. Genetic basis of endocrine disease: 5. Molecular genetics of parathyroid gland neoplasia. J Clin Endocrinol Metab 1993; 77:11081112. Copyright 1993, The Endocrine Society.)
Figure 35-3 Schematic representation of the normal RET protein (upper) and the oncoproteins created by selected RET-PTC oncogenes in papillary thyroid cancer (lower). For each indicated rearranged oncoprotein, the number of N-terminal amino acids contributed by the specified partner gene is shown, fused to the tyrosine kinase domain and C-terminus of RET. aa, amino acid; TMD, transmembrane domain. (From Pasini B, Ceccherini I, Romeo G. RET mutations in human disease. Trends Genet 1996; 12:138-144. Copyright 1996, Elsevier Science.)
**Figure 36-1** Age at onset for endocrine tumor expressions in multiple endocrine neoplasia type 1 (MEN1). Data from retrospective analysis of multiple tumor expressions in 130 inpatients with MEN1 during 15 years. Age of tumor onset was defined as the earlier of age at first symptom and age at first abnormal test result. (Modified from Marx S, Spiegel AM, Skarulis MC, et al. Multiple endocrine neoplasia type 1: clinical and genetic topics. Ann Intern Med 1998; 129:484494.)
Figure 36-2 Parathyroid gland sizes at initial parathyroidectomy for 18 cases with familial multiple endocrine neoplasia type 1. Volumes of all glands at one operation are connected by a vertical line. Dashed horizontal line is upper limit of normal gland volume) 0.075 cm$^3$, equivalent to 75 mg mass). (Modified from Marx SJ, Menczel J, Campbell G, et al. Heterogeneous size of the parathyroid glands in familial multiple endocrine neoplasia type 1. Clin Endocrinol (Oxf) 1991; 35:521526.)
Figure 36-3  Tumor multiplicity within a tissue in multiple endocrine neoplasia type 1 (MEN1). Top, Hypercellular parathyroid gland from patient with MEN1. The gland is totally replaced by diffuse sheets and two discreet nodules of chief cells. This image could reflect three or more second hits to the normal copy of the MEN1 gene in three different clone precursor cells and thus growth of three or more independent clones. An alternative pathogenesis could be stepwise evolution from one clone, that is, third hits to genes other than MEN1. Bottom, Duodenal mucosa from a second MEN1 patient, showing two large, submucosal microgastrinomas. Each tumor was positive for gastrin immunostain and negative for other peptide hormones. Possible development of these two adjacent tumors could have followed mechanisms suggested for the two parathyroid nodules at the top. (From I. Lubensky, National Institutes of Health, Bethesda, Md.)
Figure 36-4 Effect of parathyroidectomy in patients with multiple endocrine neoplasia type 1 and Zollinger-Ellison syndrome. Basal acid output and fasting serum gastrin are shown. All patients became normocalcemic except for one (case 4), who remained hypercalcemic. Dashed line is upper limit of normal for serum gastrin. Upper limit of normal for basal acid output is 15 mEq/hour. (From Jensen RT. Management of the Zollinger-Ellison syndrome in patients with multiple endocrine neoplasia type 1. J Intern Med 1998; 243:477488.)
Figure 36-5  Intact parathyroid hormone (PTH) by rapid assay during parathyroid surgery. Normal range is indicated by dashed lines. The patient had multiple endocrine neoplasia type 1 and primary hyperparathyroidism without prior parathyroidectomy. Three and one half similarly enlarged parathyroid glands (0.8 to 1.6 g; normal less than 0.08 g) and the accessible portions of the thymus were removed at the times indicated; the thymus contained no parathyroid tumor. A rapid fall of PTH below a cutoff criterion (such as 50% drop within 5 minutes) indicates that little or no hyperfunctioning parathyroid tumor remains. Note that removal of the first two parathyroid tumors was not followed by a fall in PTH. The PTH assay result for each time point was available within several minutes to help establish the time at which no hyperfunctioning parathyroid tumor remained and thereby contribute to serial decisions about extending or ending the operation.  (From S. K. Libutti, H. R. Alexander, and A. Remaley, National Institutes of Health, Bethesda, Md.)
Figure 36-6 Serum ionized calcium and peripheral midregion parathyroid hormone (PTH) levels after total parathyroidectomy and graft of parathyroid tissue in a patient with multiple endocrine neoplasia type 1. After total parathyroidectomy (PTX) and grafting of parathyroid tissue to the nondominant forearm, the ionized calcium and PTH levels fell to subnormal values. Subsequent measurements demonstrated a continuous rise of the ionized calcium and peripheral PTH levels (taken from the arm not containing the transplant) over a 60-month period, which necessitated removal of some of the grafted parathyroid tissue. The numbers in parentheses represent selected PTH values for blood taken from the brachial vein immediately downstream from the grafted parathyroid tissue. These results demonstrate continued secretion of PTH by the graft in the presence of hypercalcemia. The upper, lighter shaded area shows the normal range for ionized calcium; the lower, darker shaded area shows the normal range for serum midregion PTH. (Data from L. E. Mallette, Baylor College of Medicine.)
Figure 36-7 Facial angiofibroma in patients with multiple endocrine neoplasia type 1. A small, light pink lesion on the vermilion border of the lip (top) and a large, reddish angiofibroma on the nose (bottom) are shown. Typical lesions are smaller and may require biopsy for confirmation. (From T. Darling and M. Turner, National Institutes of Health, Bethesda, Md.)
Figure 36-8  Germ line and somatic mutations of the MEN1 gene. Unique germ line MEN1 mutations in families, sporadic cases, and nonhereditary tumors. Germ line mutations are shown as vertical check marks about the messenger ribonucleic acid (mRNA). Somatic MEN1 mutations in diverse tumors are shown separately as "flags" along the upper and lower border. MEN1 mRNA is diagrammed with exons numbered; untranslated regions are crosshatched. Truncating mutations (frameshift mutations, splice error, and nonsense [stop codon] mutations) are shown above the mRNA. Codon change mutations (missense mutations or small in-frame deletions) are shown below. Repeating mutations within the germ line or somatic category are shown only once, with a small number in parentheses to indicate total occurrences. Stippling about exon 3 represents the main zone of menin interaction with junD. One large deletion, probably of the entire MEN1 gene, is not shown.\[286\] NLS, nuclear localization sequence.
**Figure 36-9** Test categories and test methods in a hereditary tumor syndrome. Tests of the germ line carrier status (left) are largely distinguished from tests of tumor status (right). When DNA testing is not informative, carrier status can be tested by streamlined surveillance of tumors.
Figure 36-10 Bilateral medullary thyroid carcinoma in multiple endocrine neoplasia type 2A. Large bilateral foci of medullary thyroid carcinoma are located in each lobe of the thyroid gland.
**Figure 36-11** Progression of histologic changes from C-cell hyperplasia to medullary thyroid carcinoma. These sections were taken from a single thyroid lobe of a patient with hereditary medullary thyroid carcinoma and demonstrate the multicentric nature of this tumor. *A*, Nodular hyperplasia with containment of C cells within a thyroid follicle. Magnification ×250. *B*, Microscopic medullary thyroid carcinoma that is locally invasive. Magnification ×100.
Figure 36-12 A pheochromocytoma set on a background of diffuse adrenomedullary hyperplasia in multiple endocrine neoplasia type 2A. In the normal adrenal gland, the adrenal cortices are separated by a thin (less than 1 mm) band of adrenal medulla (not shown). In this pheochromocytoma there is diffuse expansion of the adrenal medulla.
**Figure 36-13a** Surveillance for pheochromocytoma in multiple endocrine neoplasia type 2 (MEN2) using catecholamines and their metabolites. Panel A, Results of 24-hour urinary norepinephrine excretion, epinephrine excretion, and ratio of epinephrine to norepinephrine in 11 prospectively surveyed patients with MEN2A proved to have pheochromocytoma. Each open square indicates a value or the mean of two or more values for a patient, and each filled square represents the mean for all the subjects in a particular year; the latter symbols are connected by a solid line. The dashed line shows the upper limit of normal. To convert epinephrine values to nanomoles, multiply by 5.5; to convert norepinephrine values to nanomoles, multiply by 5.9.
Figure 36-13b Panel B, Plasma concentrations of normetanephrine, norepinephrine, metanephrine, and epinephrine (top) and urinary excretion of norepinephrine, epinephrine, metanephrines, and vanillylmandelic acid (bottom). The values are expressed as percentages of the upper reference limit for each test. Data on individual patients are shown for three groups of patients with von Hippel-Lindau (VHL) disease and MEN2 as follows: patients with VHL disease or MEN2 in whom a pheochromocytoma was ruled out on the basis of normal computed tomography (CT-negative), patients with VHL disease who had histologically verified pheochromocytomas (VHL), and patients with MEN2 who had histologically verified pheochromocytomas (MEN2). The values for patients with pheochromocytoma were determined when the tumors were first identified by CT. The dotted horizontal line represents the upper reference limit for each test. The scales are logarithmic. (A, From Gagel RF, Tashjian AH Jr, Cummings T, et al. The clinical outcome of prospective calcitonin-based surveillance for multiple endocrine neoplasia type 2A: an 18-year experience. N Engl J Med 1988; 318:478484. B, Data from Eisenhofer G, Lenders JWM, Linehan WM, et al. Plasma normetanephrine and metanephrine for detecting pheochromocytoma in von Hippel-Lindau disease and multiple endocrine neoplasia type 2. N Engl J Med 1999; 340:18721879.)
Figure 36-14 Cutaneous and oral manifestations in multiple endocrine neoplasia type 2 (MEN2) variants. A. The characteristic clinical picture of cutaneous lichen amyloidosis associated with MEN2A. The pruritic skin lesion may cover a small area or the entire right or left upper back, as shown in this patient. B and C. Patient with MEN2B demonstrating thick bumpy lips and eversion of upper eyelids (B) and neuromas on anterior third of tongue (C). (A, From Gagel RF, Levy ML, Donovan DT, et al. Multiple endocrine neoplasia type 2A associated with cutaneous lichen amyloidosis. Ann Intern Med 1989; 111:802806; B and C, from Brown RS, Colle E, Tashjian AH Jr. The syndrome of multiple mucosal neuromas and medullary thyroid carcinoma in childhood. Importance of recognition of the phenotype for the early detection of malignancy. J Pediatr 1975; 86:7783.)
Molecular abnormalities of the RET proto-oncogene in multiple endocrine neoplasia type 2 (MEN2). Mutations of the RET proto-oncogene have been identified in MEN2A, familial medullary thyroid carcinoma (FMTC), MEN2A associated with Hirschsprung disease, MEN2A associated with cutaneous lichen amyloidosis (CLA), and as somatic mutations in sporadic MTC. Two regions of the RET tyrosine kinase are affected. The first is a cysteine-rich extracellular domain (Cys-Rich) important for dimerization of the ret receptor (codons 609, 611, 618, 620, 634). Mutations of individual cysteines at these codons cause RET dimerization, activation, autophosphorylation, and transformation. Mutations of the second region, the intracellular tyrosine kinase (TK) domain involving codons 768, 790, 791, 804, 883, 891, 918, and 922, cause activation, autophosphorylation, and transformation. A role for the cadherin-like region (Cadherin) has not been defined, although it may be involved in an interaction with the glial cell line-derived neurotrophic factor receptor. The most common germ line mutation is a codon 634 mutation that converts a cysteine to an arginine and accounts for 50% or more of all MEN2 mutations. Somatic mutations of codons 768, 804, and 918 have been identified as somatic mutations in sporadic MTC. Codon 768 and 804 mutations are rare; a somatic codon 918 mutation is identified in approximately 25% of sporadic MTCs. (TM, transmembrane domain.)
Figure 36-16 The RET tyrosine kinase and glial cell linederived neurotrophic factor receptor signaling system.  

A. The RET receptor is a tyrosine kinase receptor that couples with the glial cell linederived neurotrophic factor receptor (GFR-1) to form a receptor for glial cell linederived neurotrophic factor (GDNF). In the absence of GDNF, RET and GFR-1 exist in an undimerized form. Addition of ligand results in activation of the receptor system, autophosphorylation (P), and activation of downstream signaling pathways (phospholipase C [PLC], p38MAPK, and JNK pathways).  

B. Mutations of the extracellular cysteine-rich domain (codon 634) cause dimerization, autophosphorylation of the RET receptor complex, and activation of downstream signaling pathways.  

C. Mutations of the intracellular tyrosine kinase (codon 918) cause autophosphorylation and activation of the kinase domain in the absence of dimerization. (MAPK, mitogen-activated protein kinase; JNK, c-Jun N-terminal terminal.)
Figure 36-17 Trisomy 10 with nonrandom duplication of the mutant RET allele in pheochromocytoma associated with multiple endocrine neoplasia type 2 (MEN2). A. Representative interphase fluorescence in situ hybridization analysis of tumor touch preparation from patient 2 (tumor 2A). Three copies of chromosome 10 are shown using a centromeric satellite probe (fluorescein isothiocyanate, green signal) specific for chromosome 10. B. Combined pedigree and tumor allelic analysis of patient 2 (Pt2). Arrow, patient 2. Filled symbols, individuals with MEN2. Genotypes are shown for the chromosome 10 microsatellite marker D10S1239 linked to the RET locus. Allele 2 of D10S1239 is co-inherited with the disease in this patient's family. In patient 2, allele 2 shows greater intensity in lanes 2A and 2B (tumors) than allele 1, representing the wild-type allele, as compared with lane N2 (blood DNA). Lane N1 (blood DNA) shows equal intensities of mutant and wild-type allele in the patient's affected cousin (C). C. Representative results of microsatellite and phosphorimage analyses. After polymerase chain reaction amplification using marker D10S1239, quantitative measurement of allelic intensity was performed using phosphorimage analysis. In tumor tissue (T), allele 2 is more intense than allele 1. Phosphorimage densitometry shows a 2:1 imbalance between the two alleles in the tumor (T) compared with the normal tissue (N). D. Representative results of sequencing analysis of RET in tumor 3A. Blood DNA from an unaffected healthy individual (C, left) shows the wild-type RET sequence (codon 631 GAC). Blood DNA from patient 3 (N, middle) shows the germ line mutation (G/T). Tumor DNA (T, right) shows a higher intensity of the mutant nucleotide (T) compared with the wild type. (From Huang SC, Koch CA, Vortmeyer AO, et al. Duplication of the mutant RET allele in trisomy 10 or loss of the wild-type allele in multiple endocrine neoplasia type 2 associated pheochromocytomas. Cancer Res 2000; 60:6226-26.)
Figure 37-1 Reproduction of a plate from Addison's initial description of primary adrenal insufficiency (Addison's disease) (A) and hand of a patient with vitiligo and hyperpigmentation of Addison's disease (B). (A, From Addison T. On the Constitutional and Local Effects of Disease of the Supra-renal Capsules. London, Samuel Highley, 1855, plate XI; B, courtesy of F. Neelon.)
Figure 37-2 Autoimmune polyendocrine syndrome type II family with Addison's disease and type 1 diabetes.
**Figure 37-3** 21-Hydroxylase autoantibodies of patients with known Addison's disease, normal control subjects, and patients with type 1 diabetes mellitus. Data for the 15 patients with type 1 diabetes discovered to be 21-hydroxylase positive on screening are plotted on the right, where multiple different serum samples are available for an individual, values for each individual are connected by lines. (From Yu L, Brewer KW, Gates S, et al. DRB1*04 and DQ alleles: expression of 21-hydroxylase autoantibodies and risk of progression to Addison's disease. J Clin Endocrinol Metab 1999; 84:328335.)
Figure 37-4 Adrenal antibody (AA) titers and levels of adrenocorticotropin hormone (ACTH), cortisol, plasma renin activity (PRA), and aldosterone in three antiadrenal autoantibody-positive patients treated for 6 months with glucocorticoids for concomitant Graves' ophthalmopathy. (From De Bellis A, Bizzaro A, Rosai R, et al. Remission of subclinical adrenocortical failure in subjects with adrenal autoantibodies. J Clin Endocrinol Metab 1993; 76:10021007.)
Figure 37-5 Age at onset of mucocutaneous candidiasis, hypoparathyroidism, and adrenal insufficiency in patients with autoimmune polyendocrine syndrome type I. (From Neufeld M, Maclaren NK, Blizzard RM. Two types of autoimmune Addison's disease associated with different polyglandular autoimmune [PGA] syndromes. Medicine [Baltimore] 1981; 60:355362.)
**Figure 38-1** Essential requirement for Pax6 for glucagon-positive enteroendocrine cell formation in the murine intestine. Pax6 SEY<sup>+/−</sup> mutant mice (I) exhibit markedly reduced numbers of glucagon-immunopositive cells in the small and large intestine.
Figure 38-2 Glucagon-like peptide II receptor (GLP-2R) expression in subsets of endocrine cells in the human stomach (ST) and large bowel (LB). Most cells exhibiting positivity with antisera against the human GLP-2R also exhibited immunopositivity for an endocrine marker such as chromogranin (CHROM). In contrast, most endocrine cells in the stomach and both small and large intestine did not express the GLP-2R. Arrows denote cells positive for both the GLP-2R and chromogranin, and arrowheads denote cells positive for the GLP-2R or chromogranin. (From Yusta B, Huang L, Munroe D, et al. Endocrine localization of GLP-2 expression in humans and rodents. Gastroenterology 2000; 119:744755.)
Figure 38-3. Clinically "nonfunctioning" tumors are often found to express one or more peptide hormones after immunocytochemical analyses. The photomicrographs represent histologic sections from the identical nonfunctioning human pancreatic endocrine tumor that exhibit immunopositivity for glucagon (A) and pancreatic polypeptide (B). (Courtesy of Dr. G. Rindi, Brescia, Italy.)
Figure 38-4 Treatment algorithm for the management of a patient with gastrinoma. The dotted line indicates that in some circumstances, patients with familial gastrinoma may also be candidates for surgical resection if disease is highly limited. MEN 1, multiple endocrine neoplasia type 1; CT, computed tomography; MRI, magnetic resonance imaging; PET, positron emission tomography.
Figure 38-5  Somatostatin immunoreactivity in a human duodenal D cell tumor. The low-power micrograph illustrates the diffuse somatostatin immunoreactivity. Brunner's glands and the partly eroded mucosa are seen in the lower and upper right areas, respectively, in relation to the immunopositive endocrine tumor.
Figure 39-1 Diagrammatic representation of mechanisms by which estradiol causes breast cancer. By binding to its receptor and stimulating genes involved in cell proliferation, estradiol \( (E_2) \) increases the rate of cell division. The chances of error during deoxyribonucleic acid (DNA) replication increase as the number of dividing cells is enhanced. The number of mutations increases, and cancer ultimately develops. This mechanism can be called the cell proliferation mechanism of cancer (see Preston-Martin et al, 1993\[24\]). Increasing data suggest that metabolites of estradiol are directly genotoxic and result in deletions of DNA segments and point mutations (see Jefcoate et al, 2000\[28\]; Yager and Liehr, 1996\[29\]; and Cavalleri et al, 2000\[468\]). These two mechanisms may act in an additive or synergistic fashion to cause breast cancer. (From Santen RJ. Symposium overview. J Natl Cancer Inst Monogr 2000; 27:1516.)
Figure 39-2 Relative risk of breast cancer as a function of several factors related to long-term exposure to estradiol (see Zhang et al, 1997; Boyd et al, 1995; Cauley et al, 1996; Hulka, 1997; Huang et al, 1997; and Mouridsen, 2001). E, estrogen; E₂, estradiol; HRT, hormone replacement therapy; OOX, oophorectomy; P, progesterone.
**Figure 39-4 A**, Increase in relative risk of breast cancer in women taking estrogen alone as hormone replacement therapy (HRT). *Solid line* represents the mean increase in relative risk over time. *Shaded area* represents the 95% confidence limits of that risk. **B**, Increase in relative risk of breast cancer in women taking estrogen plus a progestin as HRT. *Solid line* represents the mean increase in relative risk over time. *Shaded area* represents the 95% confidence limits of that risk. (From Santen RJ, Pinkerton J, McCartney C, et al. Risk of breast cancer with progestins in combination with estrogen as hormone replacement therapy. J Clin Endocrinol Metab 2001; 86:1623.)
Figure 39-5 Relative risk of invasive breast cancer related to several benign breast lesions in women with long-term follow-up.
Figure 39-6. A, Reduction in risk of breast cancer in response to administration of tamoxifen (TAM) or raloxifene for a mean duration of 4 years (see Fisher et al, 1998; and Cummings et al, 1999). RR, relative risk; SERM, selective estrogen receptor modulator. B, Absolute benefit from tamoxifen expressed as percentage of women whose breast cancer was prevented as a function of the underlying risk factor present. ADH, atypical ductal hyperplasia; ER, estrogen receptor; LCIS, lobular carcinoma in situ; NSABP, National Surgical Adjuvant Breast Project.
Figure 39-7 Prognostic value of several parameters related to patients with an initial diagnosis of breast cancer. All values are presented as the percent difference in survival at the 5-year interval. This method of presentation allows one to determine the increased number of women per 100 who will be alive at 5 years if they have a favorable prognostic factor compared with those with an unfavorable factor. LI, labeling index.
Figure 39-8 Aromatase activity remaining during the administration of first-generation, second-generation, and third-generation aromatase inhibitors and inactivators. Data are expressed on a logarithmic scale to emphasize the expected log dose-response characteristics of hormone actions. With the most potent inhibitor, only 1% of aromatase activity persists during therapy. Degree of aromatase suppression determined by an isotopic kinetic method using $^3$H-labeled androstenedione and $^{14}$C-labeled estrone to assess the rho value before and during therapy. The rho value represents the percent conversion of androgens to estrogens under equilibrium conditions. ANA, anastrozole; EXE, exemestane; FAD, fadrozole; For, formestane; LET, letrozole. The numbers under bars represent daily dose in milligrams. (Data from Dowsett et al, 1995; Geisler et al, 1996; Geisler et al, 1996; Jones et al, 1992; MacNeill et al, 1992; and Geisler and Hayes.)
Figure 39-9  Cumulative frequency analysis of development of amenorrhea in women with breast cancer as a function of age in response to chemotherapy, hormonal therapy, or both in comparison with no treatment. The effect of chemotherapy on ovarian function reduces the age of menopause by an average of nearly 20 years. (Data from Goodwin PJ, Ennis M, Pritchard KI, et al. Risk of menopause during the first year after breast cancer diagnosis. J Clin Oncol 1999; 17:2365-2370.)
Figure 39-10 Absolute benefit from tamoxifen in the treatment and prevention setting. Absolute benefit is defined as the number of women per 100 who benefit from the use of tamoxifen. In the adjuvant or treatment setting, tamoxifen was used for a period of 5 years. Adj, adjuvant; DCIS, ductal carcinoma in situ; ER, estrogen receptor; Rx, treatment. (Data from Early Breast Cancer Trialists’ Collaborative Group, 1998; Abe, 1994; Carstensen et al, 2000; Cufer, 2000; Delozier et al, 2000; and Ferno et al, 2000.)
Figure 39-11 Disease-specific survival in percent in women receiving tamoxifen or a placebo as adjuvant therapy for breast cancer. Data taken from a meta-analysis of adjuvant trials for breast cancer. SD, standard deviation; -ve, node negative; +ve, node positive. (From Ovarian ablation in early breast cancer: overview of the randomized trials. Early Breast Cancer Trialists’ Collaborative Group. Lancet 1996; 348:1189-1196. Published with permission of the Lancet.)
**Figure 39-12** Results from three trials directly comparing tamoxifen with an aromatase inhibitor in postmenopausal women. CR represents complete objective tumor response, PR represents partial objective response, and clinical benefit is defined as complete and partial objective responses plus stable disease for greater than or equal to 6 months. ANA, anastrozole; LET, letrozole; SD, stable disease for greater than or equal to 6 months; TAM, tamoxifen. The numbers below each column represent the number of women in the arm of the study considered. (Data from Bonneterre et al, 2000[217]; Nabholtz et al, 2000[218]; Smith et al, 2001[220]; and Mouridsen et al, 2001[473].)
Figure 39-13 Survival data for patients undergoing a prophylactic oophorectomy in comparison with a group of women not receiving this therapy. (From Ovarian ablation in early breast cancer: overview of the randomized trials. Early Breast Cancer Trialists’ Collaborative Group. Lancet 1996; 348:11891196. Reprinted with permission of the Lancet.)
Figure 39-14 Percent reduction in tumor volume in response to either letrozole, anastrozole, or tamoxifen given for 3 months prior to the time of excisional surgery and used as neoadjuvant therapy. (From Dixon JM. Neoadjuvant endocrine therapy. In Miller WR, Santen RJ (eds). Aromatase Inhibition and Breast Cancer. New York, Marcel Dekker, 2001, p 109.)
Figure 39-16  Two analogous systems for classification of prostate cancer. The Tumor-Node-Metastasis (TNM) system is generally used for a wide range of neoplasms; it involves an estimate of tumor size and presence of distant metastases or lymph nodes that contain tumor. The Jewett classification integrates these factors into stages A to D, which indicate progressively more severe disease. Both are used by various authors, and the parallels between the two systems are demonstrated here. PSA, prostate-specific antigen; TRUS, transrectal ultrasonography.
Figure 39-17 Diagrammatic representation of the endocrinology of prostate cancer growth. Androgens are directly secreted by the testis into the circulation, are synthesized in peripheral tissues from steroidal precursors, and are formed directly in the prostate gland. AR, androgen receptor; DHA, dehydroepiandrosterone; DHT, dihydrotestosterone; HRE, hormone response element. (From Denis LJ, Griffiths K. Endocrine treatment in prostate cancer. Semin Surg Oncol 2000; 18:5254. Published with permission of Seminars in Surgical Oncology.)
Figure 39-18  Illustration of death rates in men of various ages with prostate cancer. Heavily shaded areas represent death from all causes. Lightly shaded areas indicate prostate cancer related death rates. (From Albertsen PC, Hanley JA, Gleason DF, Barry MJ. Competing risk analysis of men aged 55 to 74 years at diagnosis managed conservatively for clinically localized prostate cancer. JAMA 1998; 280:975. Reprinted with permission of the American Medical Association.)
Figure 39-19 Two trials illustrate the benefit of adjuvant androgen deprivation therapy in patients with locally advanced disease treated with radiation therapy and either placebo (open bars) or medical orchiectomy with a gonadotropin-releasing hormone agonist analogue (solid bars). EORTC, European Organization for Research and Treatment of Cancer; RTOG, Radiation Therapy Oncology Group.
**Figure 39-20** Prostate tissue levels of dihydrotestosterone (DHT) in the normal prostate gland and after surgical orchiectomy alone and in combination with suppression of adrenal androgens. n.d., nondetectable. *(From Denis LJ, Griffiths K. Endocrine treatment in prostate cancer. Semin Surg Oncol 2000; 18:5274. With permission of the publisher.)*
Figure 39-21 Results of complete versus partial androgen blockade in men with prostate cancer. Survival curves represent pooled data from a meta-analysis of multiple studies. (From Maximum androgen blockade in advanced prostate cancer: an overview of the randomised trials. Prostate Cancer Trialists’ Collaborative Group. Lancet 2000; 355:14911498. With permission of the publisher.)
Figure 39-22 Algorithm of preferred treatment strategies for localized and advanced prostate cancer. Controversial therapies are represented by an asterisk or dagger.
Figure 40-1 Plasma concentration of parathyroid hormone-related protein (PTHrP) in patients with hyperparathyroidism (HPT), normocalcemic patients with malignancy (Normocalc), and patients with hypercalcemia of malignancy caused by a solid tumor (Solid) or a hematologic malignancy (Hematol). Radioimmunoassay for amino-terminal PTHrP (left panel), an immunoradiometric assay for PTHrP (middle panel), and a radioimmunoassay for midregion PTHrP (right panel). The hatched area represents the normal ranges; the dashed line denotes the limits of detection; the numbers attached to each group indicate the number of patients. In the PTHrP (174) assay, the group Solid includes five patients classified as having local osteolytic type of hypercalcemia ( ). Note the different scale of the y-axes. (Data from Budayr et al. Burtis et al. and Blind et al.; reprinted from Blind E, Nissenson RA, Strewler GJ. Parathyroid hormone-related protein. In Becker KL, Bremner WJ, Hung W, et al [eds]. Principles and Practice of Endocrinology and Metabolism, 2nd ed. Philadelphia, JB Lippincott, 1985.)
Figure 40-2 Processing of pro-opiomelanocortin (POMC) in normal pituitary (hatched bars), intermediate lobe (open bars), and nonpituitary neoplasms (solid bars). ACTH, adrenocorticotropic hormone (corticotropin); CLIP, corticotropin-like intermediate lobe peptide; END, endorphin; LPH, lipotropin; MSH, melanocyte-stimulating hormone. (Adapted from Schteingart DE. Ectopic secretion of peptides of the proopiomelanocortin family. Endocrinol Metab Clin North Am 1991; 20:453-471.)
**Figure 40-3** Proposed explanation for the pathogenesis of hypoglycemia with nonislet cell tumors. FFA, free fatty acid; GH, growth hormone; IGF, insulin-like growth factor; IGFBP, IGF-binding protein. (Adapted from Zapf J. IGFs: function and clinical importance. 3. Role of insulin-like growth factor (IGF) II and IGF binding proteins in extrapancreatic tumour hypoglycaemia. J Intern Med 1993; 234:543552.)
Figure 41-1 Normal human intestine stained with chromogranin A (Chrom. A) to delineate neuroendocrine cells. The cells are scattered in the intestinal mucosa.
Figure 41-2. Histopathology of classic well-differentiated midgut carcinoid tumor.
Figure 41-3 Biosynthesis and metabolism of 5-hydroxytryptamine (5-HT) (serotonin).
Figure 41-4 The tachykinin family of peptides shares the same carboxyl terminus. Neuropeptide-K is a prohormone containing neurokinin-A, which can be spliced off.
Figure 41-5 Chromatography samples of plasma from a patient with carcinoid before flush (upper panel) and during flush (lower panel). Note the significant increase in eledoisin-like peptide as well as in neuropeptide-K.
Figure 41-6 The glycoprotein chromogranin-A and related peptides.
**Figure 41-7** Schematic drawing of an enterochromaffin cell. The initial step in 5-hydroxytryptamine (5-HT) synthesis is carrier transport of the amino acid tryptophan from blood into the cell across the cell membrane. Intracellular tryptophan is first converted to 5-hydroxytryptophan (5-HTP), in turn converted to 5-HT and stored in secretory granules. The transport of 5-HT into granules requires vesicular membrane transporters (VMATs). Via the basal lateral membrane, 5-HT can be released into the circulation. There is also a membrane pump mechanism in the cell membrane responsible for amine reuptake. A minor part of 5-HT can also be released into the gut lumen. Monoamine oxidase (MAO) degrades 5-HT to 5-hydroxyindoleacetic acid (5-HIAA). Peptide prohormones are synthesized in the rough endoplasmic reticulum (RER) together with chromogranin-A (CgA) and other granula proteins. The products are transported to the Golgi apparatus (GA) for packaging into prosecretory granules. On stimulation, the secretory products are released from the granules by exocytosis.
Figure 41-8 Carcinoid syndrome before and after provocation. A, Before flush provocation. B, Same patient after pentagastrin-stimulated flush.
Figure 41-9  Long-lasting chronic flushing in a patient with long-standing carcinoid disease. Note the telangiectases.
Figure 41-10 The patient has lung carcinoid and carcinoid syndrome with severe, long-standing flushing, lacrimation, and a swollen face.
Figure 41-11  Tachykinin levels (TKLI) after stimulation with pentagastrin in patients with classic midgut carcinoids. Pretreatment for 15 minutes with somatostatin causes inhibition of tachykinin release and inhibition of the flush reaction (00). P, placebo.
Figure 41-12 Plasma levels of chromogranin-A (CgA), CgB, and CgC in patients with various neuroendocrine tumors. EPTs, endocrine pancreatic tumors; MEN-1, multiple endocrine neoplasia type 1.
Figure 41-13 Bronchial carcinoid.  
A, Somatostatin-receptor scintigraphy in a patient with a bronchial carcinoid.  
B, CT scan in the same patient.
Figure 41-14  Positron emission tomography (PET) with \(^{11}C\)-5-hydroxytryptophan. Note the metastasis in the liver.
Figure 41-15 Diagnostic algorithm for patients with carcinoid tumor. CgA, chromogranin-A; $^{11}$C-5-HTP, $^{11}$C-5-hydroxytryptophan; 5-HIAA, 5-hydroxyindoleacetic acid; 5-HT, 5-hydroxytryptamine; NPK, neuropeptide-K; PET, positron emission tomography; SRS, somatostatin-receptor scintigraphy; sst 15, somatostatin-receptor subtypes 15.
Figure 41-16 Molecular structure of human somatostatin-14, octreotide acetate, and lanreotide.